Do Weak Institutions Prolong Crises?

On the identification, characteristics, and duration of declines during economic slumps[*]

Richard Bluhm[†] Denis de Crombrugghe[‡] Adam Szirmai[§]

January 11, 2016

Manuscript

Abstract

Economic slumps can be characterized statistically as interruptions of a positive growth regime by a sharp downward shift coinciding with a sequence of two trend breaks. Examining 138 countries over the period 1950-2008, we identify 58 episodes exhibiting such a pattern and investigate the duration of the decline phase. Some declines last very long and we put several likely contributing factors to the test. We find evidence that weak political institutions precede crises and that political reforms tend to follow them. Strong political institutions, such as imposing constraints on the executive, shorten the duration of declines; ethnic cleavages on the contrary prolong them. However, there is a marked interaction effect between institutions and ethnic cleavages, suggesting that the adverse effects of fractionalization can be overturned by strong political institutions.

Keywords: economic slumps, crisis duration, institutions, structural breaks, ethnic fractionalization

JEL Classification: O43, O11, C41, F43

---

[*]This paper has been presented at ETH Zurich (2015), UWarwick (Coventry, 2014), ABCDE (Washington DC, 2014), AFD (Paris, 2013), DIAL (Paris, 2013), NIPF-DEA (Neemrana, 2013) and Ifo (Dresden, 2013). We have greatly benefited from discussions with several participants. In particular, we would like to thank Agustín Casas, Oded Galor, Martin Gassebner, Melanie Krause, Nicolas Meisel, Thomas Roca, Ajay Shah, Jan-Egbert Sturm, Kaj Thomsson, and Bart Verspagen for helpful comments and suggestions. We gratefully acknowledge financial support from the Agence Française de Développement (AFD). The usual disclaimer applies.

[†]*Corresponding author.* Leibniz University Hannover, Institute of Macroeconomics, e-mail: *bluhm@mak.uni-hannover.de*

[‡]Maastricht University, School of Business and Economics, Department of Quantitative Economics, e-mail: *d.decrombrugghe@maastrichtuniversity.nl*

[§]UNU-MERIT, Maastricht University, Maastricht Graduate School of Governance, e-mail: *szirmai@merit.unu.edu*

1 Introduction

The last sixty years of growth have been far from steady. For every "growth miracle" we can easily find a counterpart in the form of a collapse. For example, the East Asian miracle was interrupted by the Asian financial crisis, China's take-off in 1978 was preceded by decades of adverse economic policies, Latin America was frequently rocked by political turmoil and economic volatility, and several African nations went from "up and coming" in the 1950s to stagnation after 1973. Moreover, during the post-war period, there is a long list of relatively short-lived advanced economy crises. The growth and development literature has often stressed the role of positive growth spurts, while in this paper we wish to direct attention to contractions. What can we learn from such abrupt changes in growth? Why do some countries deal better with negative growth shocks than others?

The instability of growth is of great concern in economics because it affects short-run and medium-run growth performance (e.g. Ramey and Ramey, 1995). A growing literature on trend breaks has established that growth is often not steady but instead characterized by switching among growth regimes (e.g. Ben-David and Papell, 1995; Jones and Olken, 2008; Papell and Prodan, 2014). Growth does usually not follow a constant trend but consists of qualitatively different episodes, such as crises, recoveries, periods of stagnation, and accelerations. This perspective provides new stylized facts. For example, growth is relatively easy to ignite (Hausmann et al., 2005; Jong-A-Pin and De Haan, 2011) but much harder to sustain (Berg et al., 2012). However, the negative implications of volatile growth are just beginning to be explored. Long-lasting slumps can nullify decades of positive growth and there is no guarantee that lost potential output is ever fully recouped (Cerra and Saxena, 2008). It thus becomes important to ask, why do some declines last so much longer than others?

A potential explanation is that the duration of declines during slumps is driven by the prevailing structure and quality of institutions. Institutions create specific political and economic incentives, solve or worsen coordination failures and define the set of feasible policies. Seminal contributions to the institutions and growth literature link stronger institutions to higher *levels* of GDP per capita (Acemoglu et al., 2001). Others have shown that strong institutions and political stability bring about reduced output volatility (e.g. Acemoglu et al., 2003; Mobarak, 2005). However, there is still a lack of evidence convincingly linking institutions to short and medium-run growth dynamics.

Each type of growth episode has distinct characteristics. We can analyze the switching among growth regimes, the rate of growth within a regime, the duration of a regime, or even the typical sequence of regime switches that makes up a growth path. Out of this plethora of possibilities, this paper focuses on three points. First, how can we identify large economic slumps empirically? Second, is there any evidence of institutional change when slumps occur? Third, conditionally on the occurrence of a slump, do weak institutions prolong the duration of the decline phase? We single out the decline phase because it is delimited by two turning points, or breaks in the growth regime, and is usually followed by a recovery phase that does often not end with yet another turning point but leads the way to continued growth.

This paper makes the following contributions to the literature. With respect to the first question, we propose a strictly statistical characterization of economic slumps as interruptions of a positive growth regime by a sharp downward shift coinciding with a sequence of two trend breaks. We demonstrate the effectiveness of the concept by identifying and recognizing 58 historical episodes with this specific pattern in the post-

1950 world. With respect to the second question, our event analysis shows that positive institutional change is triggered by the economic crises (which are preceded by weak institutions). With respect to the third issue, we find robust evidence that not only institutional features but also ethnic divisions contribute jointly to the length of crises.

We specifically focus on the *duration* of declines, as the onset of economic crises can be triggered by a variety of external or internal factors which are not (always) linked to weak institutions. However, the way a country deals with a negative shock, and whether the decline phase takes longer than in other countries, depends on the political system's ability to react with coordinated policies and avert outright social conflict. This notion derives from a large body of political economy theory putting social tension and the ability of resilient political institutions to manage such conflict at the center of development theory (e.g. Acemoglu and Robinson, 2006; North et al., 2009; Besley and Persson, 2011). Some of these theories argue that weakly institutionalized societies, or 'limited access orders', are prone to collapses, and that the declining rents associated with a crisis undermine the institutional set-up and the prevailing political arrangements (e.g. North et al., 2009). Weak institutions thus bring with them an increased vulnerability to crises and potentially much longer declines once slumps occur. Similar mechanisms are suggested in the literature on institutions and macroeconomic volatility (Acemoglu et al., 2003). Even if sound policy responses are available, a combination of coordination failures, rent seeking and power struggles combined with ethnic cleavages may lead to longer declines in weakly institutionalized environments.

Ethnic heterogeneity itself has been linked to a variety of coordination failures leading to inadequate policies, low provision of public goods and conflict. Greater diversity may also be beneficial. In fact, a particular level of heterogeneity may be optimal (Ashraf and Galor, 2013) and necessary to reap the advantages of skill complementaries in highly diversified economies (Alesina and Ferrara, 2005). Even if there are no benefits to diversity, its negative effects may become muted as "richer societies have developed institutional features that allow them to better cope with the conflict element intrinsic in diversity" (Alesina and Ferrara, 2005, p. 763). Indeed, one of our key contributions is to show that in the context of economic declines the (negative) effects of ethnic heterogeneity depend on political institutions and *vice versa*.

The 'delayed stabilizations' literature provides another lens on how conflict over the costs of adjustment always leads to delayed reform (Alesina and Drazen, 1991) and how cabinet or presidential systems affect the timing of the adjustment (Spolaore, 2004). Similarly, the model of policy non-adoption by Fernandez and Rodrik (1991) outlines a status-quo bias due to uncertainty about the benefits of reform when the losers cannot be guaranteed compensation *ex post*. These ideas naturally carry over to the agreement on policy responses during slumps where more group heterogeneity can increase delay.

Our findings support some of these theoretical perspectives. First, we present evidence that weak political institutions precede slumps and are followed by political reforms. Second, we show that longer decline phases are robustly linked to initially weak institutions and particularly strongly to a measure of ethnic divisions. Ethnic heterogeneity is especially important for understanding declines in Sub-Saharan Africa. Third, we find that political institutions and ethnic fractionalization interact. In weakly institutionalized and highly fragmented societies declines last considerably longer than in more homogeneous countries or countries with stronger constraints on the executive. This finding is at odds with the delayed reform literature. In a related contribution, Bluhm and Thomsson (2015) offer a reconciliation in the form of a theory that links the duration of the decline phase to a commitment problem between winners and losers of the recovery process that arises only when institutions are weak. Finally, we provide evidence suggesting that the effects of political institutions and ethnic fractionalization run exclusively through the duration of declines and not the average rate of contraction.

The paper is structured as follows. Section 2 outlines the identification of slumps and defines the decline phase. Section 3 briefly discusses the data and characteristics of the estimated slumps, and provides evidence of endogenous institutional change. Section 4 analyzes the duration of the decline phase and discusses the results. Section 5 concludes.

2 Identifying slumps

Restricted structural breaks

Beginning with Pritchett's (2000) classification of post-World War II growth experiences into "Hills, Plateaus, Mountains, and Plains", a growing literature analyzes the characteristics of different types of growth episodes. These papers usually employ tests of structural stability to define and identify the episode of interest.[1]

Not every change in the growth rate of GDP per capita amounts to a regime switch. The main advantage of statistical tests for multiple structural breaks over any set of predefined economic criteria is that they allow for an inferential approach to identifying growth regimes and measuring their duration. Methods based on deterministic economic criteria, such as annual versions of the NBER two-quarter rule, cannot discriminate among multiple plausible starting points or assess whether an episode truly constitutes a departure from the previous growth regime. Especially when we are interested in deep crises that exclude smaller business cycle fluctuations, we need some notion of statistical significance. Standard structural break methods accomplish this but generally leave the particular type of structural change unspecified. As a result, these tests may not identify the theoretically desired type of regime switch but rather any form of significant change which must then be classified *ex post*. Furthermore, while standard break estimators work well for identifying growth spurts, they seem to perform poorly when it comes to identifying growth collapses.

To improve the identification of what we interchangeably refer to as deep recessions, slumps, or growth collapses, Papell and Prodan (2014) propose a *two-break model with parameter restrictions*. They demonstrate that this modified structural change approach consistently identifies well-known slumps, such as the Great Depression in the United States. The key innovation is to impose features of the desired pattern directly instead of searching for *unrestricted* structural changes. Their two-break model accounts for three growth regimes (a pre-slump regime, a contraction-recovery regime, and a post-slump regime) and places sign restrictions onto the estimated coefficients to ensure the breaks occur in the desired direction. Since this approach is a version of Bai's (1999) sequential likelihood ratio test, the number of slumps – which is not known in advance – can then be estimated by recursively applying the model on ever smaller sub-samples until all breaks in the GDP per capita series have been found. While Papell and Prodan (2014) focus on the question whether growth in a few developed countries eventually returns to its pre-slump trend path, we apply a variant of this method to identify slumps in a large

[1]See, for example, Hausmann et al. (2005), Jones and Olken (2008), or Berg et al. (2012). An exception are the Markov-switching models used in Jerzmanowski (2006).

sample of countries over the period from 1950 to 2008.

The restricted structural change approach can easily be modified in principle to allow for other plausible structures, such as three-break models (e.g., to estimate a pre-slump regime, a decline, a recovery and a post-slump regime). We have experimented with variants of a three-break approach using different parameter restrictions, but we prefer the two-break approach for the following reasons: estimating three or more breaks for each slump quickly becomes computationally expensive, often identifies the same episodes, and does not necessarily provide a better estimate of the starting date than the simpler two-break model. Furthermore, with annual data, the contraction phase alone is often too brief to allow the observation of two separate breaks.[2]

We define slumps according to three criteria. First, a slump is a *departure from a previously positive trend*. Second, a slump must begin with *negative growth in the first year*. Third, all slumps should be *pronounced regime switches* and not just minor business cycle fluctuations. The precise meaning of 'pronounced' will vary and depend on the country's idiosyncratic growth process. Note that a slump as defined here is much more than a short-lived recession: it is a sharp (but often temporary) departure from the previous growth path. We focus on growth in GDP per capita, since we are primarily interested in the welfare consequences of slumps and not in aggregate output per se.

We capture these criteria in the following partial structural change model:

$$y_t = \alpha + \beta t + \gamma_0 \mathbf{1}(t > tb_1) + \gamma_1 (t - tb_1)\mathbf{1}(t > tb_1) + \gamma_2 (t - tb_2)\mathbf{1}(t > tb_2) + \sum_{i=1}^{p} \delta_i y_{t-i} + \epsilon_t \quad (1)$$

where y_t is the log of GDP per capita, β is a time trend, γ_0 is the coefficient on an intercept break occurring together with a trend change (γ_1) after the first break at time tb_1, γ_2 is the coefficient for a second trend change occurring after the second break at time tb_2, $\mathbf{1}(\cdot)$ is an indicator function selecting the regime, p is the optimal lag order determined by the Bayesian information criterion (BIC) to parametrically adjust for the presence of serial correlation, and $\{\epsilon_t\}$ is a martingale difference sequence satisfying $\mathbb{E}[|\epsilon_t|] < \infty$ and $\mathbb{E}[\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots] = 0$.

Equation (1) formalizes the notion that the evolution of GDP per capita around a slump is a simple function of time split into three different growth regimes: (1) a *pre-slump regime* from the beginning of the time series of a country until time tb_1, (2) a *slump-recovery regime* lasting from time $tb_1 + 1$ to time tb_2, and (3) a *post-slump regime* from time $tb_2 + 1$ onwards. The true location of the breakpoints is not assumed known but estimated within the model. We impose *two restrictions* to make sure we only select breaks meeting our definition of slumps. First, we require $\beta > 0$, so that growth must be positive in the years before a slump begins. Second, we also impose the condition that $\gamma_0 < 0$, so that a slump always starts with a drop in the intercept. The intercept shift implies that we assume that there is an instantaneous drop at the start of the slump. Slope shifts are left unrestricted, so that the model can catch unfinished slumps (e.g., declines from tb_1 onwards, possibly lasting until the end of a country's time series).

We implement the sequential break search algorithm as follows. First, we fit the structural change model specified in equation (1) for all possible combinations of tb_1

[2]Let $q = T - 2\tau T - h$, where T is the sample size, τ is the trimming fraction and h is the minimum distance between breaks, then the two-break model estimates $(q^2+q)/2$ regressions for the first iteration, while a three-break model already requires $\sum_{i=1}^{q}(i^2+i)/2 = (1/12)q(q+1)(2q+4) - 1$, with $q = T - 2\tau T - 2h$ to now allow for three breaks. Additional results are available on request.

and tb_2. We always exclude 5% of the observations at the beginning and end of the sample to avoid registering spurious breaks (we need at least a short time series before (after) the first (last) break to identify the adjacent regime). Second, we compute the sup-W test statistic, that is, the supremum of a Wald test of the null hypothesis of no structural change ($\mathbb{H}_0 : \gamma_0 = \gamma_1 = \gamma_2 = 0$) over all pairs of break dates implying estimates satisfying both restrictions. Third, we bootstrap the empirical distribution of the sup-W statistic. If the bootstrap test rejects the null at the 10% significance level, we record the break pair $(\widehat{tb}_1, \widehat{tb}_2)$ and split the sample into a series running until the first break and a series starting just after the second break. The process starts again on each subsample until the bootstrap test fails to reject the null hypothesis of no breaks or the sample gets too small ($T \leq 20$). A key issue in evaluating the statistical significance of breakpoints is that the individual Wald tests over which the sup-W statistic is computed are not independent. Asymptotic tests have been derived but they tend to underreject in finite samples (Prodan, 2008) and an asymptotic distribution for our particular version of restricted structural change is not available. We construct a bootstrap Monte Carlo test in order to circumvent both issues. Appendix A gives a formal description of the break search algorithm and the bootstrap.

The structural break methods applied in this paper assume that the logarithm of GDP per capita is regime-wise trend stationary. This is not a trivial requirement. Ever since the issue was first raised by Nelson and Plosser (1982), a vibrant literature has been debating the question whether most GDP series are unit-root processes or can be considered trend stationary. More recently the debate has shifted. A process that is subject to structural breaks is an intermediate case. Broken-trend stationarity only implies that, within each regime, growth can be approximated by a deterministic trend, but from one regime to the next the trend may change due to (semi-)permanent shocks. This allows for a flexible description of the growth process as several different types of breaks can occur. In fact, there is mounting evidence that once breaks are incorporated, many of the GDP series previously thought to have unit roots may in fact be broken-trend stationary (e.g. Zivot and Andrews, 1992; Ben-David and Papell, 1995). Broken trends blur the conceptual distinction. A unit root process can be thought of as a trend-stationary process with a trend that changes every year (or at another observed frequency).

We do not attempt to characterize all types of breaks an economy can experience, or to formally test for unit roots. Our approach is flexible and allows for different growth regimes occurring before, during and after an unknown number of slumps. We assume that there is some structure in the growth process, but do not assume that it is necessarily generated by neoclassical steady-state growth, endogenous growth or any other standard growth model. In fact, Aguiar and Gopinath (2007) recently highlighted that growth in emerging markets can be characterized by shocks to trend growth rather than transitory fluctuations around a stable trend. Thus, under certain conditions, broken trends are compatible with various models of aggregate output.

The duration of declines

Within a slump, we separate the decline from the recovery phase and focus solely on the decline phase. This is based on the conjecture that these two processes are potentially subject to very different dynamics and are explained by different covariates. This "to the bottom" approach stands in contrast to the earlier literature which typically focuses on the entire duration of the slump (until the pre-slump level of GDP is regained).

Instead, we argue that political institutions and ethnic cleavages matter particularly for the duration of the decline phase where policy solutions need to be agreed upon to achieve a turnaround, as opposed to the length of the recovery process which depends on the success of the chosen policy and, more generally, the type of crisis. In any case, the decline phase is naturally delimited by two turning points, or switches in the growth regime. By contrast, the recovery phase ends when a previous high of income is reattained, which does not imply a change in the growth process.

We define the end of a slump to have occurred with certainty in the first year $a > \widehat{tb}_1$ where $y_a \geq y_{\widehat{tb}_1}$. In other words, a slump is over as soon as the level of GDP per capita preceding the slump has been recovered; until then, the slump is continuing.[3] It is important to note that the end of the slump does not in general coincide with the second break and is used only as a device to identify the trough. Once the endpoint of a slump is known, *the trough is simply the year with the lowest level of GDP per capita during the slump*. The duration of the slump is censored if GDP per capita does not reach the pre-slump level again by the end of the sample. In such a case, even if GDP per capita seems to be recovering, we do not know how long the slump may last. The censoring indicator is defined as $c = \mathbf{1}(\max_{j \in (\widehat{tb}_1, T]} y_j < y_{\widehat{tb}_1})$.

Given the set of possible end years $A = \{a \mid a \in (\widehat{tb}_1, T] \text{ and } y_a \geq y_{\widehat{tb}_1}\}$, define $a_0 = \min A$, corresponding to the (certain) end of the slump. If the set A is empty, then the slump is unfinished, and the length of the episode is censored. We estimate the trough to occur at time:

$$\hat{t}_{min} = \begin{cases} \operatorname{argmin}_{j \in (\widehat{tb}_1, a_0]} y_j, & \text{if } c = 0 \\ \operatorname{argmin}_{j \in (\widehat{tb}_1, T]} y_j, & \text{if } c = 1. \end{cases} \tag{2}$$

A provisional trough occurs when y_t attains a minimum after $\widehat{tb}_1$. The duration of the decline phase lasting from the beginning of the slump to the estimated trough will be denoted $\tilde{t}_D = \hat{t}_{min} - \widehat{tb}_1$.

These definitions also imply that in some cases we date the trough after the estimated second break, which is purely a consequence of allowing for unfinished episodes. If the slump is still ongoing, the second break may have been placed at a point that maximizes the Wald statistic but does not correspond to the start of a new growth regime. A solution avoiding this problem would be to test if a restricted one-break model fits better than a restricted two-break model for those cases. Alternatively, we can (and later will) examine if our results are robust to the exclusion of these episodes. For these unfinished spells, the true trough may lie in the future, that is, beyond the end of the sample. Treating such spells as censored implies that in the later analysis we only incorporate the information that (certain) exit from the slump has not yet occurred after a duration $\tilde{t}_D$.

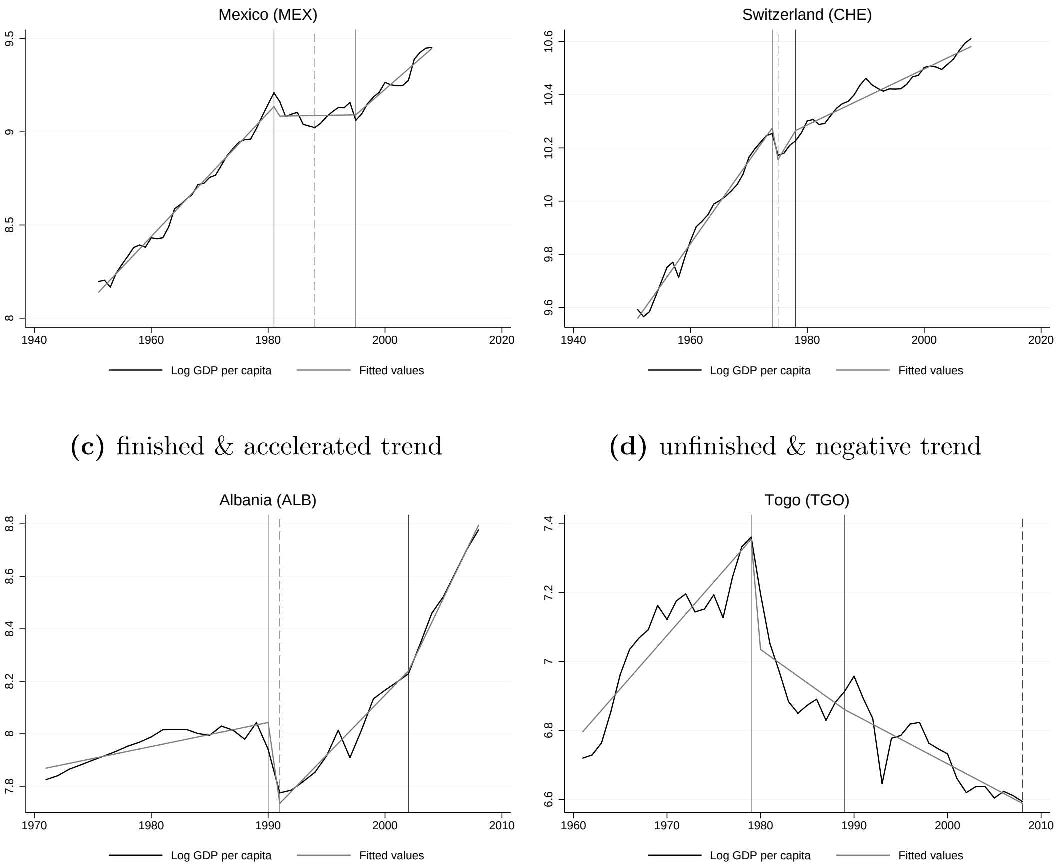
Figure 1 illustrates the diversity of slumps identified by this method. Panel (a) shows a finished slump in Mexico where the trend growth rate is nearly unchanged after the second break. The slump begins in 1982 and encompasses more than a decade of political volatility, hyperinflation, high debt and low growth. The trough is found in 1988. Another short downturn occurs during the Tequila crisis in 1994 after which the Mexican economy returns to its pre-1982 growth path. Panel (b) shows a finished slump in Switzerland where the trend growth rate decelerated after the second break. In 1975, the Swiss

[3] This also implies that we exclude a small number of episodes found by the sequential algorithm if these begin before the previous slump is certain to have ended. Avoiding this issue would imply estimating a third breakpoint (see footnote 2).

Figure 1 – Four types of slumps

(a) finished & unchanged trend **(b)** finished & decelerated trend

(c) finished & accelerated trend **(d)** unfinished & negative trend



Note(s): Models refitted using the estimated breaks $\widehat{tb}_1$ and $\widehat{tb}_2$ but without the optimal AR(p) terms to emphasize the trend breaks. The bold vertical lines are at $\widehat{tb}_1$ and $\widehat{tb}_2$, respectively. The dashed vertical line indicates $\hat{t}_{min}$.

economy was strongly affected by the oil crisis of the mid-1970s, leading to a 7.87% drop in GDP per capita within one year. After the second break, Switzerland enters a low growth regime typical for the high income economies in Western Europe of the 1980s and 1990s. Panel (c) shows a finished slump in Albania occurring at the time of the post-communist transition with an accelerated trend after the second break. The estimated first break occurs in 1990, the trough is located in 1991, and the second break occurs in 2002, a few years after the end of the slump. While the duration of the decline phase is only one year, output contracted considerably. GDP per capita in 1991 was 15.32% lower than in 1990. Last but not least, panel (d) shows an unfinished slump with a continuing decline in Togo. Togo grew rapidly for over a decade following independence from France in 1960 but then experienced a dramatic collapse. The first break occurs in 1979, while the second break occurs when the economy partially recovers and then declines even further (which is not very meaningful but maximizes the Wald statistic). Togo's GDP per capita did not recover its pre-slump level for the next 29 years. At the end of the observed period, the decline is still ongoing and the provisional trough coincides with the censoring cutoff in 2008.

3 Data and characteristics of slumps

We apply the sequential algorithm to the entire Penn World Table (v7.0) yielding a total of 58 slumps between 1950 and 2008.[4] We deliberately stop in 2008 to avoid the global recession of 2009 which is too close to the end of the sample for reliable break estimation. The mean duration from the first break to the trough is about 7.7 years and the median duration is three years. Ten out of the 58 slumps are censored. For these spells the location of the trough is not yet definitive. Table B-2 in the Appendix lists all episodes and provides some summary statistics.

Many covariates used in the sequel are from well-known sources and will be discussed only summarily here. We include four major categories of variables: 1) a variety of measures for different aspects of *institutions, politics and social conflict*, 2) macroeconomic indicators of *prices, trade and exports*, 3) a set of variables for domestic and international *finance*, and 4) several *other growth determinants* (such as life expectancy or years of schooling). Table C-1 in the Appendix provides an exhaustive list of all variable names, data sources and a basic set of summary statistics for the data used throughout the paper. Some data may not be entirely satisfactory but are simply the best available. For example, we use the Polity IV database as our primary proxy for institutional development because of a lack of other time series data capturing the characteristics of political and economic institutions.

We observe several well-known growth collapses and deep recessions. Most slumps begin between the 1970s and the early 1990s. Seven downbreaks occur following the oil shock in 1973–1974, eleven declines begin between 1979 and 1981 during the debt crisis of the early 1980s, and nine slumps follow the post-communist transitions of 1989–1990. Due to trimming and the cut-off in 2008, we find no beginnings of slumps in the period of the early 2000s and tranquil mid-2000s. Generally, the period between the 1970s and early 1980s is marked by heightened volatility, as has been documented in a number of studies (Easterly et al., 1993; Rodrik, 1999; Pritchett, 2000; Jones and Olken, 2008).

Table 1 summarizes the distributions of depth, duration, and number of spells across income groups and geographical regions. For this purpose, we define the depth of a decline as the percent decrease of GDP per capita at the trough relative to its pre-slump level. We detect considerably deeper slumps in low-income and middle-income countries than in high-income (OECD) countries. The spread of depth and duration is very large. High-income (OECD) countries experience relatively short declines with a comparatively soft landing. The median duration is only one year with a mean depth of about -7.1%. In the middle, there is little substantial variation between non-OECD high-income countries and upper/lower-middle-income countries. In all of these three groups, the mean depth is in the range of -20.8% to -27.4% and the mean (median) duration varies between about 5.4 to 6 (2 to 3) years. Low-income countries experience the most dramatic declines. Both mean and median duration are about 16 years, with an associated average depth of -34.2%. Interestingly, the number of spells itself is distributed relatively evenly across the different income groups, suggesting that developed countries, too, experience their

[4] We run the algorithm on countries with a population of at least one million and at least 20 years of data. In addition, we discard some episodes that are driven by positive breaks in the slope coefficient but fail the negative growth criterion due to the presence of the AR(p) terms. A simple rule is applied to these cases. We define a valid episode as an interval of two break dates $\widehat{tb}_1, \widehat{tb}_2 \in [\tau T, (1-\tau)T]$ satisfying: $\exists j \in (\widehat{tb}_1, \widehat{tb}_2]$ such that $\min y_j < y_{\widehat{tb}_1}$, where τ is the trimming fraction and T is the length of the estimation sample. This rule *only* requires that a actual contraction occurs within the range of the two estimated breaks, otherwise there is no slump.

Table 1 – Depth and Duration by Income Level and Geographical Region

	Mean Depth	Mean Duration	Median Duration	Number of Spells	Censored Spells	Number of Countries
Income Level						
High Income (OECD)	-7.12	2.00	1	12	0	29
High Income (Other)	-20.84	5.38	2	8	1	12
Upper Middle Income	-21.20	5.39	2	16	2	30
Lower Middle Income	-27.40	6.00	3	11	3	34
Low Income	-34.17	15.75	16	11	4	33
Geographical Region (detailed)						
East Asia & Pacific	-13.63	2.30	2	10	0	17
Eastern Europe & Central Asia	-19.70	3.40	2	5	0	10
Europe (excl. Eastern Europe)	-8.37	1.50	1	6	0	22
Latin America & Caribbean	-17.34	5.27	3	15	1	23
Middle East & North Africa	-33.24	8.66	9	7	3	17
North America	-2.51	1.00	1	1	0	2
South Asia	-5.33	1.00	1	1	0	6
Sub-Saharan Africa	-37.14	17.74	16	13	6	41
Total	-21.87	7.69	3	58	10	138

Note(s): Depth is defined as the percent decrease in GDP per capita at the trough relative to GDP per capita before the slump (not log difference). Mean and median duration are expressed in years. As a result of some spells being censored, both mean duration and depth are underestimated. The number of countries refers to countries with more than one million inhabitants and more than 20 observations of GDP per capita in a particular income group or region.

fair share of (milder) volatility.

The geographical distribution reveals three interesting patterns. First, Sub-Saharan Africa and the Middle East & North Africa are the two regions experiencing both the deepest and longest declines. Their experience is striking in comparison to other regions. The mean decline is -37.1% in Sub-Saharan Africa and -33.2% in the Middle East & North Africa, which is about double of the average decline in Latin America & the Caribbean. The duration is longest in Sub-Saharan Africa, with the median spell lasting 16 years and the mean spell lasting over 17 years. Declines are shorter in the Middle East & North Africa, where the mean and median do not exceed nine years. Both regions also have the most censored/unfinished spells due to their long duration. Second, countries in Latin America & the Caribbean experienced slumps most frequently, but the average decline was only moderately deep and lasted for about five years. Third, when comparing Eastern Europe & Central Asia to the East Asia & Pacific region we find similar mean and median durations but much deeper slumps in the former. As expected, there are comparatively few, short and mild declines in North America, Europe (excluding Eastern Europe), and – more surprisingly – South Asia.

Table 1 suggests a relatively strong association of both the mean duration and mean depth of the decline phase with different income levels. This is particularly interesting, since we subsequently model these observed differences in duration between high and low income economies with more fundamental factors such as institutions and ethnic fractionalization. Furthermore, there is substantial regional heterogeneity which will have to be taken into account in the analysis.

Is there evidence of institutional change occurring before, during or after a slump? To study this question descriptively, we employ an event methodology often used in the literature on currency and banking crises (e.g. Eichengreen et al., 1995; Gourinchas and Obstfeld, 2012). The basic idea is to use dummy variables indicating the imminence or recentness of the start of the slump as a means of detecting changes in the relative mean of each time-varying covariate. The coefficients of the dummies measure if and how the covariate changes around the time the slump hits, and their standard errors quantify the associated uncertainty.

We run the following regression for each measure of institutions: $x_{it} = \sum_{s=-5}^{5} \delta_{t,\widehat{tb}_1+s}\beta_s + \mu_i + \epsilon_{it}$ where $\delta_{t,\widehat{tb}_1+s}$ is the Kronecker delta which is equal to one if $t = \widehat{tb}_1 + s$ and zero otherwise, β_s are coefficients, μ_i is an unobserved country effect and ϵ_{it} is an idiosyncratic error term. We set $s \in \{-5, \ldots, 0, \ldots, 5\}$, so that the result is an 11-year window around the break date $\widehat{tb}_1$. The first year of the slump is $s = 1$ corresponding to $t = \widehat{tb}_1 + 1$. The standard errors are robust to heteroskedasticity and also autocorrelation within both country *and* time clusters (Cameron, Gelbach, and Miller, 2011). We plot the estimates of the coefficients (including 95% confidence bands) as they represent *the conditional expectation of x_{it} at time s relative to "normal" times.*[5]

Figure 2 – Institutions & Politics

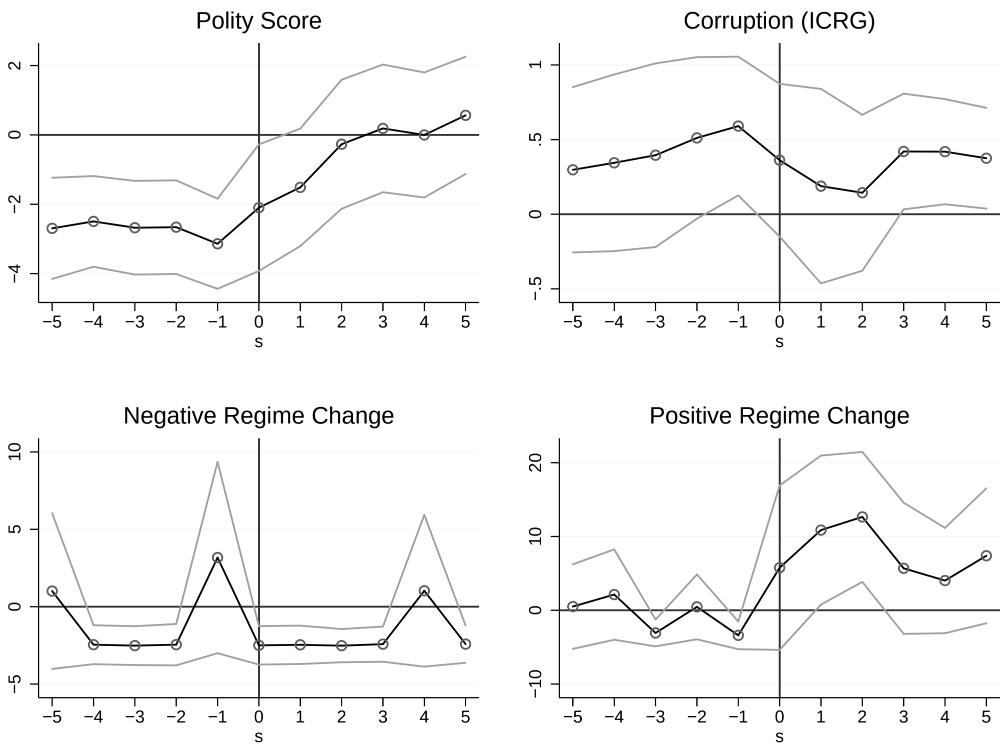


Figure 2 shows how certain institutional and political dynamics evolve around the downbreak. The Polity score is much lower before a slump occurs, but increases towards normal levels thereafter. In the five years before a slump, the conditional expectation is between 2.5 and 3.1 points lower than in normal times and until the break date these differences are significant at the 5%-level. This suggests that prior deficiencies in institutions increase vulnerability to slumps *and* institutions start to improve when slumps occur. All the subcomponents of the combined Polity score, including *constraints on the executive*, exhibit very similar trends (not shown, available on request). Conversely, the ICRG's 6-point corruption indicator shows a moderate, yet insignificant, decrease in

[5] In this case, "normal" refers to all observations other than the 11 years around the downbreak. The working paper version of this paper also reports this analysis for a host of other growth determinants.

corruption in the first two years of a slump. The ICRG series suffers from low coverage; it begins only in 1984 while a majority of the slumps in our sample start earlier.

The association of reforms and slumps is confirmed by the time profile of the probabilities of negative or positive regime changes, measured as a minimum three-point downward or upward change in the Polity score. There is little evidence that negative regime changes precede downbreaks or systematically occur thereafter. Interestingly, there is an upward trend in the probability of positive regime changes from the eve of the slump onwards. The probability is 10-12% higher in the first and second year of a slump.

This pattern suggests a new stylized fact: slumps are often preceded by weak institutions and then abrupt negative growth creates room for political and economic reforms. Deep crises seem to increase the pressure on governments to pursue institutional change, illustrating the *endogenous nature of reforms*. This is not a trivial finding. The aggregate costs of reforms are generally easier to bear in good times than in bad, but a deep slump may alter the power balance and weaken the opposition towards reforms. It also means that we have to rule out endogenous feedback from crises to institutions in the empirical analysis that follows.

4 The duration of declines

Estimation strategy

We use parametric accelerated failure time (AFT) models to analyze the duration of declines. A key advantage of AFT models is their straightforward interpretation: they are akin to a classical log-linear regression model for the survival time. The hazard function and survival function are only indirectly characterized by the distribution of the error terms in the log-linear model. Like all survival models, they can easily deal with right censoring. If a country is observed to exit the decline phase at some time, its contribution to the likelihood is the probability of the recovery starting at that particular time (conditional on the decline phase lasting until that time). If there is no observed exit from the decline phase, then the observation is censored and only the survival probability enters the likelihood.

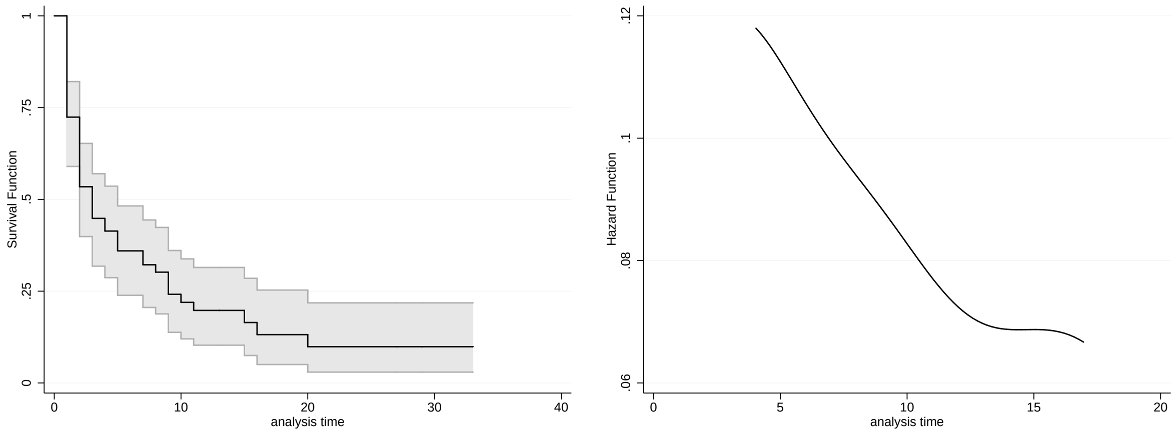
All parametric models assume a certain shape of the baseline hazard. We have no strong theoretical prior that the hazard function must follow a particular shape. We may expect some countries to exit rather quickly and others to take longer, but it is difficult to determine *ex ante* if remaining in the decline phase for very long leads to a deterioration of fundamentals and thus a decreasing hazard, or if the probability of exit is actually increasing because countries are bound to enter the recovery phase eventually.

Figure 3 shows the non-parametric Kaplan-Meier estimate of the unconditional survival function and the (smoothed) Gaussian hazard function. About 47% of the spells in our sample end after only two years of decline and the unconditional probability of exiting the decline phase is monotonically decreasing. Nevertheless, the shape of the conditional hazard may be very different. We take a flexible approach by first relying on a log-normal parameterization and then testing the robustness of our preferred specification under different distributional assumptions. We provide a more detailed description of how log-normal AFT models are estimated in Appendix D.

Let analysis time be $\tilde{t}$, where $\tilde{t} \equiv t - t_0$ and $t_0 = \widehat{tb}_1$, so that we can refer to the calendar times t and t_0 when necessary. We specify the following regression equation for

Figure 3 – Unconditional Survival and Hazard Functions

(a) Kaplan-Meier Survival Function **(b)** Gaussian Hazard Function



Note(s): The Kaplan-Meier survival curve is a non-parametric estimate of the probability of remaining in the decline state at each unit of analysis time. 95% confidence intervals are shown in grey. The corresponding hazard function has been smoothed using a Gaussian kernel with boundary adjustment and bandwidth 3.

crisis durations in AFT form:

$$\ln \tilde{t} \equiv \ln(t - t_0) = \alpha + \beta INS_0 + \gamma ELF + \mathbf{x}_0' \boldsymbol{\xi} + \mathbf{z}_t' \boldsymbol{\zeta} + \epsilon_t \tag{3}$$

where INS_0 is a measure of institutions fixed at t_0, ELF is a time-invariant measure of ethnic fractionalization, $\mathbf{x}_0 = (x_{0,1}, x_{0,2}, \ldots, x_{0,k})'$ is a $k \times 1$ vector of controls fixed at t_0 often including region fixed effects, $\mathbf{z}_t = (z_{t,1}, z_{t,2}, \ldots, z_{t,m})'$ is a $m \times 1$ vector of strictly exogenous time-varying controls, and – for the log-normal model – ϵ_t is distributed $\mathcal{N}(0, \sigma_\epsilon^2)$. All parameters, including σ_ϵ^2, are estimated with Maximum Likelihood (ML). Unfinished spells are accounted for but censoring is assumed to be independent of the duration process. Our coefficients of interest are β and γ. We suppressed the country-spell index to simplify the exposition.

The estimated coefficients are semi-elasticities of the expected duration with respect to the covariates, or elasticities if the covariate is in logs. The term 'accelerated failure time' derives from the interpretation of the implied effects. If the coefficient of the covariate is positive, then the expected duration until the event is prolonged by larger realizations of the covariate. In our case, this is equivalent to delayed exit from the decline phase (later start of the recovery). If the coefficient is negative, then the expected duration is shortened and the recovery will start earlier.

A complication of using time-varying covariates is feedback from the duration to the covariates. If this occurs, then estimated coefficients are biased and the usual test statistics are invalid (Lancaster, 1990; Kalbfleisch and Prentice, 2002). In order to avoid this problem, we simply take the last pre-slump value of all potentially endogenous covariates at t_0, including our measure of political institutions, so that no feedback from slumps to the covariates is possible. Hence, we can rule out simultaneous causality. This is particularly important given that the previous section highlights that political institutions may endogenously respond to crises. On the other hand, ethnic fractionalization is assumed to be strictly exogenous; we do not expect the ethnic configuration of a country to change as a short-run response to a crisis.

The fact that countries can have several recurrent slumps is a minor concern in our application; only eight of the 58 spells in our data are not the first spell for a given country. To account for this dependence, we allow the variances of the parameter estimates to be correlated across spells in the same country. This procedure assumes that the sequence of repeated spells does not matter. We show in the robustness section that our results hold when this assumption is relaxed.

Dealing with at most 48 exits in 58 decline spells over the entire period of 1950 to 2008 requires a careful approach to model selection, since we have to match these episodes with data over the almost six decades spanned by them. Including many control variables with different patterns of missing data then easily results in small samples, so that including a large set of controls is not feasible. Even at more moderate sample sizes, care needs to be taken to guard against overfitting. To arrive at a parsimonious specification, we employ a two-step approach. First, we fit variable-by-variable regressions and reduce the set of controls based on statistical significance (p-value $< .1$). We select only those variables clearly exhibiting a correlation with the duration of declines. Results of this step are relegated to Appendix E. Second, using the smaller set of controls, we then extend our base specification in several ways and examine its robustness.

Results

We model the duration of declines as a function of executive constraints, ethno-linguistic fractionalization, initial GDP, and the real US interest rate. Constraints on the executive is our preferred proxy of institutional quality for two reasons. First, it is widely used in the empirical literature as a measure of institutional constraints placed on political actors and has already been linked to macroeconomic volatility (e.g. Acemoglu et al., 2003; Acemoglu and Johnson, 2005). Second, it is conceptually rooted in the economic theory of institutions, more so than any of the broader measures capturing wider aspects of the political regime (e.g. democracy or autocracy). Controlling for initial GDP matters, as executive constraints are correlated with the level of development and both potentially determine the duration of declines. The real US interest rate serves as a proxy for "good" or "bad" times in the global economy. It is especially important since we cannot parametrize duration dependence and include a full set of time effects at the same time.

For fractionalization, we use a measure from Desmet et al. (2012), who recently developed a very detailed set of estimates of linguistic diversity. They compute the probability that two randomly chosen individuals in a country belong to different ethno-linguistic groups at 15 levels of 'the language tree' (a genealogy). These new measures of fractionalization capture the historical nature of ethnic and linguistic differentiation into increasingly narrower groups over time. We use two variables at both extremes of the spectrum. ELF1 is the most aggregate level, measuring only crude distinctions such as Indo-European versus non-Indo-European languages. ELF15 is the most disaggregate level, differentiating among the language groups known today. Desmet et al. (2012) show that aggregate fractionalization matters more for civil conflict while the disaggregate level strongly predicts growth differentials. Hence, we use the latter as our primary measure.

The variable selection results reported in Appendix E show that the basic correlations are mostly as expected. One notable exception is the lack of correlation between conflict and the duration of declines. Stronger institutions are associated with shorter declines, regardless of the measure. Higher initial GDP in 1950 or at the first observed value predicts shorter declines. Conversely, higher fractionalization and a higher US interest rate predict longer declines. Yet these findings could be driven by omitted variables.

Table 2 – Additive effects of institutions and fractionalization

VARIABLES	(1) $\ln \tilde{t}$	(2) $\ln \tilde{t}$	(3) $\ln \tilde{t}$	(4) $\ln \tilde{t}$	(5) $\ln \tilde{t}$	(6) $\ln \tilde{t}$
Executive Constraints (INS_0)	-0.144***	-0.139**	-0.155**	-0.017	-0.069	-0.183***
	(0.048)	(0.055)	(0.061)	(0.094)	(0.102)	(0.065)
Fractionalization ($ELF15$)	0.014***	0.017***	0.011**	0.019***	0.020***	0.018***
	(0.004)	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)
Initial log GDP	-0.065	-0.009	0.132*	0.082	0.026	-0.043
	(0.063)	(0.071)	(0.071)	(0.121)	(0.120)	(0.071)
Real US Interest Rate	0.084*	0.079*	0.055*	0.059	0.059	0.063
	(0.047)	(0.041)	(0.033)	(0.059)	(0.048)	(0.041)
Trade Openness (de jure)			-0.130			
			(0.270)			
Trade Openness (de facto)			0.015***			
			(0.005)			
Manufactures (% Exports)				-0.006		
				(0.009)		
Export Diversification				-0.007		
				(0.009)		
Private Credit					0.003	
					(0.008)	
Education (All)						0.091
						(0.071)
Region FE	NO	YES	YES	YES	YES	YES
Exits	48	48	43	24	28	46
Spells	58	58	52	31	35	56
Years of Decline	348	348	316	236	198	327
Log-$\mathcal{L}$	-74.028	-69.324	-55.867	-31.403	-36.783	-65.675
Pseudo-R^2	0.157	0.211	0.302	0.296	0.278	0.219

Note(s): The standard errors are cluster-robust at the country level. *** p<0.01, ** p<0.05, * p<0.1.

Table 2 first presents our base specification and then addresses the issue of omitted variables by adding groups of variables which passed the selection step. All variables, except *de facto* trade openness and private credit, enter with the expected sign. The broad patterns are very interesting. Above all, the effect of fractionalization is extremely robust in all specifications and varies only within a narrow band. A one percentage point increase in fractionalization is estimated to prolong the decline phase by about 1-2% depending on the specification used. Further, the coefficient of executive constraints has a negative sign throughout and is often significant. Most specifications imply that a one point improvement in executive constraints leads to a 14-18% reduction in the duration of the decline phase. The coefficient weakens when we control for export variables and private credit, but this is simply the result of losing roughly half the sample. In both cases, we find no evidence suggesting that these additional covariates are important. Even if they were significant, modern diversified economies can be characterized as outcomes of institutional development (Acemoglu et al., 2003).

Column (1) is our preferred specification. This model captures most of the effects we are interested in and uses all available spells. The coefficients point in the expected directions and our effects of interest are highly significant. In column (2) we add region fixed effects to the base specification to capture a range of unobserved factors. Apart from a slight loss of efficiency, this induces no qualitative change and shows these results are robust to only using within-region variation. Column (3) shows that *de jure* trade openness has an insignificant negative effect on expected duration and *de facto* openness has a significant positive effect. The latter suggests that greater import and/or export reliance is not necessarily an advantage in a crisis and could be an indicator of vulnerability to contagion. Both the coefficients and standard errors of political institutions and fractionalization are unaffected. Column (4) illustrates that the effect of fractionalization is also robust to the inclusion of the share of manufactured exports and a Herfindahl index of export diversification. However, the effect of executive constraints decreases in absolute magnitude and becomes insignificant. A regression using the same sample without the added controls shows that this is owed to the diminished sample size (not reported). Turning to financial factors, column (5) reveals that including private credit reduces the coefficient of executive constraints but also decreases the sample size substantially. The effect of private credit is very weak statistically, so we have no reason to prefer this specification. Here too, the reduction in the coefficient on executive constraints is due to the smaller sample size. Column (6) illustrates that human capital (years of schooling) has no discernible effects on the expected duration. Most interestingly, the coefficient of initial GDP is not very stable. Sometimes it is negative, sometime positive, but it is usually insignificant. This suggests that, provided we take GDP at its first observed value to rule out the obvious reverse causality from crisis to output[6], there is no systematic income effect. Conditional on institutions and fractionalization, the duration of declines does not seem to depend on income levels. Taken together, the regressions in Table 2 show that the effect of fractionalization is very robust and the effect of political institutions only suffers when we control for export variables and private credit – measures that we can only include with considerable data loss.

Table E-3 in the Appendix assesses the effects of adding each control variable separately to our preferred specification and reveals some additional insights. Complementing the results of Desmet et al. (2012), we find no evidence of an effect of aggregate measures of fractionalization (ELF1) on the duration of declines. In other words, historical group divisions do not seem to be relevant once we condition on contemporary linguistic diversity. Several of the variables that passed the selection step have effects that are not robust in a multivariate setting. The coefficients of the share of manufactured exports, export diversification, and education point in the expected direction but are insignificant by a large margin. Otherwise, the same pattern as in Table 2 emerges. The coefficient of institutions only becomes insignificant when the sample size is diminished substantially. In fact, the effect size of executive constraints remains relatively stable in most specifications, even when significance suffers as the sample size decreases.

The estimated effects of political institutions are quite large in economic terms. It is most natural to evaluate their substantive implications by examining the range of predicted durations until the recovery starts. In the log-normal model, log mean and log median duration are both estimated by the exponentiated linear prediction, while mean and median durations can be obtained from the implied survival function. Based on the specification with region fixed effects in Table 2, a country with the lowest score on the executive constraints measure is expected to decline for about 10.9 years on average, while a country with the highest score is expected to decline for "only" about 4.7 years on average. The mean of executive constraints in the estimation sample is about 2.4, implying a duration of 8.9 years. Clearly our models capture a significant portion of the different crisis experiences of more and less developed economies, but the actually observed range is still much larger. While other (unobserved) factors could be at play, one option is that the effect of institutions on the duration depends on another variable.

[6]This is especially important in the duration context where we pool earlier and later GDP levels in the same regression. If there is a time trend in the durations, then its effect may be spuriously attributed to GDP.

We conjecture that the models presented so far are misspecified in the sense that they all lack an interaction effect between political institutions and fractionalization. The rationale for this hypothesis is simple. Given a political economy in which ethnic tension challenges the ability of political actors to take coordinated action, more cohesive institutions may help to overcome this vulnerability by internalizing these disputes and limiting the downside risks for the involved groups. Hence, countries with a high degree of ethnic fractionalization may require strong institutions just to compensate. Conversely, countries with a greater degree of ethnic homogeneity may make do with less developed institutions to achieve a similar degree of social coordination. This hypothesis is a less restrictive variant of the idea that there is a multiplicative effect between social conflict (broadly defined) and institutions in response to external shocks (Rodrik, 1999). As argued in the introduction, such an effect is commonly proposed in the literature (Alesina and Ferrara, 2005) but there is limited empirical evidence along these lines, especially when it comes to the more precise channels through which stronger institutions can "mute" the adverse effects of ethnic heterogeneity. We now examine if we can find broad evidence of such coordination issues when it comes to managing downturns.

Table 3 – Interaction effects of institutions and fractionalization

Variables	(1) $\ln \tilde{t}$	(2) $\ln \tilde{t}$	(3) $\ln \tilde{t}$	(4) $\ln \tilde{t}$	(5) $\ln \tilde{t}$	(6) $\ln \tilde{t}$
Executive Constraints ($\widetilde{INS_0}$)	-0.247***	-0.234***	-0.323***	-0.013	-0.310***	-0.294***
	(0.060)	(0.068)	(0.079)	(0.092)	(0.097)	(0.078)
Fractionalization ($\widetilde{ELF}15$)	0.015***	0.019***	0.013***	0.019***	0.024***	0.021***
	(0.004)	(0.004)	(0.003)	(0.007)	(0.004)	(0.005)
$\widetilde{INS_0} \times \widetilde{ELF}15$	-0.004***	-0.004***	-0.005***	0.000	-0.007***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Initial log GDP	-0.071	-0.023	0.146**	0.076	-0.131	-0.062
	(0.062)	(0.072)	(0.074)	(0.127)	(0.119)	(0.072)
Real US Interest Rate	0.094**	0.081*	0.056*	0.061	0.032	0.059
	(0.048)	(0.041)	(0.031)	(0.057)	(0.043)	(0.040)
Trade Openness (de jure)			-0.131			
			(0.236)			
Trade Openness (de facto)			0.020***			
			(0.006)			
Manufactures (% Exports)				-0.006		
				(0.009)		
Export Diversification				-0.007		
				(0.009)		
Private Credit					0.014**	
					(0.007)	
Education (All)						0.107
						(0.073)
Region FE	NO	YES	YES	YES	YES	YES
Exits	48	48	43	24	28	46
Spells	58	58	52	31	35	56
Years of Decline	348	348	316	236	198	327
Log-$\mathcal{L}$	-71.232	-66.878	-49.515	-31.393	-30.300	-62.551
Pseudo-R^2	0.189	0.239	0.381	0.296	0.405	0.256

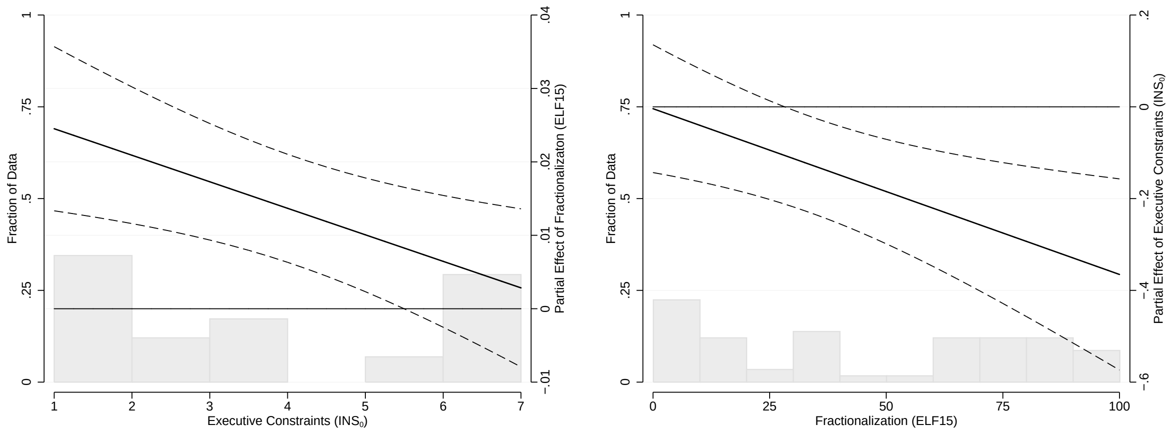
Note(s): The standard errors are cluster-robust at the country level. *** p<0.01, ** p<0.05, * p<0.1

Table 3 reports the corresponding results. In order to ease the interpretation, we subtract the sample average of the institutions and fractionalization variables from their observed values before estimating each model. We denote the demeaned variables by $\widetilde{INS_0}$ and $\widetilde{ELF}15$. This has the following effect. If either one of the two variables is at its mean, then the interaction term is zero and the only relevant coefficient is the non-interacted variant. As a result, the coefficient of the executive constraints variable directly measures the effect of institutions at the average level of fractionalization, and *vice versa*. For values other than the mean, the coefficient on the interaction term needs to be taken into account. Note that the statistical significance of the interaction term is not affected by this transformation.

We find convincing evidence of an interaction effect. In the same specifications where we find a robust effect of political institutions, we also see a significant interaction effect between executive constraints and ethno-linguistic fractionalization. In each case, the partial effect of one variable at the mean of the other is at least as significant as in the corresponding specification without an interaction effect. Since our earlier preferred specification is nested in column (1), testing the null that the interaction term is zero is equivalent to a test that this model fits the data better. A likelihood ratio test also prefers the interaction model and the pseudo-R^2 improves from 0.157 to 0.189. Several earlier results are strengthened. For example, we now find coherent results even in the small sample produced by the inclusion of private credit. The interaction term is negative and significant at the 5% level throughout all perturbations but column (4). In column (4) there is simply not enough data to estimate this effect, since we are left with only half of the observed exits. Here too, an auxiliary regression confirms that this is an issue of sample size and not the added controls. Overall, the interaction effect is strong, considering the underlying scale of the variables, and we still find little – if any – evidence of omitted variable bias.

Figure 4 – Average partial effects on the expected log duration until recovery

(a) Fractionalization ($ELF15$) **(b)** Executive Constraints (INS_0)



Note(s): The average partial effects are based on column (2) in Table 3 and are computed over the entire range of the variable on the horizontal axis while all other variables take on their respective realizations. The dashed lines are upper and lower 95% confidence limits.

Figure 4 show that the effects estimated in the interaction model are both economically and statistically significant across a large range of values. It plots the average partial effect with respect to one variable of the interaction term over representative values of the other, including a 95% confidence interval. The vertical axis on the right measures the average predicted semi-elasticity of the expected duration with respect to fractionalization or executive constraints. The expected log of time is still the dependent variable, so that the effects can be read just like the coefficients. For example, when the executive constraints index is at unity ('unlimited authority'), then a one percentage point increase in fractionalization leads to an expected increase in the duration until the recovery starts of about 2.5%. Both graphs are based on the interaction specification with region fixed effects (Column (2) in Table 3).

Three findings stand out. First, the effect of executive constraints clearly depends on linguistic heterogeneity (and *vice versa*), second, both partial effects are significant over most of the distribution, and, third, both partial effects consistently have the expected sign. In the background, Figure 4 also shows histograms of the sample. Executive constraints scores cover the entire range from 1 to 7, and ethno-linguistic fractionalization ranges from near-zero (0.07%) to near-total heterogeneity (96%). As expected, the predictions now cover a much wide range of the observed durations. At the average score of executive constraints, a country with the highest (lowest) degree of ethnic heterogeneity is expected to decline for about 15.4 years (2.2 years). In Figure 4, the effect of ethnic diversity ranges from being indistinguishable from zero, to the 2.5% mentioned above. Hence, it would be difficult to understand the effects of institutions without taking fractionalization into account. Stronger institutions also have the potential to overcome the adverse effects of high levels of ethnic fractionalization. At the 75^{th} percentile of ethnic heterogeneity ($ELF15 = 89.7$), a country with the highest (lowest) score of executive constraints is expected to decline for about 2.8 years (20.30 years). Or, as panel (b) shows, the partial effect of increasing executive constraints at perfect homogeneity is practically zero, while it peaks at about 37% at perfect heterogeneity.

One way to interpret the effects of fractionalization and institutions is through the lens of the delayed stabilizations literature (Alesina and Drazen, 1991). When (ethnic) groups engage in a 'war of attrition' over the burden of reform and are uncertain about how the reform will benefit all other groups (hence their willingness to bear the costs), then policy reform is delayed until the weakest group concedes. The expected time until stabilization occurs is expected to increase with the number of groups involved in the decision-making process and the veto points they possess, so that the adjustment speed depends on the political system (Spolaore, 2004). However, this interpretation does not explain the strong interaction effect between executive constraints and ethnic heterogeneity very well.

Bluhm and Thomsson (2015) propose a different mechanism and theory of how ethnic heterogeneity leads to delayed responses during crises. Groups facing a crisis have to decide on a policy response under uncertainty about post-crisis outcomes. When the executive is unconstrained, then some groups may have an incentive to delay cooperation as they fear boosting the strength of the independent executive and its power to expropriate them in the aftermath of a crisis should they become too weak. Conversely, if institutions are sufficiently strong, then the risk of expropriation practically disappears and only the uncertainty due to the crisis remains. There is still plenty of room for coordination failures to occur but the political friction that *always* induces non-cooperative behavior disappears. If groups can fortify their position through blocking agreement on different policies (e.g. a nationalization or taking conditional loans), then such a mechanism could generate the observed interaction. We show that the problem gets worse with increasing group diversity, but can be resolved by stronger constraints on the executive at all levels of heterogeneity.

Robustness

We now briefly illustrate that our main conclusions are unaffected by the choice of the baseline hazard, extending the sample and adding unobserved heterogeneity, the exclusion of influential regions, dropping censored spells, and different ways of accounting for recurrent spells. We also analyze the depth of the decline and the average rate of contraction to show that our variables of interest primarily work through crisis duration.

Table 4 – Robustness: functional form

VARIABLES	(1) Log-normal $\ln \tilde{t}$	(2) Log-logistic $\ln \tilde{t}$	(3) Exponential $\ln \tilde{t}$	(4) Weibull $\ln \tilde{t}$	(5) Gompertz $\ln \tilde{t}$	(6) Cox PH $\ln \tilde{t}$
	Coefficients		*Hazard Ratios* ($\mathbb{H}_0$: HR = 1)			
Executive Constraints (INS_0)	-0.144***	-0.150**	1.200***	1.240***	1.196***	1.195***
	(0.048)	(0.060)	(0.067)	(0.087)	(0.069)	(0.080)
Fractionalization ($ELF15$)	0.014***	0.014**	0.984***	0.982***	0.985***	0.986**
	(0.004)	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)
Initial log GDP	-0.065	-0.053	1.114	1.149	1.109	1.122
	(0.063)	(0.070)	(0.090)	(0.115)	(0.090)	(0.109)
Real US Interest Rate	0.084*	0.081*	0.948	0.928	0.949	0.948
	(0.047)	(0.049)	(0.059)	(0.063)	(0.059)	(0.066)
$\ln \sigma$ (Log-normal)	-0.061					
	(0.083)					
$\ln \gamma$ (Log-logistic)		-0.555***				
		(0.089)				
$\ln p$ (Weibull)				0.215***		
				(0.082)		
γ (Gompertz)					-0.008	
					(0.031)	
Exits	48	48	48	48	48	48
Spells	58	58	58	58	58	58
Years of Decline	348	348	348	348	348	348
Log-$\mathcal{L}$	-74.028	-75.849	-76.974	-75.396	-76.946	-147.582
AIC	160.056	163.698	163.948	162.793	165.892	303.165
Pseudo-R^2	0.157	0.147	0.203	0.208	0.154	0.083

Note(s): The standard errors are cluster-robust at the country level. *** p<0.01, ** p<0.05, * p<0.1.

Table 4 tackles the issues of functional form and model selection. Focusing on first-order effects, we report our preferred specification without the interaction in the first column and then show estimates based on five alternative forms of the hazard function. Column (2) uses a log-logistic hazard instead of the log-normal shape. The estimated shape parameter ($\ln \gamma$) is negative, implying that the hazard is first increasing then decreasing as in the log-normal model. This lends itself to the following interpretation. In the first few years of a decline, some countries are able to recover quickly. However, the longer the decline lasts, the more economic fundamentals deteriorate making it increasingly difficult to enter a recovery. Columns (3) to (6) have a different interpretation. We no longer report coefficients but instead *hazard ratios*, since these models are proportional hazards (PH) models by nature. They are interpreted as follows. A hazard ratio greater than one implies a higher instantaneous probability of exiting the decline. A hazard ratio smaller than one implies a lower instantaneous probability of exiting the decline. Column (3) is the exponential or constant hazard model. The results remain very similar (given the altered interpretation), but the log-likelihood decreases somewhat. Column (4) uses a Weibull parameterization which allows for monotonically increasing or decreasing hazard rates. This model also has a shape parameter (p) which allows testing for whether the rate increases, decreases or is constant.

The estimate suggests that the baseline hazard is increasing over time. In contrast, the Gompertz model in column (5) suggests a shape that is monotonically decreasing ($\gamma < 0$). Among these parametric models, the AIC is lowest for the log-normal distribution; that is, our preferred model fits the data best. In column (6), we specify a semi-parametric Cox model which does not restrict the shape of the baseline hazard. The Cox model suggests that the probability of exiting a spell first increases very briefly and then decreases and increases in turns. However, the imposed proportional hazard restriction comes at a great cost in terms of fit.

Table 5 – Robustness: sample, heterogeneity, dropping regions, and multiple failures

Variables	(1) Country RE $\ln \tilde{t}$	(2) Decade FE $\ln \tilde{t}$	(3) No Africa $\ln \tilde{t}$	(4) No Censored $\ln \tilde{t}$	(5) Single Spells $\ln \tilde{t}$	(6) PWP $\ln \tilde{t}$
Executive Constraints (INS_0)	-0.111**	-0.124**	-0.117**	-0.077*	-0.161***	1.235***
	(0.049)	(0.052)	(0.055)	(0.047)	(0.054)	(0.094)
Fractionalization ($ELF15$)	0.008**	0.014***	0.004	0.007*	0.012***	0.988**
	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)
Initial log GDP	-0.060	-0.051	-0.035	-0.016	-0.074	1.104
	(0.066)	(0.060)	(0.063)	(0.057)	(0.070)	(0.104)
Real US Interest Rate	0.097**	-0.055	0.093**	0.092**	0.082	0.943
	(0.039)	(0.071)	(0.044)	(0.039)	(0.059)	(0.069)
VCE	–	cluster	cluster	cluster	cluster	cluster
Frailties	shared	–	–	–	–	–
Strata	–	–	–	–	–	spell #
Exits	70	48	40	48	41	48
Spells	82	58	44	48	51	58
Years of Decline	466	348	170	197	314	348
Log-$\mathcal{L}$	-112.660	-70.835	-53.300	-57.399	-65.454	-127.891
Pseudo-R^2	0.065	0.194	0.092	0.094	0.154	0.089

Note(s): *** p<0.01, ** p<0.05, * p<0.1.

The previous section has already shown that the findings are robust to the inclusion of regional fixed effects. Column (1) in Table 5 goes two steps further: it extends the sample by running the break search algorithm with a significance level of 20% to detect more episodes, and it includes country-level effects. We now identify 82 spells in total, out of which 70 are completed. Each country now also has a so-called gamma distributed frailty; the duration analysis equivalent of random effects in linear models. As usual, they are assumed to be uncorrelated with any of the covariates (which is unlikely to hold in practice). Our results are robust to these two modifications; the coefficients of interest hardly change. Interestingly, there is only some evidence in favor of unobserved effects altogether. A Likelihood Ratio test for the presence of shared frailties fails to reject the null at 5% significance ($p = 0.111$).

Next we examine if trying to also account for time effect on top of duration dependence makes a difference. Column (2) shows that the inclusion of decade fixed effects creates only negligible differences. Note that we cannot include a full set of time dummies in duration regressions. To examine if results are driven by specific regions, we re-estimate our preferred model and remove the region with the longest spells. Column (3) drops all episodes in Africa and reveals an interesting additional finding. While the coefficient of fractionalization (ELF15) is very robust in the previous models, its size and significance is clearly driven by African observations. Without those, the coefficient keeps the same sign but shrinks substantially and becomes insignificant at conventional levels. The interaction model proposed earlier may thus be particularly relevant to understanding the effects of institutions and fractionalization in Africa. The coefficient of institutions remains large and significant. Since the African continent has the greatest ethno-linguistic heterogeneity of all regions, this result is hardly a surprise. Column (3) throws away a large part of the relevant variation.

Column (4) deletes the censored spells. Recall that censoring basically confounds two types of cases. On the one hand, we have episodes where there is a preliminary trough and the country is well on the way of recovery but pre-crisis GDP per capita has not been regained yet. On the other hand, there could be cases where the trough coincides with the end of the sample, so that we only observe perpetual decline. In our sample, only Togo fits into the latter category. Deleting both types of episodes means that our effects of interest weaken somewhat, both in magnitude and statistically, but still point in the same direction. This is also not particularly surprising; we have limited data and the whole purpose of running survival regressions is to be able to take censoring into account. Excluding them creates a selected sample of countries that managed to recover within a certain time span. Reassuringly, when we only exclude the cases that are qualitatively different (i.e. Togo), then the effects are very close to our preferred specification.

Until now, we assumed that multiple spells of the same type are interchangeable. The last two columns of Table 5 investigate if this relatively strong form of conditional independence is a reasonable assumption. Column (5) shows that our findings are robust to excluding all spells other than the first, which rules out any dependency across recurrent spells. The coefficient of executive constraints becomes even larger and the effect of fractionalization is virtually unchanged. Column (6) takes a different approach and specifies a conditional risk set model or stratified Cox model due to Prentice et al. (1981). The model accounts for ordering of the events but assumes that a subject cannot experience another event until the previous event has occurred. This is a natural assumption, as – by definition – a country cannot exit a second decline phase before having left the first. The results (hazard ratios) are qualitatively similar.

Last but not least, we construct a simple test to examine whether the effects of institutions and ethnic cleavages on the depth of a slump run solely through the duration process or if they also affect the rate of contraction. Since the depth of the decline is the product of the estimated duration and the average rate of contraction, we can define $\bar{g}_i \equiv \left(y_{i,\hat{t}_{min}} - y_{i0}\right)/(\hat{t}_{min} - t_0) \equiv \left(y_{i,\hat{t}_{min}} - y_{i0}\right)/\tilde{t}_D$ as the average rate of decline and $\tilde{t}_D \times \bar{g}_i \equiv y_{i,\hat{t}_{min}} - y_{i0}$ as the overall depth of the decline. Here y_{it} is still the log of GDP per capita in country i at time t, and $\bar{g}_i$ is by construction negative. We have already analyzed the duration of declines. Now we run OLS regressions explaining the rate of contraction ($\bar{g}_i$) and depth ($\tilde{t}_D \times \bar{g}_i$) to isolate the channel through which the previously estimated effects run. We scale both outcomes by one hundred for readability.

Column (1) in Table 6 shows that we find only weak evidence in favor of an effect of either executive constraints or fractionalization on the average rate of decline. Column (2) repeats this exercise for the specification with an interaction term. While some of the coefficients are significant, two out of three are estimated to be virtually zero. Hence, in substantive terms, the estimated effects on the average rate of decline are very small (e.g. around 0.4 p.p. in the case of a unit change in executive constraints) and explain very little of the variation (the adjusted R^2's are around 1-2%). Both variables that make up the interaction are mean-centered in all columns. The remaining columns of Table 6 examine the depth of the slump. Columns (3) and (4) illustrate that we now recover the previously estimated effects with similar significance levels and, naturally, with reversed signs since $\bar{g}_i$ is negative. These results are robust to the inclusion of region fixed effects

Table 6 – Robustness: Average rate of decline and total depth

VARIABLES	(1) $\bar{g}_i$	(2) $\bar{g}_i$	(3) $\tilde{t}_D \times \bar{g}_i$	(4) $\tilde{t}_D \times \bar{g}_i$	(5) $\tilde{t}_D \times \bar{g}_i$	(6) $\tilde{t}_D \times \bar{g}_i$
Executive Constraints ($\widetilde{INS_0}$)	0.341*	0.408**	3.141***	3.827***	2.830***	3.575***
	(0.170)	(0.172)	(0.795)	(0.749)	(1.052)	(1.056)
Fractionalization ($\widetilde{ELF}15$)	-0.012	-0.006	-0.261***	-0.200***	-0.251**	-0.214***
	(0.014)	(0.012)	(0.089)	(0.073)	(0.101)	(0.078)
$\widetilde{INS_0} \times \widetilde{ELF}15$		0.008**		0.086***		0.089***
		(0.004)		(0.020)		(0.024)
Initial log GDP	0.028	0.045	1.325	1.496	0.583	0.988
	(0.343)	(0.336)	(1.255)	(1.146)	(1.458)	(1.436)
Real US Interest Rate	-0.157	-0.204	0.042	-0.433	0.147	-0.329
	(0.157)	(0.157)	(0.852)	(0.821)	(0.885)	(0.846)
Region FE	NO	NO	NO	NO	YES	YES
Spells	58	58	58	58	58	58
Log-$\mathcal{L}$	-158.159	-157.363	-254.477	-251.303	-253.902	-250.935
Adjusted R^2	0.014	0.022	0.290	0.351	0.247	0.306

Note(s): Standard errors are cluster-robust at the country level. *** p<0.01, ** p<0.05, * p<0.1.

as shown in columns (5) and (6). The interaction models explain about 30–35% of the variation in depth, highlighting the relevance of the estimated effects. Together with the duration analysis and the analysis of the average rate of decline, this leads us to conclude that a) political institutions and ethnic heterogeneity have robust effects on the overall depth of slumps, and b) these effects run primarily through the duration of the decline phase and not the rate of contraction.

5 Concluding remarks

This paper makes several contributions to a burgeoning literature on structural breaks in growth performances and the political economy of crises. First, we show that a restricted structural change approach, as in Papell and Prodan (2014), works well as an inferential method for identifying slumps, big recessions or growth collapses in a large sample of countries. We find a substantial number of slumps of varying length in developing and developed countries alike. Severe downward volatility seems to be an ubiquitous phenomenon in the post-war period. We then seek explanations for what drives these slumps and determines their durations. To our best knowledge, we are the first to analyze the duration of the decline phase. We provide systematic evidence of weak political institutions before slumps hit and positive institutional change during and in the immediate aftermath of slumps. Our interpretation of this stylized fact is that, while institutions may cause growth, volatility can in turn contribute to endogenous institutional change. Severe economic crises raise the pressure for institutional reform in a very broad sense.

Our main empirical finding is that the duration of declines depends on the strength of political institutions and also particularly strongly on the level of ethnic diversity. An important qualification is that the effect of executive constraints is non-linear and depends on the level of ethnic fractionalization. Once this interaction is taken into account, these two variables alone (without regional dummies etc.) are able to explain why some declines are very brief and others last more than a decade. We also show that effects of political institutions and ethnic cleavages on the depth of declines run primarily through the *duration* until the recovery starts and not through the *pace* of decline. This highlights why we have to focus on the duration process to begin with and need to suggest a different line of explanations than earlier contributions.

It is well-established that both institutions and ethnicity matter for long-run development, but we still lack evidence on how precisely they affect contemporary economic growth. This paper shows that effective coordination and responses to slumps are hampered by high degrees of social tension as captured by ethno-linguistic fractionalization. Conversely, particularly strong political institutions can put in place coordination mechanisms that are able to contain or resolve these conflicts within the institutional framework. At the same time, our findings imply that in less ethnically fragmented societies political institutions are a less critical determinant of the length of declines. This interplay between institutions and group diversity is not well captured by current theories of policy reform and delay, which typically focus on information asymmetries or uncertainty about the benefits of reform.

While the previous literature has stressed the role of positive growth spurts, we show that slumps matter a lot and that the decline phase can last very long in some cases. Hence, a key function of stronger institutions is limiting downside risks, perhaps much more than the literature emphasizes so far. A comparison of the relative effects of slumps versus accelerations on long-run GDP levels would be an interesting extension of our findings, but more work should go into creative identification strategies to uncover the specific causal effects of ethnic diversity on crisis management in different political environments.

References

Acemoglu, D. and S. Johnson (2005). Unbundling institutions. *Journal of Political Economy 113*(5), 949–95.

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review 91*(5), 1369–1401.

Acemoglu, D., S. Johnson, J. A. Robinson, and Y. Thaicharoen (2003). Institutional causes, macroeconomic symptoms: volatility, crises and growth. *Journal of Monetary Economics 50*(1), 49–123.

Acemoglu, D. and J. A. Robinson (2006). *Economic Origins of Democracy and Dictatorship*. Cambridge University Press.

Aguiar, M. and G. Gopinath (2007). Emerging market business cycles: The cycle is the trend. *Journal of Political Economy 115*(1), 69–102.

Alesina, A. and A. Drazen (1991). Why are stabilizations delayed? *American Economic Review 81*(5), 1170–1188.

Alesina, A. and E. L. Ferrara (2005). Ethnic diversity and economic performance. *Journal of Economic Literature 43*(3), 762–800.

Ashraf, Q. and O. Galor (2013). The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development. *American Economic Review 103*(1), 1–46.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory 13*, 315–352.

Bai, J. (1999). Likelihood ratio tests for multiple structural changes. *Journal of Econometrics 91*(2), 299–323.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*, 47–78.

Barro, R. J. and J. W. Lee (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics 104*, 184–198.

Beck, T., A. Demirgüç-Kunt, and R. Levine (2010). Financial institutions and markets across countries and over time: The updated financial development and structure database. *World Bank Economic Review 24*(1), 77–92.

Ben-David, D. and D. H. Papell (1995). The great wars, the great crash, and steady state growth: Some new evidence about an old stylized fact. *Journal of Monetary Economics 36*(3), 453 – 475.

Berg, A., J. D. Ostry, and J. Zettelmeyer (2012). What makes growth sustained? *Journal of Development Economics 98*(2), 149–166.

Besley, T. and T. Persson (2011). The logic of political violence. *Quarterly Journal of Economics 126*(3), 1411–1445.

Bluhm, R. and K. Thomsson (2015). Ethnic divisions, political institutions and the duration of declines: A political economy theory of delayed recovery. UNU-MERIT Working Papers 2015–003, UNU-MERIT.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Cerra, V. and S. C. Saxena (2008). Growth dynamics: The myth of economic recovery. *American Economic Review 98*(1), 439–57.

Chinn, M. D. and H. Ito (2006). What matters for financial development? Capital controls, institutions, and interactions. *Journal of Development Economics 81*(1), 163–192.

Desmet, K., I. Ortuno-Ortín, and R. Wacziarg (2012). The political economy of linguistic cleavages. *Journal of Development Economics 97*(2), 322–338.

Diebold, F. X. and C. Chen (1996). Testing structural stability with endogenous breakpoint a size comparison of analytic and bootstrap procedures. *Journal of Econometrics 70*(1), 221–241.

Easterly, W., M. Kremer, L. Pritchett, and L. Summers (1993). Good policy or good luck? *Journal of Monetary Economics 32*(3), 459–483.

Eichengreen, B., A. K. Rose, and C. Wyplosz (1995). Exchange market mayhem: the antecedents and aftermath of speculative attacks. *Economic Policy 10*(21), 249–312.

Fernandez, R. and D. Rodrik (1991). Resistance to reform: Status quo bias in the presence of individual-specific uncertainty. *American Economic Review 81*(5), 1146–1155.

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002). Armed conflict 1946-2001: A new dataset. *Journal of Peace Research 39*(5), 615–637.

Goemans, H. E., K. S. Gleditsch, and G. Chiozza (2009). Introducing archigos: A dataset of political leaders. *Journal of Peace Research 46*(2), 269–283.

Gourinchas, P.-O. and M. Obstfeld (2012). Stories of the twentieth century for the twenty-first. *American Economic Journal: Macroeconomics 4*(1), 226–265.

Hansen, B. E. (2000). Testing for structural change in conditional models. *Journal of Econometrics 97*(1), 93–115.

Hausmann, R., J. Hwang, and D. Rodrik (2007). What you export matters. *Journal of Economic Growth 12*(1), 1–25.

Hausmann, R., L. Pritchett, and D. Rodrik (2005). Growth accelerations. *Journal of Economic Growth 10*(4), 303–329.

Jerzmanowski, M. (2006). Empirics of hills, plateaus, mountains and plains: A markov-switching approach to growth. *Journal of Development Economics 81*(2), 357–385.

Jones, B. F. and B. A. Olken (2008). The anatomy of start-stop growth. *Review of Economics and Statistics 90*(3), 582–587.

Jong-A-Pin, R. and J. De Haan (2011). Political regime change, economic liberalization and growth accelerations. *Public Choice 146*(1-2), 93–115.

Kalbfleisch, J. D. and R. L. Prentice (2002). *The statistical analysis of failure time data* (2nd ed.). New York: John Wiley.

Lancaster, T. (1990). *The econometric analysis of transition data.* Cambridge University Press.

Lane, P. R. and G. M. Milesi-Ferretti (2007). The external wealth of nations mark ii: Revised and extended estimates of foreign assets and liabilities, 1970-2004. *Journal of International Economics 73*(2), 223–250.

MacKinnon, J. (2009). Bootstrap hypothesis testing. In D. Belsley and E. Kontoghiorghes (Eds.), *Handbook of Computational Econometrics*, Chapter 6, pp. 183–213. John Wiley & Sons, Ltd.

Mobarak, A. (2005). Democracy, volatility, and economic development. *Review of Economics and Statistics 87*(2), 348–361.

Nelson, C. R. and C. R. Plosser (1982). Trends and random walks in macroeconmic time series: some evidence and implications. *Journal of Monetary Economics 10*(2), 139–162.

North, D. C., J. Wallis, and B. Weingast (2009). *Violence and Social Orders A Conceptual Framework for Interpreting Recorded Human History.* Cambridge University Press.

Papell, D. H. and R. Prodan (2014). Long-run time series tests of constant steady-state growth. *Economic Modelling 42*, 464 – 474.

Prentice, R. L., B. J. Williams, and A. V. Peterson (1981). On the regression analysis of multivariate failure time data. *Biometrika 68*(2), 373–379.

Pritchett, L. (2000). Understanding patterns of economic growth: searching for hills among plateaus, mountains, and plains. *World Bank Economic Review 14*(2), 221–250.

Prodan, R. (2008). Potential pitfalls in determining multiple structural changes with an application to purchasing power parity. *Journal of Business & Economic Statistics 26*(1), 50–65.

Ramey, G. and V. A. Ramey (1995). Cross-country evidence on the link between volatility and growth. *American Economic Review 85*(5), 1138–1151.

Rodrik, D. (1999). Where did all the growth go? External shocks, social conflict, and growth collapses. *Journal of Economic Growth 4*(4), 385–412.

Solt, F. (2009). Standardizing the World Income Inequality Database. *Social Science Quarterly 90*(2), 231–242. SWIID Version 3.0, July 2010.

Spolaore, E. (2004). Adjustments in different government systems. *Economics & Politics 16*(2), 117–146.

Wacziarg, R. and K. H. Welch (2008). Trade liberalization and growth: New evidence. *World Bank Economic Review 22*(2), 187–231.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). The MIT press.

Zivot, E. and D. W. K. Andrews (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics 10*(3), 251–270.

A Appendix: Estimation of Structural Breaks

A.1 Sequential procedure for testing and dating breaks

The procedure described here is a modification of Bai's (1997) sequential likelihood ratio tests for structural change – see also the extensions in Bai and Perron (1998) and in Bai (1999). We make an important simplifying assumption, namely, that all output series are regime-wise trend-stationary. Verifying this assumption is beyond the scope of this paper, as testing for unit roots in the presence of structural breaks (with sufficient power and size) is still contested territory and our output series have only a moderate time dimension ($T < 60$ years). We implement the sequential procedure in six steps.

1. Determine the optimal $AR(p)$ trend model using the Bayesian information criterion to adjust for serial correlation up to a maximum lag count (p_{max}). We set $p_{max} = 4$.

2. Specify the partial structural change model:

$$y_t = \alpha + \beta t + \gamma_0 \mathbf{1}(t > tb_1) + \gamma_1 (t - tb_1)\mathbf{1}(t > tb_1) + \gamma_2 (t - tb_2)\mathbf{1}(t > tb_2) + \sum_{i=1}^{p} \delta_i y_{t-i} + \epsilon_t$$

   where y_t is the log of GDP per capita in year t, tb_i are the possible break dates, $\mathbf{1}(\cdot)$ is an indicator function, and p is the lag order as determined by the optimal $AR(p)$ model. We require that $tb_2 \geq tb_1 + h$ for $h = 4$. In other words, the period between two successive breaks making up the same episode is at minimum 4 years.

3. Define trimming parameter τ, where typically $\tau \in [0.05, 0.25]$. The breaks are in the ranges $tb_1 \in [\tau T, (1-\tau)T - h]$ and $tb_2 \in [\tau T + h, (1-\tau)T]$. We set $\tau = 0.05$. Let Λ_τ denote the set of all possible episodes $[tb_1, tb_2] \subset [\tau T, (1-\tau)T]$.[7]

4. Compute the sup-W test statistic of the null of no break versus at least one break ($\mathbb{H}_0 : \gamma_0 = \gamma_1 = \gamma_2 = 0$). The supremum is taken over all episodes in Λ_τ with a positive estimate of β and a non-positive estimate of γ_0:

$$\sup_{[tb_1, tb_2] \in \Lambda_\tau} W(tb_1, tb_2) = \sup_{[tb_1, tb_2] \in \Lambda_\tau} \left(\frac{T-K}{3}\right) \frac{SSR^r - SSR^u}{SSR^u}$$

   where K is the number of parameters, SSR^r denotes the sum of squared residuals from a regression imposing $\mathbb{H}_0$, and SSR^u the sum of squared residuals from a regression imposing only $\beta > 0$ and $\gamma_0 \leq 0$.

5. The critical value and empirical p-value of sup-W statistic is bootstrapped.[8]

6. If the sup-W statistic is significant at the desired level, the remaining sample is split into two new sub-samples from the beginning to the first break and from the second break to the end, then the procedure restarts at (4) using the estimated AR-order from before. If the bootstrapped sup-W^* test fails to reject in each sub-sample, or the sub-samples are too small ($T \leq 20$), then the procedure stops.

[7] For simplicity of exposition, we suppress an additional index running over the sub-samples (defined in Step 6). T refers to the number of observations of the currently active sample. The notation neglects the discontinuity of actual observation times.

[8] In finite samples comparable asymptotic tests often have poor size and power (see Prodan, 2008).

A.2 Bootstrapping the sup-Wald statistic

There have been several suggestions on how to best bootstrap structural change tests in particular or other popular time series tests in general. For example, Hansen (2000) suggests employing a fixed-design bootstrap allowing for non-stationarity, lagged dependent variables and conditional heteroskedasticity. MacKinnon (2009), on the contrary, shows that the recursive bootstrap of Diebold and Chen (1996) gives results superior to most other bootstrap types (fixed-parameter, sieve, pairs, block, double block) as well as the asymptotic test in a simple application of an AR(1) model with an endogenous break. Papell and Prodan (2014) also favor a recursive bootstrap though they do not compare it to other methods. We use a recursive bootstrap similar to Diebold and Chen (1996). Comparing methods systematically is beyond the scope of this paper.[9] In line with usual notation, we denote all bootstrap quantities with the superscript '$*$'. The bootstrap procedure is as follows.

1. Specify the optimal break model under the $\mathbb{H}_0$ of no structural breaks in the specified sample using the BIC as before and obtain the residuals:

$$\hat{e}_t = y_t - \hat{\alpha} - \hat{\beta}t - \sum_{i=1}^{p} \hat{\delta}_i y_{t-i}$$

2. Draw new residuals: $\hat{e}_t^* = u_t$, with $u_t \sim$ i.i.d. $\mathcal{N}(0, \hat{\sigma}_{\hat{e}}^2)$

3. Construct a bootstrap sample of equal size as the original sample:

$$y_t^* = \hat{\alpha} + \hat{\beta}t + \sum_{i=1}^{p} \hat{\delta}_i y_{t-i}^* + \hat{e}_t^*, \qquad \forall t = 1+p, \ldots, T$$

   where y_{t-i}^* is the observed y_{t-i} only in the case of a *fixed-design* bootstrap, otherwise y_t^* must be constructed *recursively* (conditional on p observed initial values).

4. Rerun the break search algorithm on the bootstrap series $\{y_t^*\}$, including determination of the optimal AR(p) model, and compute bootstrapped test statistics $\sup_{[tb_1^*, tb_2^*] \in \Lambda_\tau} W_j^*$, where j indexes the current bootstrap iteration.

5. Repeat from Step (2) until $j = B$, where B is the total number of bootstrap replications. We set $B = 1000$.

6. The bootstrap p-value ($\hat{p}^*$) is obtained by counting the proportion of the estimated bootstrap test statistics that are greater than the originally calculated test statistic.

$$\hat{p}^* = \frac{1}{B} \sum_{j=1}^{B} \mathbf{1} \left(\sup_{[tb_1^*, tb_2^*] \in \Lambda_\tau} W_j^* > \sup_{[tb_1, tb_2] \in \Lambda_\tau} W(tb_1, tb_2) \right)$$

   The critical value is the $(1-\alpha^s)B^{th}$ largest bootstrapped sup-W^* statistic, where α^s is the desired significance level (10% throughout the text, unless otherwise noted).

[9] We use a parametric recursive bootstrap, but informally compared the results to other techniques. Hansen's fixed-design bootstrap generates (too) many questionable slumps and the Wild bootstrap rejects (too) often. Residual and parametric bootstraps give similar results.

B Appendix: List of Episodes

Table B-1 – Global Parameters

Data:	PWT	Max AR (p_{max}):	4
Sample start:	1950	Bootstrap replications:	1000
Sample end:	2008	Bootstrap errors:	parametric
Trimming (τ):	0.05	Bootstrap type:	recursive
Min. tb_i distance (h):	4	Bootstrap significance (α^s):	0.1

Table B-2 – Estimated and Filtered Breaks with Troughs: 58 Episodes*

Code	T_0	$\widehat{tb}_1$	$\widehat{t}_{min}$	$\widehat{tb}_2$	T	Sup-W	Critical W	p-value	Drop (%)	Duration	c
ALB	1970	1990	1991	2002	2008	18.5	13.6	0.007	-15.32	1	0
ARE	1986	1990	1999	2002	2008	29.1	14.5	0.003	-10.90	9	0
AUS	1950	1954	1957	1966	2008	8.3	8.7	0.064	-0.72	3	0
AUS	1967	1989	1991	1998	2008	10.1	10.7	0.059	-2.29	2	0
BDI	1960	1971	1972	1988	2008	9.9	11.3	0.089	-3.23	1	0
BEL	1950	1957	1958	1973	2008	12.8	12.1	0.029	-2.24	1	0
BGR	1970	1988	1997	1997	2008	16.3	12.8	0.010	-23.79	9	0
BHR	1970	1980	1987	1986	2008	14.4	11.0	0.010	-44.12	7	1
BRA	1950	1980	1983	2003	2008	12.5	12.3	0.043	-14.60	3	0
CAF	1960	1978	2005	2005	2008	8.3	8.7	0.060	-46.38	27	1
CHE	1950	1974	1975	1978	2008	10.7	10.6	0.047	-7.87	1	0
CHL	1951	1953	1954	1972	1973	12.0	8.5	0.017	-9.06	1	0
CHL	1951	1974	1975	1979	1980	13.3	10.8	0.021	-16.50	1	0
CHL	1951	1981	1983	1995	2008	12.6	11.4	0.025	-21.22	2	0
CHN	1952	1960	1962	1977	2008	13.9	12.9	0.029	-23.71	2	0
CMR	1960	1986	1995	1990	2008	12.0	12.3	0.055	-40.46	9	1
COG	1960	1974	1977	1982	2008	11.9	12.5	0.069	-21.35	3	0
CRI	1950	1955	1956	1963	1979	11.4	11.3	0.048	-4.39	1	0
CRI	1950	1980	1982	2002	2008	17.2	10.6	0.002	-17.47	2	0
CUB	1970	1988	1993	1995	2008	11.4	12.5	0.072	-34.70	5	0
CYP	1950	1973	1975	1977	2008	15.5	9.7	0.001	-31.40	2	0
CYP	1978	1990	1991	1995	2008	11.6	14.6	0.098	-10.19	1	0
DNK	1950	1954	1955	1965	2008	12.9	11.7	0.022	-1.56	1	0
DZA	1960	1984	1994	1996	2008	10.9	8.2	0.013	-14.09	10	0
ETH	1950	1972	1992	1993	2008	11.5	10.2	0.020	-30.68	20	0
FIN	1950	1989	1993	2006	2008	10.6	10.8	0.057	-16.34	4	0
GAB	1960	1976	1987	1997	2008	10.6	11.2	0.062	-50.56	11	1
GMB	1960	1982	1998	2002	2008	16.4	11.2	0.006	-25.33	16	0
GRC	1951	1973	1974	1994	2008	17.9	11.6	0.003	-6.92	1	0
GTM	1950	1980	1988	1984	2008	15.1	12.3	0.015	-19.14	8	0
HUN	1970	1990	1992	2004	2008	15.6	13.5	0.018	-10.56	2	0
IDN	1960	1997	1999	2001	2008	13.5	10.6	0.013	-17.49	2	0
IRN	1955	1976	1981	1980	2008	15.9	11.6	0.004	-56.78	5	1
IRQ	1970	1990	2003	1994	2008	9.1	8.9	0.046	-66.43	13	1
JPN	1950	1973	1974	1990	2008	13.5	13.4	0.050	-2.85	1	0
MEX	1950	1981	1988	1995	2008	11.9	11.0	0.038	-17.03	7	0
MNG	1970	1990	1993	2003	2008	46.5	11.7	0.000	-41.81	3	0
MOZ	1960	1981	1986	1995	2008	12.6	12.0	0.037	-24.99	5	0
MYS	1955	1984	1986	1993	2008	9.1	10.5	0.093	-7.47	2	0
NPL	1960	1979	1980	2000	2008	10.6	8.9	0.025	-5.33	1	0
NZL	1950	1974	1978	1992	2008	9.9	10.5	0.070	-9.03	4	0

Continued on next page

Table B-2 – *Continued from previous page*

Code	T_0	$\widehat{tb}_1$	$\widehat{t}_{min}$	$\widehat{tb}_2$	T	Sup-W	Critical W	p-value	Drop (%)	Duration	c
OMN	1970	1979	1980	1985	2008	12.4	9.0	0.007	-21.61	1	0
PER	1950	1958	1959	1966	1976	11.9	9.3	0.022	-6.91	1	0
PER	1950	1977	1992	1992	2008	11.0	10.3	0.037	-29.30	15	0
PHL	1950	1983	1985	2003	2008	12.8	10.2	0.007	-16.78	2	0
POL	1970	1979	1982	1993	2008	13.8	12.1	0.027	-22.55	3	0
PRY	1980	1989	2002	2002	2008	8.8	8.8	0.049	-14.24	13	1
RWA	1960	1993	1994	1997	2008	18.0	7.9	0.001	-45.38	1	0
SAU	1986	1992	1999	2002	2008	14.6	13.3	0.039	-18.75	7	0
SLE	1961	1995	1999	2006	2008	14.2	11.1	0.011	-41.65	4	1
SLV	1950	1978	1983	1987	2008	18.2	10.2	0.002	-25.82	5	0
TGO	1960	1979	2008	1989	2008	9.6	10.1	0.065	-53.60	29	1
THA	1950	1996	1998	2003	2008	10.7	7.8	0.003	-14.17	2	0
TTO	1950	1961	1963	1969	1981	16.8	14.9	0.020	-0.78	2	0
TTO	1950	1982	1993	2006	2008	12.4	12.6	0.054	-28.96	11	0
UGA	1950	1977	1986	1987	2008	11.6	10.5	0.029	-30.27	9	0
USA	1950	1957	1958	1966	2008	8.7	9.3	0.075	-2.51	1	0
ZMB	1955	1968	2001	2000	2008	15.0	10.9	0.007	-68.99	33	1

* Out of a total of 70 episodes identified by the sequential algorithm, 12 are invalid slumps. The invalid episodes are [country code (spell number)]: AUT (1), AUT (2), CHN (1), FIN (1), HKG (1), IRN (1), MRT (1), PRY (1), TZA (1).

C Appendix: Data Sources and Summary Statistics

Table C-1 – Summary Statistics: break date to trough

Variable	Mean	Std. Dev.	$N \times T$	Source
Institutions, Politics & Conflict				
Polity Score	-1.88	7.00	347	Polity IV
Democracy	2.75	3.61	331	Polity IV
Autocracy	4.67	3.75	331	Polity IV
Executive Recruitment	4.92	2.27	331	Polity IV
Executive Constraints	3.19	2.29	331	Polity IV
Political Competition	4.12	3.38	331	Polity IV
Regime Duration	18.14	22.70	347	Polity IV
Corruption (ICRG)	2.63	1.10	193	ICRG
Fractionalization (ELF1)	18.36	18.69	348	Desmet et al. (2012)
Fractionalization (ELF15)	63.68	30.71	348	Desmet et al. (2012)
Inequality (Gini)	45.83	11.65	192	Solt (2009)
Leader Exit	0.39	0.49	344	Goemans et al. (2009)
War/Conflict (major)	0.12	0.33	348	Gleditsch et al. (2002)
War/Conflict (any)	0.24	0.43	348	Gleditsch et al. (2002)
Macro I: Prices, Trade & Exports				
Inflation ($\ln(1+\delta)$)	22.89	43.97	292	WDI/IFS
RER Undervalue	0.07	0.54	348	PWT 7.0
Current Account Balance (% of GDP)	-3.98	6.70	254	WDI
Δ Terms of Trade	-4.11	17.72	224	WDI/IFS
Manufactures (% of Exports)	22.65	24.27	264	WITS/ COMTRADE
Trade Openness (de facto)	67.85	37.43	348	PWT 7.0
Trade Openness (de jure)	0.23	0.42	306	Wacziarg and Welch (2008)
Export Sophisticaton	8.43	0.42	234	Hausmann et al. (2007)
Export Diversification	65.91	24.58	264	WITS/ COMTRADE
Macro II: Finance				
Capital Account Openness	-0.49	1.28	304	Chinn and Ito (2006)
Financial Integration	115.30	88.18	309	Lane and Milesi-Ferretti (2007)
Financial Depth	32.35	18.68	245	Beck et al. (2010)
Financial Development	68.40	22.18	271	Beck et al. (2010)
Private Credit (% of GDP)	26.25	23.53	248	Beck et al. (2010)
FDI Liabilities (% of GDP)	15.11	15.66	309	Lane and Milesi-Ferretti (2007)
External Debt Liabilities (% of GDP)	65.22	59.18	309	Lane and Milesi-Ferretti (2007)
External Leverage[a]	165.29	327.09	307	Lane and Milesi-Ferretti (2007)
Other Determinants				
Initial log GDP	15.74	1.66	348	PWT 7.0
Real US Interest Rate[c]	1.90	2.44	348	FRED
Infant Mortality[d]	73.37	40.23	348	World Population Prospects
Life Expectancy[d]	58.63	10.55	348	World Population Prospects
Telephones (per 100 people)	5.24	9.78	312	WDI
Education (primary)	3.14	1.71	327	Barro and Lee (2013)
Education (secondary)	1.12	0.83	327	Barro and Lee (2013)
Education (all)	4.44	2.47	327	Barro and Lee (2013)

[a] Following Gourinchas and Obstfeld (2012), external leverage is $l_i = (\tau + A_i/Y_i)(\tau + NA_i/Y_i + E_{ij}/Y_i)^{-1}$, where τ is the market value of assets to output (set to 3) and j is the rest of the world, A_i/Y_i is assets over GDP, NA_i/Y_i is net foreign assets over GDP and E_{ij}/Y_i equity over GDP. The ratio is always > 0 if $NA_i > -300$, this condition is not satisfied in very few cases; we set these missing.

[b] Initial refers to the first observed GDP value in the Penn World Tables.

[c] Deflated three months treasury bill rate.

[d] Converted into annual data by interpolation. If the average is for the years 1950-55, we assume it is reached in the 1952 and linearly interpolate to the middle of the next group (1957), and so on. The data is from the 2010 edition of the Word Population Prospects (medium-fertility variant).

D Appendix: Duration Method

Log-normal Accelerated Failure Time (AFT) models

Given the model $\ln(\tilde{t}) = \beta_0 + \mathbf{x}'\boldsymbol{\beta} + \epsilon$, log-normality implies the following relationships. Setting all covariates zero, the expected survival time is $\mathrm{E}[\ln \tilde{t}|\mathbf{x} = \mathbf{0}] = \beta_0$. Hence, the baseline survival and hazard functions are

$$S_0(\tilde{t}) = 1 - \Phi\left((\ln \tilde{t} - \beta_0)\sigma^{-1}\right) \quad \text{and} \quad \lambda_0(\tilde{t}) = \frac{\phi\left((\ln \tilde{t} - \beta_0)\sigma^{-1}\right)}{\left(1 - \Phi\left((\ln \tilde{t} - \beta_0)\sigma^{-1}\right)\right)\sigma\tilde{t}}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf, respectively.

Including (time-invariant) covariates is equivalent to scaling the baseline survival functions. The conditional survival curve is defined as $S(\tilde{t}|\mathbf{x}) = S_0(\tilde{t})\left(\exp(-\mathbf{x}'\boldsymbol{\beta})\tilde{t}\right)$. This implies $S(\tilde{t}|\mathbf{x}) = 1 - \Phi\left((\ln \tilde{t} - (\beta_0 + \mathbf{x}'\boldsymbol{\beta}))\sigma^{-1}\right)$; that is, the intercept can be absorbed into $\boldsymbol{\beta}$. The density and cumulative probability functions are defined implicitly.[10]

Time-varying covariates introduce two complications. First, the hazard rate at each unit of analysis time $\tilde{t}$ is not independent from previous realizations of the time-varying covariates. Second, the covariates must be *strictly exogenous*, as otherwise feedback may occur from the duration to future realizations of the covariates. Following Lancaster (1990) and Kalbfleisch and Prentice (2002) these issues can be formalized as follows. For time-varying covariates $\mathbf{x}(\tilde{t})$, let $\mathbf{x}^H(\tilde{t})$ denote the covariate path up until time $\tilde{t}$, so that $\mathbf{x}^H(\tilde{t}) \equiv \{\mathbf{x}(u), 0 \leq u \leq \tilde{t}\}$ for all $\tilde{t} \geq 0$, then the conditional hazard function is:

$$\lambda(\tilde{t}|\mathbf{x}^H) = \lim_{d\tilde{t}\to 0} \frac{\Pr(\tilde{t} \leq \tilde{T} < \tilde{t} + d\tilde{t} \mid \tilde{T} \geq \tilde{t}, \mathbf{x}^H(\tilde{t} + d\tilde{t}))}{d\tilde{t}}$$

Lancaster (1990, pp. 26–30) and Kalbfleisch and Prentice (2002, p. 196) define strict exogeneity as $\Pr(\mathbf{x}^H(\tilde{t}) \mid \mathbf{x}^H(u), \tilde{T} \geq u) = \Pr(\mathbf{x}^H(\tilde{t}) \mid \mathbf{x}^H(u), \tilde{T} = u)$ for all $0 < u \leq \tilde{t}$. The condition states that the future path of the time-varying covariate is not affected by the event occurring at present.

We can now derive the partial likelihood.[11] Suppose we know the event occurs at $\tilde{t}_i$, the likelihood contribution of an observation i at time $j = \tilde{t}_i$ then is $\mathcal{L}_i = S(j)\lambda(j)$. The likelihood contribution of an observation that has not failed at time j, so that $j < \tilde{t}_i$, is just the probability of survival until j: $\mathcal{L}_i = S(j)$. Hence, right-censoring is essentially nothing else than an observation at analysis time j that is still in the sample but has not yet failed and thus extends easily to (exogenous) time-varying covariates.

Using the notation for grouped data from Wooldridge (2010, p. 1016), the log-likelihood of the log-normal model with time-varying covariates can be expressed as:

$$\ln \mathcal{L}(\boldsymbol{\beta}, \sigma) = \sum_{i=1}^{N} \left[\sum_{j=1}^{\tilde{t}_i - 1} \ln \alpha_j(\mathbf{x}'_{ij}\boldsymbol{\beta}, \sigma) + (1 - c_i)\ln\left(1 - \alpha_{\tilde{t}_i}(\mathbf{x}'_{i\tilde{t}_i}\boldsymbol{\beta}, \sigma)\right)\right]$$

where $\alpha_j(\cdot) = \exp[-\int_{\alpha_j - 1}^{\alpha_j} \lambda(s, \cdot)ds]$ measures survival over the given interval and c_i indicates if observation i is censored. The inner sum (first term) is the probability of survival until $\tilde{t}_i - 1$ and the second term is the conditional probability of failure at $\tilde{t}_i$.

[10] It follows that an expression for the hazard function conditional on the covariates is $\lambda(\tilde{t}|\mathbf{x}) = \lambda_0\left(\tilde{t}\exp(-\mathbf{x}'\boldsymbol{\beta})\right)\exp(-\mathbf{x}'\boldsymbol{\beta})$; these hazards are not proportional.

[11] This does not apply to frailty models where the likelihoods are more involved.

E Appendix: Variable Selection and Robustness

Table E-1 – Base Models

	Coefficient	SE	p-value	Exits	Spells	Years	$\log \mathcal{L}$
Constant Only	1.346	0.180	0.00	48	58	348	-87.86
Initial log GDP	-0.179	0.089	0.05	48	58	348	-85.83
Real US Interest Rate	0.096	0.047	0.04	48	58	348	-86.55

Note(s): All models include a constant. The standard errors are cluster-robust at the country level.

Table E-2 – Variable Selection

	Coefficient	SE	p-value	Exits	Spells	Years	$\log \mathcal{L}$
Inflation ($\ln(1+\delta)$)	-0.002	0.004	0.67	38	45	234	-62.88
RER Underval	-0.139	0.314	0.66	48	58	348	-84.33
Trade Openness (de jure)	-0.884	0.286	0.00	43	52	316	-73.33
Trade Openness (de facto)	0.009	0.005	0.05	48	58	348	-81.83
Current Account Balance	0.004	0.027	0.89	27	34	222	-47.14
Manufactures (% Exports)	-0.016	0.007	0.03	24	31	236	-41.75
Δ Terms of Trade	-0.008	0.015	0.58	24	27	164	-33.60
Export Diversification	-0.015	0.009	0.08	24	31	236	-41.75
Export Sophistication	-0.649	0.503	0.20	28	34	241	-48.41
Capital Account Openness	0.014	0.119	0.91	32	41	275	-58.86
Financial Integration	-0.001	0.003	0.80	35	43	271	-60.36
Financial Depth	-0.007	0.006	0.27	26	33	195	-44.23
Financial Development	0.001	0.008	0.88	31	39	266	-57.08
External Debt Liabilities	-0.002	0.007	0.73	35	43	271	-60.32
External Leverage	-0.002	0.013	0.87	35	43	271	-60.38
FDI Liabilities	-0.015	0.017	0.36	35	43	271	-60.04
Private Credit	-0.009	0.005	0.05	28	35	198	-46.96
Polity IV Score	-0.057	0.016	0.00	48	58	348	-79.55
Democracy Score	-0.099	0.029	0.00	48	58	348	-80.08
Autocracy Score	0.122	0.035	0.00	48	58	348	-79.36
Executive Recruitment	-0.166	0.054	0.00	48	58	348	-80.57
Executive Constraints (INS_0)	-0.184	0.056	0.00	48	58	348	-79.34
Political Competition	-0.106	0.034	0.00	48	58	348	-80.52
Regime Durability	0.002	0.005	0.63	47	57	346	-82.91
Corruption (ICRG)	-0.456	0.150	0.00	14	18	98	-19.08
Fractionalization ($ELF1$)	0.014	0.008	0.10	48	58	348	-83.05
Fractionalization ($ELF15$)	0.016	0.004	0.00	48	58	348	-77.65
Inequality (Gini)	0.031	0.022	0.16	22	27	137	-34.88
Leader Exit	0.355	0.332	0.29	47	57	346	-82.31
War/Conflict (major)	0.265	0.847	0.75	48	58	348	-84.35
War/Conflict (any)	0.436	0.505	0.39	48	58	348	-83.92
Infant Mortality	0.005	0.004	0.18	48	58	348	-83.56
Life Expectancy	-0.019	0.018	0.27	48	58	348	-83.57
Education (Primary)	-0.192	0.088	0.03	46	56	327	-78.10
Education (Secondary)	-0.218	0.127	0.09	46	56	327	-79.73
Education (All)	-0.121	0.056	0.03	46	56	327	-78.46
Telephones per capita	-0.015	0.012	0.21	30	38	257	-52.29

Note(s): All models also include initial GDP, the real US interest rate, and a constant. The standard errors are cluster-robust at the country level.

Table E-3 – Preferred Specification: Variable-by-Variable Models

Variables	(1) $\ln \tilde{t}$	(2) $\ln \tilde{t}$	(3) $\ln \tilde{t}$	(4) $\ln \tilde{t}$	(5) $\ln \tilde{t}$	(6) $\ln \tilde{t}$	(7) $\ln \tilde{t}$
Executive Constraints (INS_0)	-0.107*	-0.168***	-0.141	-0.171*	-0.036	-0.148***	-0.188***
	(0.054)	(0.051)	(0.091)	(0.092)	(0.070)	(0.048)	(0.064)
Fractionalization ($ELF15$)	0.014***	0.011**	0.007	0.005	0.015***	0.013**	0.015***
	(0.005)	(0.004)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Initial log GDP	-0.056	-0.032	-0.122	-0.147	-0.029	-0.060	-0.101
	(0.074)	(0.059)	(0.108)	(0.109)	(0.099)	(0.064)	(0.073)
Real US Interest Rate	0.059	0.083*	0.133	0.145*	0.096*	0.085*	0.070
	(0.046)	(0.046)	(0.086)	(0.086)	(0.051)	(0.047)	(0.047)
Trade Openness (de jure)	-0.522*						
	(0.298)						
Trade Openness (de facto)		0.008**					
		(0.004)					
Manufactures (% Exports)			-0.009				
			(0.009)				
Export Diversification				-0.007			
				(0.010)			
Private Credit					-0.009*		
					(0.005)		
Fractionalization ($ELF1$)						0.004	
						(0.008)	
Education (All)							0.061
							(0.069)
Exits	43	48	24	24	28	48	46
Spells	52	58	31	31	35	58	56
Years of Decline	316	348	236	236	198	348	327
Log-$\mathcal{L}$	-65.843	-72.019	-39.044	-39.145	-41.996	-73.914	-69.630
Pseudo-R^2	0.178	0.180	0.125	0.122	0.176	0.159	0.172

Note(s): The standard errors are cluster-robust at the country level. *** p<0.01, ** p<0.05, * p<0.1.