Do Weak Institutions Prolong Crises?

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

Richard Bluhm* Denis de Crombrugghe† Adam Szirmai‡

April 2018

Revised Manuscript

Abstract

Economic slumps can be characterized 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 up to 86 episodes exhibiting such a pattern using a restricted structural change approach. We analyze the behavior of several covariates around the downbreak and then focus on the characteristics of the decline phase. Our results show (*i*) that weak political institutions precipitate crises, and (*ii*) that the length and depth of economic slumps is robustly correlated with constraints on the executive and ethnic heterogeneity. Moreover, we document a robust interaction between these two variables, suggesting that strong political institutions are particularly important for stabilizing growth in heterogeneous societies.

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

JEL Classification: O43, O11, C41, F43

* *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

Modern economic growth since the 1950s has been far from steady. Every growth "miracle" is easily matched by a spectacular growth 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 or decline in the 1970s and 1980s. Moreover, there is a long list of relatively short-lived advanced economy crises. Yet many economic slumps in developing countries over this period were considerably deeper and longer than even the 'Great Depression' of the 1930s. What can we learn from such abrupt changes in growth? Why do some countries deal better with negative shocks than others?

The instability of growth is of great concern in economics because it has been suggested to affect output in the short run (e.g. Ramey and Ramey, 1995), have serious welfare implications (Pritchett et al., 2016), and even play an important role in sustaining long-run growth (e.g. Broadberry and Wallis, 2017). A growing literature on trend breaks has confirmed that growth is often not steady but instead characterized by switching among growth regimes (e.g. Ben-David and Papell, 1995, Jerzmanowski, 2006, Jones and Olken, 2008, Kerekes, 2012, Papell and Prodan, 2014). This perspective offers new stylized facts. For example, positive growth is relatively easy to ignite (Hausmann et al., 2005) but much harder to sustain (Berg et al., 2012). However, the implications of negative regime switches are only beginning to be explored. Long-lasting slumps can nullify decades of positive growth, with no guarantee that lost potential output is ever fully recouped (Cerra and Saxena, 2008), and they seem to be driven by different factors than growth accelerations (Jones and Olken, 2008, Hausmann et al., 2008).

A potential explanation for why some declines last so much longer than others is that their duration is driven by the prevailing structure of institutions. Institutions create specific political and economic incentives, solve or worsen coordination failures and define the set of feasible policies. Seminal contributions to this literature link stronger institutions with 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-run growth, just as there is an ongoing discussion about the relative importance of first improving political institutions versus prioritizing macroeconomic policies (Acemoglu et al., 2003, Fatás and Mihov, 2013).

This paper focuses on three aspects in particular. First, 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 (as in Papell and Prodan, 2012, 2014). A key difference to earlier approaches is that this superimposes the desired break structure *ex ante*, rather than classifying episodes *ex post*. We demonstrate the effectiveness of this approach by identifying many historical episodes following such a pattern in the post-1950 world. Second, we conduct an event analysis which reveals that political institutions are weaker than normal just before the start of a large economic crisis, but the crisis itself seems to trigger positive institutional change. This change happens across all indicators of political institutions and shows that institutions are not as static as is often thought. Third, we single out the decline phase of a slump which is defined as the time from the first break until the empirical trough and analyze its duration. We identify constraints on the executive, ethnic heterogeneity, and an interaction between these two variables as robust correlates of the duration of the decline phase. Furthermore, those same factors determine the total depth of the decline, but exclusively through the duration and not through the speed of contraction.

Our findings support the view that the length of the decline phase depends on the political system's ability to quickly react to a negative shock with coordinated policies in order to avert outright social conflict. A large body of political economy theory asserts the ability of resilient political institutions to internalize social conflict is essential to development (e.g. Acemoglu and Robinson, 2006, North et al., 2009, Besley and Persson, 2011). Some of these theories specifically argue that weakly institutionalized societies are especially prone to collapses because declining rents during a negative shock undermine the prevailing political arrangements (e.g. North et al., 2009). In fact, Broadberry and Wallis (2017) recently decomposed several hundreds of years of economic growth and come to the conclusion that a reduction in the frequency and rate of shrinking are an integral part of forging ahead. Broadberry and Wallis (2017) suggest that this is related to a transition from 'identity rules' among powerful groups to a system of impersonal rule for all. A key feature of societies based on identity rules is that they place few constraints on the ruling elite which creates a commitment problem towards other less powerful groups. Weak institutions thus bring with them the seeds of increased vulnerability to crises, as well as potentially much longer and deeper declines once crises occur. Similar mechanisms are suggested in the literature on institutions and macroeconomic volatility (Acemoglu et al., 2003, Mobarak, 2005).

The number and relative strength of groups with whom the ruling elite needs to coordinate plays a significant role in identity rule societies (Bluhm and Thomsson, 2015). This is especially important in Sub-Saharan Africa where power is typically concentrated in the executive branch and cabinet posts are distributed across ethnic groups in line with their population shares (Francois et al., 2015). Ethnic heterogeneity itself has been linked to a variety of coordination failures leading to inadequate policies, low provision of public goods and conflict. At the same time, greater diversity can be beneficial and may be necessary to reap the advantages of skill complementarities in highly diversified economies (Alesina and Ferrara, 2005). The negative effects of diversity may also 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). One of our key findings is that in the context of economic declines the (negative) effects of ethnic heterogeneity depend on certain political institutions and *vice versa*.

Few macroeconomic policy indicators are correlated with the length of economic declines once we account for constraints on the executive, ethnic heterogeneity and their interaction. This does not imply that macroeconomic policy is unimportant. In fact, reducing policy volatility for an extended period of time can reduce the duration of declines by a sizable amount. However, the mechanism we have in mind is more general: even if sound policy responses are available, coordination failures, rent seeking and power struggles combined with ethnic cleavages lead to substantially longer declines in heterogeneous countries with weak institutions.

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 reports the results of the event study. 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 employs tests of structural stability to identify and subsequently analyze the growth episode of interest.[1]

Standard structural break approaches (e.g. Bai and Perron, 2003) work well for identifying growth spurts but perform poorly when it comes to identifying growth collapses. A well-known advantage of these methods is that they allow us to distinguish abrupt regime changes from ordinary 'year-to-year' business cycle fluctuations. By definition, a structural break requires a departure from an ongoing growth process to be statistically significant at some reasonable level. This often discounts apparently large movements in economic activity which are actually in line with the observed fluctuations of a particular country, but can also deem smaller fluctuations in tranquil times as important.[2] An unappreciated weakness of standard structural break approaches, at least in our setting, is that they leave the particular type of structural change unspecified. As a result, they often do not identify the theoretically desired pattern of breaks but instead record all significant changes which must then be classified *ex post*.

To improve the identification of what we interchangeably refer to as deep recessions, slumps, or growth collapses, Papell and Prodan (2012, 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 first. 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 on the estimated coefficients to ensure the breaks occur in the desired direction.[3] Whereas they focus on the question whether growth in a few developed countries eventually returns to its pre-slump trend path, we apply a variant of their method to identify slumps in a large sample of countries over the period from 1950 to 2008.

Using a restricted structural change approach has another notable advantage: it can take meaningful restrictions on the growth process before and after the desired episode into account. For instance, a slump always interrupts a period of positive growth according to our definition. To see why these restrictions are useful, consider a sharp upward break in economic growth followed quickly by a deep recession. If the unrestricted Bai and Perron (2003) approach detects the upbreak first, it is likely to miss the downbreak which follows too soon. The appropriately restricted approach would instead dismiss the upbreak and detect the downbreak instead. Conversely, in the case of a double dip recession where the second dip is much deeper, the unrestricted

[1]See, for example, Hausmann et al. (2005), Jones and Olken (2008), Berg et al. (2012), or Pritchett et al. (2016). The Markov-switching models in Jerzmanowski (2006) and Kerekes (2012) are an exception.

[2]By contrast, uniform economic criteria would treat all series as if they were generated by the same process. Such criteria can therefore not discriminate among multiple plausible starting points or assess whether an episode truly constitutes a departure from the previous growth regime.

[3]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. Note that Papell and Prodan (2012) also refer to these episodes as 'economic slumps'.

approach might register only the second downbreak, whereas the restricted approach would locate the starting date of the slump at the first downturn, where the positive trend is interrupted. This is not just pure conjecture. We have encountered these problems when attempting to identify economic slumps by "inverting" the episodes of sustained growth from Berg et al. (2012). Online Appendix B shows that such a strategy is at best moderately successful.

Based on this discussion, 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 significant *regime switches* and not just ordinary business cycle fluctuations. What precisely constitutes a significant regime switch will vary and depend on the country's own, idiosyncratic growth process – a feature which we argue is desirable. 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* (although we also report results for aggregate output).

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

$$y_t = \alpha + \beta t + \gamma_0 \mathbf{1}(t > t_{b1}) + \gamma_1 (t - t_{b1}) \mathbf{1}(t > t_{b1}) + \gamma_2 (t - t_{b2}) \mathbf{1}(t > t_{b2}) + \sum_{i=1}^{p} \delta_i y_{t-i} + \epsilon_t \quad (1)$$

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 t_{b1}, γ_2 is a second trend change occurring after the second break at time t_{b2}, $\mathbf{1}(\cdot)$ 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.

Eq. (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 t_{b1}, (2) a *slump-recovery regime* lasting from time $t_{b1} + 1$ to time t_{b2}, and (3) a *post-slump regime* from time $t_{b2} + 1$ onwards. The second break (t_{b2}) is necessary to allow the return to the historical growth path after the recovery phase, or to some new relatively steady state. The location of the breakpoints is endogenous. 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. The other coefficients are left unrestricted, so that the model can catch slumps of various shapes and durations, even if they are unfinished (e.g., a decline starting at t_{b1} and lasting until the end of the time series).[4] This approach can easily be extended to permit other plausible structures, such as three-break models (including, e.g., a pre-slump regime, a contraction, a recovery and a post-slump regime). However, allowing for more than two breaks adds layers of complexity but has few, if any, additional benefits.[5]

[4]An alternative approach would be to impose the restriction $\beta + \gamma_1 + \gamma_2 > 0$, so that growth must be positive after the second break. A case like Togo in Figure 1d would then not qualify as a slump.

[5]We have experimented with three-break models using different parameter restrictions. This often identified exactly the same episodes as the simpler two-break model without providing a better estimate of the starting date. Moreover, with annual data, the contraction phase alone is often too brief to allow the observation of two separate breaks. Single break models with parameter restriction are an alternative but then require more ex post classification in return.

Following Papell and Prodan (2012), we implement the sequential break search algorithm as follows. First, we fit the structural change model specified in eq. (1) for all possible combinations of t_{b1} and t_{b2}. We always exclude 5% of the observations at the beginning and end of the sample to avoid registering spurious breaks. 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, since asymptotic break tests perform poorly in smaller samples (Antoshin et al., 2008, Prodan, 2008). If the bootstrap test rejects the null at the desired significance level, we record the break pair $(\widehat{t}_{b1}, \widehat{t}_{b2})$ and split the sample into a series running until the first break and a series starting just after the second break. We always report two sets of results for a nominal size of 10% and 20%, so as to arbitrate between committing type I and type II errors. The process starts again on each sub-sample until the bootstrap test fails to reject the null hypothesis of no breaks or the sample gets too small ($T \leq 20$).[6]

We do not attempt to characterize all types of breaks an economy can experience. Broken trends blur the conceptual distinction between unit-root and trend-stationary series, and are compatible with various models of aggregate output. Our approach is flexible since it allows for multiple breaks and for different growth regimes occurring before, during and after slumps. Although we assume that there is some structure in the growth process, it need not be generated by neoclassical steady-state growth, endogenous growth, or any other specific model of economic growth.

The duration of declines

As long as the level of GDP per capita preceding the slump has not been recovered, the economy is experiencing a loss of output. A slump is therefore only over when the level of GDP per capita has caught up again with its own past. If that point is reached within the sample, we define the recovery to have been completed in the first year $t_c > \widehat{t}_{b1}$ where $y_{t_c} \geq y_{\widehat{t}_{b1}}$. Within a slump, we identify the trough separating the decline from the recovery phase, and then focus squarely on the decline phase. This 'to the bottom' approach stands in contrast to the related literature which typically focuses on the entire duration of the slump (until the economy has fully recovered). Our rationale is that the decline and recovery processes are subject to very different dynamics and depend on different covariates. 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 income peak is reattained, which often passes without a change in the growth regime.

Naturally, the pre-slump level of GDP per capita is not always reached again within the sample period. In that case, the duration of the slump is censored. Even though GDP per capita may be recovering, we do not know how long it will take to restore the earlier peak. A provisional trough is then observed when y_t attains a minimum after $\widehat{t}_{b1}$. To cover all cases, we estimate the trough to have occurred at time:

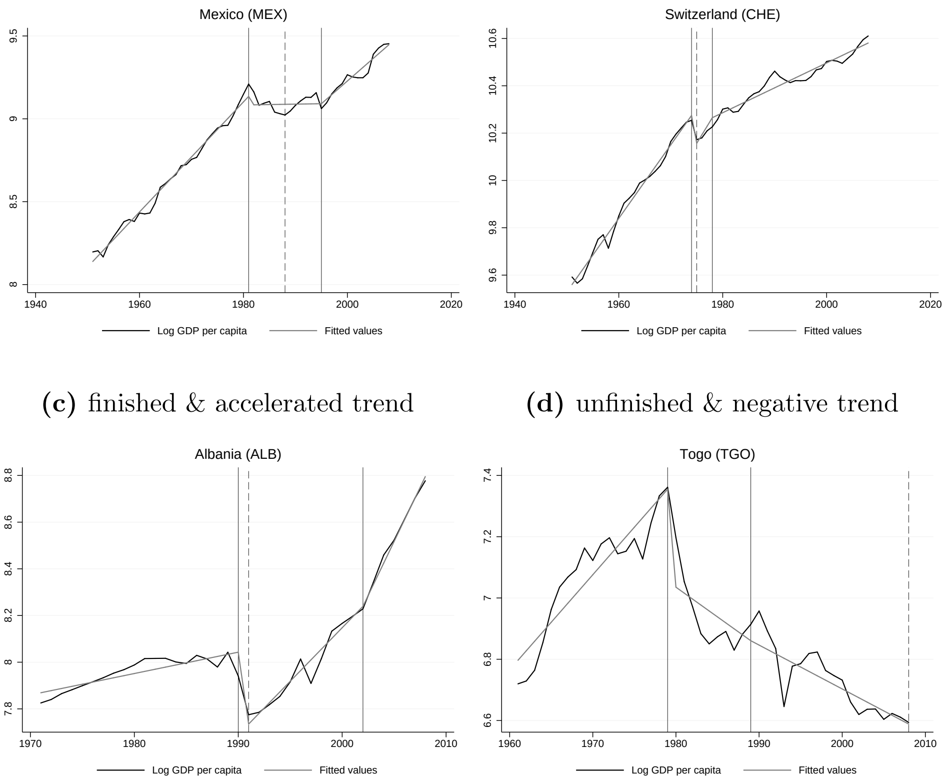
$$\widehat{t}_{min} = \begin{cases} \text{argmin}_{j \in (\widehat{t}_{b1}, t_c]} \, y_j & \text{if the spell is completed in year } t_c, \\ \text{argmin}_{j \in (\widehat{t}_{b1}, T]} \, y_j & \text{if the spell is censored.} \end{cases} \tag{2}$$

We denote the duration of the contraction phase, lasting from the initial break ($\widehat{t}_{b1}$) to the estimated trough ($\widehat{t}_{min}$), by $\widetilde{t}_D = \widehat{t}_{min} - \widehat{t}_{b1}$.

[6] Online Appendix B provides a formal description of the break search algorithm and the bootstrap.

It is important to note that the end of the slump when output has recovered (year t_c) does not in general coincide with the start of a new growth regime (year $\widehat{t}_{b2}$) and is used only as a device to identify the trough. In unfinished episodes it is even possible that the trough is dated after the estimated second break (see Figure 1d). For these unfinished spells, the true trough may lie in the future, that is, beyond the end of the sample period, and $\widetilde{t}_D$ is only a lower bound for the duration of the contraction. The analysis will treat such spells as censored, thereby fully taking this qualification into account.[7]

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 $\widehat{t}_{b1}$ and $\widehat{t}_{b2}$ but without the optimal AR(p) terms to emphasize the trend breaks. The bold vertical lines are at $\widehat{t}_{b1}$ and $\widehat{t}_{b2}$, respectively. The dashed vertical line indicates $\widehat{t}_{min}$.

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, inflation, high debt and low growth. The trough is found in 1988. Another short downturn occurs during the Tequila crisis in 1994 after which from 1995 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 economy is 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

[7] 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 (see Figure 1d). In a small number of such cases, the sequential algorithm will detect the start of a "new" slump before the previous one has ended. These are not distinct slumps and they are not included in the sample.

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 first break is located in 1990, the trough in 1991, and the second break in 2002, a few years after the end of the slump. Although the duration of the decline phase is only one year, GDP per capita contracted 15.32% in that year. 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. Togo's GDP per capita does 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 Characteristics of slumps

We apply the sequential algorithm to the entire Penn World Table (v7.0, series *rgdpch*) yielding 59 or 86 slumps between 1950 and 2008, depending on how permissive we are towards type I errors.[8] 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 (but later vary all potentially important parameters, including the end year). Online Appendix A lists all episodes and Online Appendix C provides summary statistics.

We observe several well-known growth collapses and deep recessions. Most slumps begin between the 1970s and the early 1990s. Several downbreaks occur following the oil shock in 1973–1974, another peak occurs between 1979 and 1981 during the debt crisis of the early 1980s, and several slumps follow the post-communist transitions of 1989–1990. We find no slumps starting in the period of the early 2000s and the tranquil mid-2000s.[9] 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 continents. 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 ('peak-trough ratio'). The spread of depth and duration is very large. We detect considerably longer declines and deeper slumps in low-income and middle-income countries than in high-income (OECD) countries, which is hardly surprising since the history of slumps directly affects the subsequent income category. The geographical distribution reveals interesting patterns. Africa, the Americas and Asia experience the deepest and longest declines. African slumps are striking in comparison to those in other regions. The *average* slump in Africa is deeper and longer than the 'Great Depression' in the United States. Due to their long duration, the continent is home to the most censored (unfinished) spells. Declines in Asia have been very deep as well but are generally much shorter.

The movement of other variables around the starting date of a slump is another

[8] 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, which only requires that an actual contraction occurs within the range of the two estimated breaks, otherwise there is no slump.

[9] Online Appendix C reports the annual frequency of starting dates for both samples.

Table 1 – Depth and duration, by income level and region

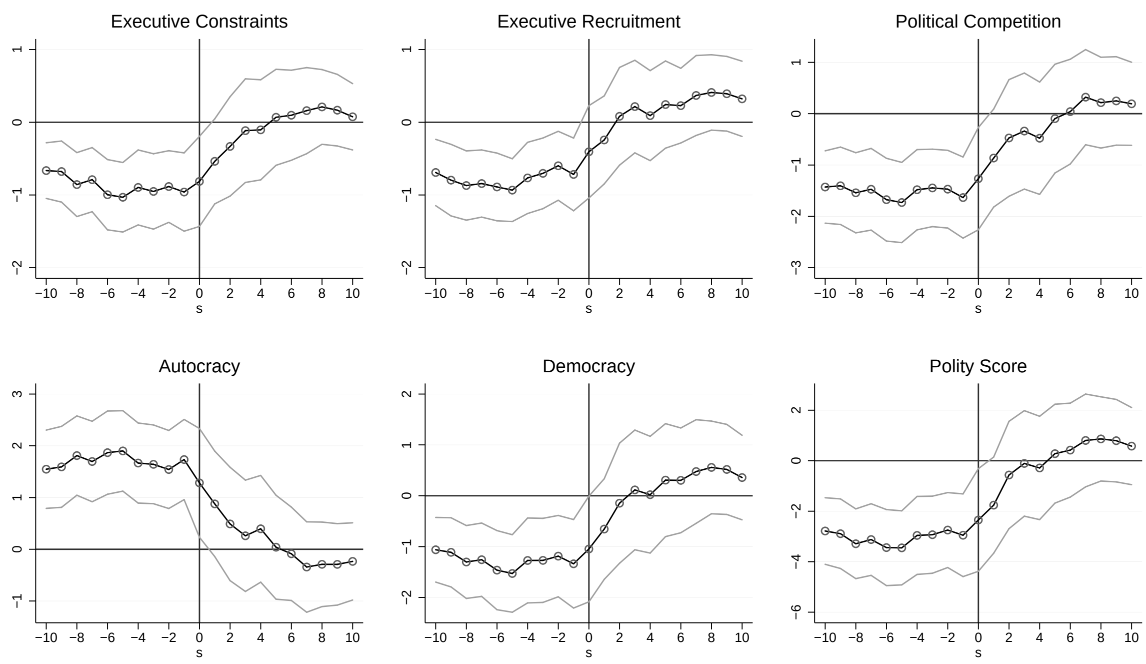
	Mean Depth	Median Depth	Mean Duration	Median Duration	Number of Spells	Censored Spells	Number of Countries
	Panel (a) Structural breaks estimated with size = 0.10						
Income Level							
High Income (OECD)	-7.1	-4.9	2.0	1	12	0	29
High Income (Other)	-19.4	-18.7	5.0	2	9	1	12
Upper Middle Income	-22.1	-17.0	5.6	3	15	2	30
Lower Middle Income	-25.6	-20.2	6.1	3	12	3	34
Low Income	-34.2	-30.7	15.8	16	11	4	33
Geographical Region							
Africa	-33.5	-30.7	16.2	10	15	6	46
Americas	-15.9	-16.5	4.8	2	17	1	25
Asia	-25.5	-18.7	4.5	2	15	3	35
Europe	-11.9	-10.6	2.6	1	9	0	29
Oceania	-4.0	-2.3	3.0	3	3	0	3
Total	-21.6	-17.5	7.7	3	59	10	138
	Panel (b) Structural breaks estimated with size = 0.20						
Income Level							
High Income (OECD)	-6.0	-3.3	1.9	2	19	0	29
High Income (Other)	-20.1	-18.7	5.2	3	9	1	12
Upper Middle Income	-19.7	-16.5	6.4	3	21	3	30
Lower Middle Income	-20.6	-17.1	6.1	3	20	4	34
Low Income	-29.3	-25.3	12.0	8	17	4	33
Geographical Region							
Africa	-24.0	-20.3	11.3	7	28	7	46
Americas	-16.6	-16.5	6.2	3	19	2	25
Asia	-24.3	-17.1	4.0	2	20	3	35
Europe	-8.9	-6.9	2.3	2	15	0	29
Oceania	-3.8	-2.7	2.5	2	4	0	3
Total	-18.8	-14.4	7.0	3	86	12	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.

interesting feature of their characteristics. We are especially interested in whether there is 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. Gourinchas and Obstfeld, 2012). The basic idea is to use dummy variables indicating the distance to the start of the slump as a means of detecting changes in the relative mean of each time-varying covariate.

We run the following regression for each measure of political institutions: $x_{it} = \sum_{s=-10}^{10} \delta_{t,\widehat{t}_{b1}+s}\beta_s + \mu_i + \epsilon_{it}$ where $\delta_{t,\widehat{t}_{b1}+s}$ is the Kronecker delta which is equal to one if $t = \widehat{t}_{b1} + s$ and zero otherwise, β_s are coefficients, μ_i is an unobserved country effect and ϵ_{it} is an idiosyncratic error term. We set $s \in \{-10, \dots, 0, \dots, 10\}$, so that the result is a 21-year window around the break date $\widehat{t}_{b1}$. The first year of the slump is $s = 1$ corresponding to $t = \widehat{t}_{b1} + 1$. The standard errors are robust to heteroskedasticity and autocorrelation within both country and time clusters (Cameron et al., 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.[10]

Figure 2 – Institutions and politics



All indicators describing a country's political institutions from the Polity IV database are significantly lower (or higher) in the decade before a slump occurs and then quickly return to normal levels thereafter. Figure 2 shows that, ten years before a slump, countries tend to score about one point less on indexes of executive constraints or executive recruitment, about two points less (or more) on indexes of political competition, autocracy or democracy, and about three points less on the combined Polity score.[11] These sizable differences vanish quickly after the downbreak. Just two years after the downbreak, we can no longer reject the null that they take on values comparable to "normal" times.

This remarkably consistent pattern suggests two stylized facts: (i) slumps are preceded by weak institutions, and (ii) abrupt negative growth creates room for political reform. Deep crises seem to increase the pressure on governments to pursue political change, illustrating the endogenous nature of reforms. Whereas it should be easier to bear the cost of reform in good times, it is often in bad times that the power balance shifts and the opposition to reform weakens. An important implication is that we have to rule out feedback from crises to institutions in the empirical analysis of slump duration.

Online Appendix E reports the results from similar tests for an array of macroeconomic policy indicators. The general pattern is as expected: the size of the government increases during a slump, just as investment falls, while prices and exchange rates react predictably. Other variables are trending over the entire period, but none exhibit such stark changes at the start of a slump as the pattern of institutional change documented here.

[10]In this case, "normal" refers to all observations other than the 21 years around the downbreak.

[11]These indexes have different ranges: 'Executive Constraints' range from 1 (unlimited authority) to 7 (parity or subordination), 'Executive Recruitment' from 1 (hereditary) to 8 (free and fair), 'Political Competition', 'Autocracy', and 'Democracy' from 1 to 10, and the 'Polity Score' from −10 to 10.

4 Explaining the duration of declines

Estimation strategy

We model the duration of declines in an 'accelerated failure time' (AFT) framework, where 'failure time' refers to time until the turnaround at $\widehat{t}_{min}$, that is, the end of the contraction phase. An AFT model is analogous to a classical log-linear regression model for duration and similarly easy to interpret. The hazard function and survival function are characterized indirectly by the distribution of the error terms, for which a normal distribution is only one among several possibilities.

AFT models are estimated by Maximum Likelihood (ML) which can easily incorporate incomplete and censored spells. If a contraction is observed from start ($\widehat{t}_{b1}$) to finish ($\widehat{t}_{min}$), its contribution to the likelihood is the probability of the recovery beginning at $\widehat{t}_{min}$, conditionally on the contraction having lasted until $\widehat{t}_{min}$. If the end of the contraction is not observed within the sample, then the spell is censored and only the so-called survival probability enters the likelihood. The difference between complete and incomplete spells is crucial. Persistent contractions are much more likely to be censored than brief contractions. Simply truncating unfinished spells biases their estimated duration downward. Dropping all unfinished spells is even worse, as persistent spells are then underrepresented and the sample biased towards short-lived contractions.

We have no strong theoretical prior concerning the error distribution or the implied shape of the hazard function. We may expect some countries to exit rather swiftly and others to take longer, but we leave open whether a prolonged decline phase will lead to a deterioration of fundamentals with a decreasing hazard, or whether the probability of a turnaround may be increasing over time as the contraction lasts.

Let $\widetilde{t} \equiv t - \widehat{t}_{b1}$ denote the time lapsed within a decline spell ('spell time'). We specify the following regression equation for crisis durations in AFT form:

$$\ln \widetilde{t} = \beta_1 INS_0 + \beta_2 ELF + \gamma(INS_0 \times ELF) + \mathbf{x}_0'\boldsymbol{\xi} + \mathbf{z}_t'\boldsymbol{\zeta} + \epsilon_t \tag{3}$$

where INS_0 is a measure of institutions fixed at $t_0 \equiv \widehat{t}_{b1}$, ELF is a time-invariant measure of ethno-linguistic fractionalization, $\mathbf{x}_0 = (x_{0,1}, x_{0,2}, \ldots, x_{0,k})'$ is a $k \times 1$ vector of controls fixed at $t_0 \equiv \widehat{t}_{b1}$, usually including region fixed effects and a constant, and $\mathbf{z}_t = (z_{t,1}, z_{t,2}, \ldots, z_{t,m})'$ is an $m \times 1$ vector of time-varying controls assumed strictly exogenous. In the case of a log-normal model, ϵ_t is distributed $\mathcal{N}(0, \sigma_\epsilon^2)$. Although we start off with this log-normal parameterization, we later test the robustness of our specification under different distributional assumptions.[12] Our main coefficients of interest are β_1, β_2 and γ. We suppress 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 covariates are in logs. They act as modifiers of the scale on which spell time is measured (which explains the term 'accelerated' failure time). A negative coefficient means that higher values of the covariate shorten the expected duration of the contraction phase and hasten recovery. Conversely, a positive coefficient means that higher values of the covariate delay recovery and prolong the expected duration of contraction (in effect 'decelerating' spell time).

The presence of the time-varying covariates $\mathbf{z}_t$ complicates estimation somewhat and becomes problematic if there is feedback from the spell duration to the covariates. In

[12] For details, see Table F-6 in Online Appendix F.

that case, the estimated coefficients are biased and the usual test statistics are invalid (Lancaster, 1990, Kalbfleisch and Prentice, 2002). We are mostly worried about the endogeneity of political institutions, given that the previous section highlights that they endogenously respond to crises. We are less concerned about ethnic diversity, since it is time-constant in our data and historically predetermined.[13] The interaction term is also less of a concern. Interactions of an endogenous variable with an exogenous variable are identified under fairly mild conditions (see Bun and Harrison, 2014).[14] Therefore, to prevent feedback, we set potentially endogenous covariates at their last pre-slump value, corresponding to $t_0 = \widehat{t}_{b1}$. In effect, we only rely on the temporal ordering to identify these relationships. The estimates should therefore be interpreted as partial correlations rather than causal effects.[15] Unobserved factors could still be determining political institutions and the duration of declines jointly, regardless of the proxies used in the regressions.

Results

Our baseline specification models the duration of declines as a function of executive constraints, ethnic fractionalization, initial GDP, and the real US interest rate. Constraints on the executive are our preferred proxy of political institutions for two reasons. First, the index 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). Our proxy for the number and relative strength of other identity groups in society is ethnic fractionalization. This variable can be interpreted as the probability that two randomly chosen individuals in a country belong to different ethno-linguistic groups and is often used in the literature on conflict (e.g. Esteban and Ray, 2011). We use the finest level of linguistic disaggregation from Desmet et al. (2012). Controlling for initial GDP is important, 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 (see Berg et al., 2012). This control too is important, since we cannot parametrize duration dependence and include a full set of time effects at the same time.

There are strong first-order effects of executive constraints and ethnic heterogeneity on the duration of declines. Column (1) to (3) in Table 2 present the corresponding results where we sequentially add region fixed effects and initial decade fixed effects to our base specification. Panel (a) reports the results for a false positive break detection rate of 10% and panel (b) increases this parameter to 20% in the hope of committing fewer type II errors. The coefficients of executive constraints and fractionalization are significant throughout. An improvement in executive constraints by one standard deviation (2.51

[13]The ethnic configuration of a country rarely changes as a short-run response to an economic decline, which is not to deny some tragic exceptions.

[14]Countries can have several recurrent slumps, but this is a minor concern in our application. To account for this dependence, we allow the error terms 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.

[15]Online Appendix D shows simple split sample plots where we plot the duration of declines over constraints of the executive or ethnic heterogeneity for different quantiles of the other variable. The lines are crossing at some point, already suggesting the presence of an interaction effect in the raw data.

points) is associated with a reduction in the duration of the decline phase of 19-30%. A difference in fractionalization of one standard deviation (33.88 percentage points) corresponds to a difference in duration of about 36-71%. Neither the inclusion of region fixed effects nor that of initial decade fixed effects has much effect on the results, although the latter attenuate the coefficient of executive constraints and weaken the influence of the US interest rate (which also proxies for time effects).

Table 2 – Main results

	Additive			*Interacted*		
	Fixed effects on ...					
	None (1)	Region (2)	Region + Time (3)	None (4)	Region (5)	Region + Time (6)
	Panel (a) Structural breaks estimated with size = 0.10					
Executive Constraints ($\widetilde{INS_0}$)	-0.140***	-0.135**	-0.084*	-0.232***	-0.217***	-0.188***
	(0.047)	(0.054)	(0.051)	(0.056)	(0.064)	(0.061)
Fractionalization ($\widetilde{ELF}$)	0.014***	0.016***	0.016***	0.015***	0.018***	0.019***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$\widetilde{INS_0} \times \widetilde{ELF}$				-0.003***	-0.003**	-0.004***
				(0.001)	(0.001)	(0.001)
Real US Interest Rate	0.100**	0.091**	0.063	0.107**	0.093**	0.046
	(0.048)	(0.042)	(0.045)	(0.049)	(0.042)	(0.045)
Initial log GDP	-0.068	-0.010	0.042	-0.075	-0.019	0.033
	(0.062)	(0.070)	(0.060)	(0.061)	(0.070)	(0.059)
Exits	48	48	48	48	48	48
Spells	58	58	58	58	58	58
Years of Decline	353	353	353	353	353	353
Log-$\mathcal{L}$	-73.286	-68.958	-65.651	-71.035	-67.026	-62.377
Pseudo-R^2	0.166	0.215	0.252	0.191	0.237	0.290
	Panel (b) Structural breaks estimated with size = 0.20					
Executive Constraints ($\widetilde{INS_0}$)	-0.127***	-0.135***	-0.095**	-0.193***	-0.190***	-0.154***
	(0.039)	(0.045)	(0.047)	(0.047)	(0.046)	(0.049)
Fractionalization ($\widetilde{ELF}$)	0.009**	0.012***	0.013***	0.010***	0.013***	0.014***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
$\widetilde{INS_0} \times \widetilde{ELF}$				-0.003***	-0.003***	-0.004***
				(0.001)	(0.001)	(0.001)
Real US Interest Rate	0.100***	0.089***	0.078*	0.102***	0.090***	0.064
	(0.037)	(0.034)	(0.044)	(0.037)	(0.034)	(0.041)
Initial log GDP	-0.089	-0.044	-0.007	-0.066	-0.011	0.029
	(0.057)	(0.062)	(0.060)	(0.060)	(0.068)	(0.064)
Exits	69	69	69	69	69	69
Spells	81	81	81	81	81	81
Years of Decline	465	465	465	465	465	465
Log-$\mathcal{L}$	-110.349	-107.193	-103.548	-107.192	-103.848	-99.549
Pseudo-R^2	0.103	0.128	0.158	0.128	0.155	0.190

Notes: All models include a constant. The standard errors are clustered at the country level to account for repeated spells. Some countries are not included in the Polity IV data, so that we analyze at most 58 (out of 59) or 81 (out of 86) spells. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A key point of our paper is that these first few models miss a potentially important interaction effect. If greater ethnic diversity 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 (e.g. see Rodrik, 1999, Bluhm and Thomsson, 2015). 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 achieve a similar degree of social coordination with less developed institutions. 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). Although such an effect has been proposed in the literature (e.g. Alesina and Ferrara, 2005), there is little empirical evidence of it, especially when it comes to the more precise channels through which political institutions can "mute" the adverse effects of ethnic heterogeneity.

We find strong, albeit indirect, evidence suggesting that coordination matters when it comes to managing downturns. Columns (4) to (6) of Table 2 present the same specifications as before with the addition of an interaction between ethnic heterogeneity and constraints on the executive. In order to ease the interpretation, the institutions and fractionalization variables have been demeaned by subtracting their sample average from the observed values.[16] The demeaned variables are denoted $\widetilde{INS_0}$ and $\widetilde{ELF}$. The interaction effect is negative (as expected), highly significant in each regression, and remarkably stable. Since the earlier specifications are nested in these columns, testing the significance of the interaction basically amounts to testing that the interaction model fits the data better. Indeed, likelihood ratio tests also prefer the interaction model and the pseudo-R^2s improve in every column.

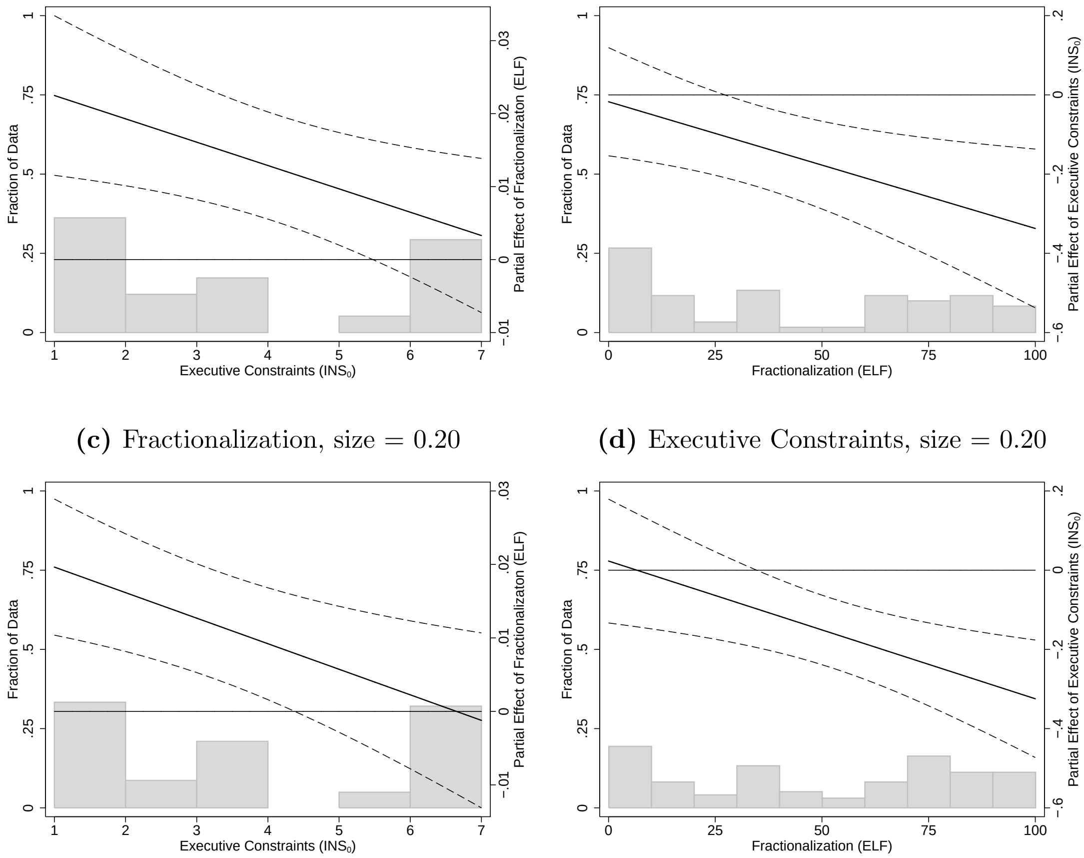
The estimated effects are both economically and statistically significant across a wide range of situations. Figure 3 plots the average partial effect with respect to one variable of the interaction term over representative values of the other, including a 95% confidence interval. The vertical axis on the right measures the average predicted semi-elasticity of the expected duration with respect to fractionalization or executive constraints. The effect sizes can be interpreted just like the coefficients. For example, when the executive constraints index is at unity ('unlimited authority'), then a one percentage point higher fractionalization corresponds to a 2% longer duration of contractions (and a one standard deviation rise in fractionalization almost doubles the expected duration). In the background, Figure 3 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).

The predictions also cover a 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 15.5 years (2.6 years). Hence, it would be difficult to understand the effects of institutions without taking fractionalization into account. Countries with stronger institutions can overcome the adverse effects of high fractionalization. At the 75^{th} percentile of ethnic heterogeneity (= 89.7), a country with the highest (lowest) score of executive constraints is expected to decline for about 3.2 years (19.7 years). Or, as panel (b) shows, the partial effect of a unit change in executive constraints at perfect homogeneity is practically zero, while it peaks at about -34% at perfect heterogeneity. The estimates for the larger sample are nearly identical.

For our second set of main results we investigate whether the effects of institutions and ethnic cleavages on the depth of slumps run solely through the duration of the decline phase or if they also affect the rate of contraction. This is an important distinction to

[16]This has the following effect. If either one of the two variables is at its mean, then the interaction term is zero. As a result, the coefficient of the institutions variable directly measures the effect of executive constraints at the average level of fractionalization, and vice versa. Away from the means, the coefficient on the interaction term becomes important. Note that the statistical significance of the interaction term is not affected, and that demeaning is irrelevant when the interaction term is absent.

Figure 3 – Average partial effects



(a) Fractionalization, size = 0.10

(b) Executive Constraints, size = 0.10

(c) Fractionalization, size = 0.20

(d) Executive Constraints, size = 0.20

Note(s): The average partial effects are based on column (5) in panel (a) and (b) of Table 2 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.

make. We ultimately care about the *net present value of lost output* (Pritchett et al., 2016) which is a function of these two variables. We use a simple decomposition of the depth of declines which abstracts from the potential income gains a country could experience in absence of a slump and simply takes pre-slump GDP per capita (y_{i0}) as the reference point. The depth of the decline in percent of pre-slump GDP per capita is the product of the estimated duration and the average rate of contraction. Hence, we may define $\bar{g}_i \equiv \left(y_{i,\widehat{t}_{min}} - y_{i0}\right) / (\widehat{t}_{min} - t_{b1}) \equiv \left(y_{i,\widehat{t}_{min}} - y_{i0}\right) / \widetilde{t}_D$ as the average rate of decline and $\widetilde{t}_D \times \bar{g}_i \equiv y_{i,\widehat{t}_{min}} - y_{i0}$ as the overall depth of the decline. Note that $\bar{g}_i$ is negative by construction. We scale both outcomes by one hundred for readability.

We find only weak evidence in favor of an effect of either executive constraints or fractionalization on the average rate of decline, but effects comparable to those from the duration regressions when we consider the overall depth of declines. Columns (1) and (2) in both panels of Table 3 report the results for the model without and with mean-centered interaction terms. Column (3) adds initial decade effects. While several coefficients are significant, they are economically small (e.g. about 0.4 percentage points of $\bar{g}_i$ in the case of a unit change in executive constraints) and explain little of the variation on top of the regional (and decade) fixed effects. The remaining columns of Table 3 examine the overall depth of the slump. Columns (3) and (4) illustrate that we now recover the previously

Table 3 – Average rate of decline and total depth

	Dependent variable					
	$\bar{g}_i$ (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)
	Panel (a) Structural breaks estimated with size = 0.10					
Executive Constraints ($\widetilde{INS}_0$)	0.375*	0.449**	0.361	2.727**	3.403***	3.268***
	(0.200)	(0.208)	(0.247)	(1.058)	(1.024)	(1.082)
Fractionalization ($\widetilde{ELF}$)	-0.006	-0.001	0.002	-0.282***	-0.235***	-0.199***
	(0.016)	(0.015)	(0.016)	(0.092)	(0.074)	(0.069)
$\widetilde{INS}_0 \times \widetilde{ELF}$		0.010**	0.012**		0.089***	0.106***
		(0.004)	(0.005)		(0.022)	(0.026)
Initial log GDP	0.278	0.316	0.211	0.702	1.047	0.834
	(0.328)	(0.335)	(0.379)	(1.446)	(1.424)	(1.484)
Real US Interest Rate	-0.202	-0.253	-0.286	0.107	-0.365	0.215
	(0.167)	(0.171)	(0.279)	(0.902)	(0.874)	(1.641)
Region FE	YES	YES	YES	YES	YES	YES
Initial Decade FE	NO	NO	YES	NO	NO	YES
Spells	58	58	58	58	58	58
Log-$\mathcal{L}$	-154.543	-153.482	-151.873	-253.874	-250.889	-247.146
Adjusted R^2	0.066	0.081	0.051	0.249	0.308	0.337
	Panel (b) Structural breaks estimated with size = 0.20					
Executive Constraints ($\widetilde{INS}_0$)	0.306	0.330*	0.279	2.731***	2.920***	2.691***
	(0.186)	(0.190)	(0.196)	(0.920)	(0.911)	(0.962)
Fractionalization ($\widetilde{ELF}$)	0.002	0.007	0.008	-0.184**	-0.144**	-0.125*
	(0.017)	(0.016)	(0.018)	(0.078)	(0.067)	(0.069)
$\widetilde{INS}_0 \times \widetilde{ELF}$		0.009**	0.010**		0.073***	0.078***
		(0.004)	(0.004)		(0.021)	(0.022)
Initial log GDP	0.255	0.175	0.116	0.990	0.346	0.080
	(0.323)	(0.336)	(0.349)	(1.260)	(1.318)	(1.376)
Real US Interest Rate	-0.051	-0.069	0.005	0.269	0.121	0.744
	(0.124)	(0.123)	(0.191)	(0.694)	(0.655)	(1.291)
Region FE	YES	YES	YES	YES	YES	YES
Initial Decade FE	NO	NO	YES	NO	NO	YES
Spells	81	81	81	81	81	81
Log-$\mathcal{L}$	-208.648	-206.608	-205.432	-353.725	-349.940	-348.261
Adjusted R^2	0.093	0.125	0.099	0.145	0.210	0.197

Notes: All models include a constant. The standard errors are clustered at the country level to account for repeated spells. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

estimated effects with similar significance. (The coefficients have reversed signs since $\bar{g}_i$ is negative.) These results are robust to the inclusion of initial decade fixed effects as shown in column (6). Moreover, the interaction models explain 20–30% of the variation in depth, highlighting the relevance of the estimated effects. As in the previous tables, the results remain qualitatively similar when we consider the larger set of slumps.

Taken together, this leads us to conclude that (i) political institutions and ethnic heterogeneity are robustly correlated with the overall depth of slumps, and (ii) these correlations run primarily through the duration of the decline phase.

One way to interpret this finding 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 increases 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. In fact, a stronger executive should reduce the duration of the decline phase in such models, since it implies that other groups possess less veto power.

Bluhm and Thomsson (2015) propose a different 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, some groups 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. Conversely, if institutions are sufficiently strong, 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 commitment problem 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 generates the observed interaction. The commitment problem gets worse with increasing group diversity, but can be resolved by stronger constraints on the executive at all levels of heterogeneity.

As discussed in the introduction, the argument that impersonal rules and limits on executive power can help to mitigate the risk of experiencing deeper economic contractions parallels recent historical work. Broadberry and Wallis (2017) trace historical improvements in the frequency and rate of shrinking and the subsequent stabilization of positive growth to the replacement of identity rules by a system of impersonal rule. Societies based on identity rules place few constraints on the power of the elite vis-à-vis the rest of society but grant some rule of law within elites. Most importantly, unconstrained power makes it difficult for rulers to credibly commit to rule-based dealings with less powerful groups. This is precisely how we think ethnic heterogeneity enters the picture. It reflects the presence of other identity groups with whom the ruling elite needs to coordinate (cf. Bluhm and Thomsson, 2015).

Robustness

Our main conclusions are unaffected by a variety of robustness checks, such as dropping influential observations, varying the national accounts data, and the inclusion of policy variables. We only present a selection of the most insightful perturbations here. A comprehensive battery of robustness checks can be found in Online Appendix F.

Table 4 takes the estimated breaks and data as given and focuses on unobserved heterogeneity and influential observations. Column (1) allows for country-level random effects (frailties) on top of region fixed effects, while column (2) adds a set of initial income level dummies. Both perturbations do not affect our findings. Column (3) takes a more radical step and drops all of Africa. The estimated effect sizes are a bit smaller and we lose precision on the interaction term in the smaller data set, but not the expanded version. This loss of significance is hardly surprising. The African continent is home to some of the most ethnically diverse countries with the weakest institutions. Column (4) deletes the censored spells. Now we observe a sizable drop in the estimated effect of political institutions, a halving of the effect of fractionalization, and the interaction term becomes insignificant. A look into the underlying data reveals what is going on. Countries with completed spells score 3.3 on average on the index of executive constraints and have an average degree of heterogeneity close to 50 out of 100. Countries with ongoing slumps score 1.45 in terms of executive constraints and have an average level of heterogeneity close to 82. The values refer to the smaller sample, but are not much different for the data underlying panel (b). This strongly suggests that it is the composition of the two groups that matters and not the fact that they are censored. The whole point of duration analysis is that we can use these observations without introducing bias. Note that if we determine the slumps in terms of GDP rather than GDP per capita, there are much fewer censored slumps and our results hold in both samples without these data points.

Table 4 – Robustness: heterogeneity, outliers and multiple spells

	Current modification					
	Country RE (1)	Income FE (2)	No Africa (3)	No Censored (4)	Single Spells (5)	PWP (6)
	Panel (a) Structural breaks estimated with size = 0.10					
Executive Constraints ($\widetilde{INS_0}$)	-0.217***	-0.240***	-0.149***	-0.087	-0.243***	1.470***
	(0.070)	(0.071)	(0.054)	(0.055)	(0.068)	(0.142)
Fractionalization ($\widetilde{ELF}$)	0.018***	0.019***	0.010*	0.010***	0.016***	0.977***
	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003**	-0.003**	-0.002	-0.001	-0.004**	1.005**
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Region FE	YES	YES	YES	YES	YES	YES
Income Level FE	NO	YES	NO	NO	NO	NO
Exits	48	48	39	48	41	48
Spells	58	58	43	48	51	58
	Panel (b) Structural breaks estimated with size = 0.20					
Executive Constraints ($\widetilde{INS_0}$)	-0.188***	-0.190***	-0.108**	-0.087**	-0.216***	1.270***
	(0.056)	(0.046)	(0.051)	(0.039)	(0.052)	(0.074)
Fractionalization ($\widetilde{ELF}$)	0.013***	0.015***	0.008*	0.008**	0.011**	0.987***
	(0.004)	(0.004)	(0.005)	(0.003)	(0.004)	(0.005)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003**	-0.003**	-0.003**	-0.001	-0.004***	1.005***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Region FE	YES	YES	YES	YES	YES	YES
Income Level FE	NO	YES	NO	NO	NO	NO
Exits	69	69	50	69	55	69
Spells	81	81	55	69	64	81

Notes: All models include the real US interest rate, initial GDP, and a constant. The standard errors are clustered at the country level to account for repeated spells. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Until now, we have assumed that multiple spells of the same type are interchangeable. The last two columns of Table 4 investigate if this particular 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 (Prentice et al., 1981). The model accounts for the ordering of events based on the natural assumption that a country cannot exit a second decline phase before having left the first. Here too, the results are qualitatively similar but are now reported in terms of hazard ratios. A hazard ratio greater than one implies a higher probability of exiting the decline, and vice versa.

A major concern in the cross-country growth literature is the fragility of regression estimates with respect to different GDP series. Johnson et al. (2013) show that new renditions of the Penn World Tables incorporating more recent PPP benchmarks can imply vastly different historical growth rates. Worse still, they point out that the findings of several prominent studies, including Ramey and Ramey (1995) and Hausmann et al. (2005) cited in the introduction, are not robust to different versions of the PWT.[17] In fact, this has prompted major changes in how the PWT data are prepared and made available to the public. Starting from version 8 different time series are reported for level and growth rate comparisons.[18] We take these concerns seriously and use three very different data sets in the restricted structural break search to examine the robustness of our findings: the PWT version 7, the PWT version 9, and the World Development Indicators. Each time we consider GDP per capita and GDP, sometimes in PPPs, sometimes in local currencies. We prefer GDP per capita throughout the paper since it is closer to a welfare relevant metric but expect to obtain similar results with GDP.

Table 5 – Robustness: different data sets for GDP and GDP per capita

	GDP per capita series			*GDP series*		
	PWT7 *rgdpch* (1)	PWT9 *rgdpna/pop* (2)	WDI *gdplcu/pop* (3)	PWT7 *rgdpch* × *pop* (4)	PWT9 *rgdpna* (5)	WDI *gdplcu* (6)
	Panel (a) Structural breaks estimated with size = 0.10					
Executive Constraints ($\widetilde{INS_0}$)	-0.217***	-0.201***	-0.162**	-0.184***	-0.146***	-0.132***
	(0.064)	(0.073)	(0.077)	(0.046)	(0.048)	(0.048)
Fractionalization ($\widetilde{ELF}$)	0.018***	0.013**	0.010*	0.011***	0.010***	0.009***
	(0.004)	(0.005)	(0.006)	(0.003)	(0.002)	(0.003)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003**	-0.003**	-0.006**	-0.003***	-0.003***	-0.004***
	(0.001)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)
Exits	48	45	40	59	50	44
Spells	58	54	54	62	50	45
	Panel (b) Structural breaks estimated with size = 0.20					
Executive Constraints ($\widetilde{INS_0}$)	-0.190***	-0.200***	-0.048	-0.126***	-0.156***	-0.092**
	(0.046)	(0.055)	(0.061)	(0.037)	(0.045)	(0.044)
Fractionalization ($\widetilde{ELF}$)	0.013***	0.012***	0.006	0.008***	0.010***	0.008***
	(0.004)	(0.004)	(0.005)	(0.002)	(0.002)	(0.003)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003***	-0.004***	-0.004**	-0.004***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Exits	69	57	56	74	62	61
Spells	81	69	72	77	63	63

Notes: All models include region FEs, the real US interest rate, initial GDP, and a constant. The standard errors are clustered at the country level to account for repeated spells. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Our main findings are virtually unaffected by the choice of data sets and preference for a particular national income series. Table 5 takes our preferred specification including the mean-centered interaction term and region fixed effects but replaces the dependent variable with the duration estimated by running our break search algorithm on these different data sets. Columns (1) to (3) focus on GDP per capita, while columns (4) to (6) report results for GDP. As usual, we consider two type I error rates in panels (a) and (b). In all but one instance the results are qualitatively (and often even numerically) similar: only in the case of GDP per capita in local currency units from the WDI in panel (b) the coefficients on the mean-centered constituent terms are not statistically significant. This is hardly cause for concern since the interaction effect is robust and the averages

[17] We are indebted to an anonymous referee for pointing this out to us.

[18] The PWT version 7 already offered a new series, *rdgpl2*, based on the growth rate of domestic absorption, as a response to the work of Johnson et al. (2013). We obtain very similar results when using this series instead of *rgdpch*.

shift around in different samples. Note that the underlying list of slumps differs for each output series we consider, in part because the growth rates are computed differently and in part because these data sets have different temporal and geographical coverage. For comparability, we therefore always stop in 2008 even if more data points are available.[19]

The relationships reported here survive many other useful tests. They are robust to (*i*) using different bootstrap techniques, (*ii*) varying the interbreak period, (*iii*) using longer time series, (*iv*) adding other controls individually or in groups, (*v*) changing the functional form, and (*vi*) dropping individual countries one-by-one. Using different institutional indicators shows that – in line with theory – a narrow measure of constraints on the executive generates stronger interactions than broader measures of political institutions. All corresponding results are reported in Online Appendix F.

Table 6 – Robustness: policy variables

	Policy variable					
	Policy Volatility (1)	Government Size (2)	Investment Share (3)	Inflation ($\ln[1+\delta]$) (4)	RER Underval (5)	Trade Openness (d.f.) (6)
	Panel (a) Structural breaks estimated with size = 0.10					
Executive Constraints ($\widetilde{INS_0}$)	-0.221***	-0.217***	-0.227***	-0.230***	-0.216***	-0.261***
	(0.068)	(0.063)	(0.064)	(0.064)	(0.064)	(0.064)
Fractionalization ($\widetilde{ELF}$)	0.018***	0.018***	0.018***	0.021***	0.018***	0.013***
	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)	(0.003)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003**	-0.003**	-0.004***	-0.004***	-0.003**	-0.005***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Policy Variable	0.039*	-0.000	0.010	0.000	0.217	0.016***
	(0.022)	(0.017)	(0.009)	(0.003)	(0.310)	(0.004)
Exits	48	48	48	37	48	48
Spells	58	58	58	44	58	58
	Panel (b) Structural breaks estimated with size = 0.20					
Executive Constraints ($\widetilde{INS_0}$)	-0.181***	-0.188***	-0.191***	-0.165***	-0.189***	-0.193***
	(0.044)	(0.047)	(0.046)	(0.054)	(0.046)	(0.041)
Fractionalization ($\widetilde{ELF}$)	0.014***	0.013***	0.014***	0.016***	0.013***	0.009***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003***	-0.003***	-0.004***	-0.004**	-0.003***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Policy Variable	0.050**	-0.006	0.011	-0.004	-0.058	0.012***
	(0.022)	(0.016)	(0.009)	(0.003)	(0.275)	(0.003)
Exits	69	69	69	53	69	69
Spells	81	81	81	62	81	81

Notes: All models include region FEs, the real US interest rate, initial GDP, and a constant. The standard errors are clustered at the country level to account for repeated spells. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Finally, we examine if accounting for macroeconomic policies changes the nature of our conclusions with respect to political institutions and heterogeneity.[20] Much of the earlier literature argues that (bad) policies are mostly a function of (weak) institutions. Fatás and Mihov (2013), however, show that the volatility of macroeconomic policy matters for growth, even within the *same* set of institutions. Table 6 reports the results from adding several proxies for economic policies to our regression. As in Fatás and Mihov

[19] Online Appendix F reports additional results using data up to 2014 which do not differ substantially from those reported here. If anything, the results using the WDI data become stronger.

[20] We are indebted to an anonymous referee for pointing us in this direction.

(2013), policy volatility is defined as the (rolling) standard deviation of the residuals from country-level regressions of the log-difference of real government consumption spending on the log difference of real GDP. Few of the added indicators turn out to be significant predictors of the duration of declines, apart from policy volatility and, perhaps less surprisingly, trade openness. We are therefore able to confirm that greater variation in macroeconomic policies increases the duration of declines, but the presence of this effect does not compete with our main results. Interestingly, its partial effect is comparable to the effect of institutions in moderately to highly fractionalized countries. A standard deviation increase in policy volatility increases the duration of declines by about 27% to 36%, depending on the sample. Establishing a stable policy environment for an extended period of time thus seems just as important as constraining executive power in heterogeneous countries.

5 Concluding remarks

Severe downward volatility is an ubiquitous phenomenon in the post-war period. While the literature has often stressed the role of positive growth spurts, our paper emphasizes that slumps can quickly undo the gains from growth and that, in some cases, the decline phase lasts very long. As it turns out, economic slumps are harder to identify than growth accelerations but also have a distinct shape. We show that a restricted structural change approach directly incorporating this shape works well as an inferential method for identifying economic slumps in a large sample of countries. We find a substantial number of slumps of varying length in developing and developed countries alike.

In the title of this paper we ask if weak institutions prolong crises – a question we believe can only be meaningfully answered with cross-country data. Our answer is yes, constraining leaders is beneficial for limiting the downside of negative shocks. However, our analysis suggests that there are, at least, two important qualifications to be made.

First, our event study illustrates that weak political institutions precipitate crises and positive institutional change occurs during and in the immediate aftermath of slumps. Our interpretation of this stylized fact is that, while good institutions may sustain growth, growth collapses can in turn contribute to endogenous institutional change. Severe economic crises seem to raise the pressure for institutional reform in a very broad sense.

Second, the length and depth of economic slumps is negatively correlated with constraints on the executive but the favorable effects of strengthening constraints are greater in ethnically heterogeneous societies. We take this as evidence that effective coordination and responses to slumps are hampered by the existence of many identity groups with whom the ruling group needs to coordinate. However, the coordination problems implied by ethnic heterogeneity are not immutable and can be overcome by strong impersonalized rules. We also provide evidence that these effects run primarily through the duration until the recovery starts and not through the pace of decline.

This interplay between weakly constrained leaders 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. Our results are more in line with recent historical accounts which stress the emergence of impersonal rules, though without recognizing the role of ethnic or interest group diversity. Hence, a key reason to strengthen political institutions is the stabilization of growth, particularly in diverse countries and perhaps more so than the literature emphasizes so far.

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.

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

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.

Antoshin, S., A. Berg, and M. R. Souto (2008, March). Testing for structural breaks in small samples. IMF Working Papers 08-75, International Monetary Fund.

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 (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*(1), 1–22.

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.

Broadberry, S. and J. J. Wallis (2017, April). Growing, shrinking, and long run economic performance: Historical perspectives on economic development. Working Paper 23343, National Bureau of Economic Research.

Bun, M. J. and T. D. Harrison (2014, December). OLS and IV estimation of regression models including endogenous interaction terms. UvA-Econometrics Working Papers 14-02, Universiteit van Amsterdam, Dept. of Econometrics.

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.

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

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

Esteban, J. and D. Ray (2011). Linking conflict to inequality and polarization. *American Economic Review 101*(4), 1345–74.

Fatás, A. and I. Mihov (2013). Policy volatility, institutions, and economic growth. *Review of Economics and Statistics 95*(2), 362–376.

Francois, P., I. Rainer, and F. Trebbi (2015). How is power shared in africa? *Econometrica 83*(2), 465–503.

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

Hausmann, R., L. Pritchett, and D. Rodrik (2005). Growth accelerations. *Journal of Economic*

Growth 10(4), 303–329.

Hausmann, R., F. Rodriguez, and R. Wagner (2008). Growth collapses. In C. Reinhart, C. Vegh, and A. Velasco (Eds.), *Money, Crises and Transition*, pp. 376–428. Cambridge, Mass.: MIT Press.

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

Johnson, S., W. Larson, C. Papageorgiou, and A. Subramanian (2013). Is newer better? Penn World Table revisions and their impact on growth estimates. *Journal of Monetary Economics 60*(2), 255–274.

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

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

Kerekes, M. (2012). Growth miracles and failures in a markov switching classification model of growth. *Journal of Development Economics 98*(2), 167 – 177.

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

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

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 (2012). The Statistical Behavior of GDP after Financial Crises and Severe Recessions. *The B.E. Journal of Macroeconomics 12*(3), 1–31.

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.

Pritchett, L., K. Sen, S. Kar, and S. Raihan (2016). Trillions gained and lost: Estimating the magnitude of growth episodes. *Economic Modelling 55*, 279–291.

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.

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

Online appendix

A List of breaks

Table A-1 – Global parameters

Data:	PWT ($rgdpch$)	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. t_{bi} distance (h):	4	Bootstrap significance (α^s):	0.2 (0.1 is a subset)

Table A-2 – Estimated slumps: 86 episodes[a,b,c]

Code	T_0	$\widehat{tb}_1$	$\widehat{t}_{min}$	$\widehat{tb}_2$	T	Sup-W	Critical W	p-value	Drop (%)	Duration	c
AFG	1970	1986	1994	1996	2008	8.5	10.9	0.136	-69.57	8	0
ALB	1970	1990	1991	2002	2008	18.5	13.9	0.009	-15.32	1	0
ARE	1986	1990	1999	2002	2008	29.1	14.8	0.006	-10.90	9	0
AUS	1950	1954	1957	1966	2008	8.3	8.4	0.051	-0.72	3	0
AUS	1967	1981	1982	1985	1988	6.5	9.4	0.175	-3.12	1	0
AUS	1967	1989	1991	1998	2008	10.1	10.2	0.057	-2.29	2	0
BDI	1960	1971	1972	1988	2008	9.9	11.2	0.085	-3.23	1	0
BEL	1950	1957	1958	1973	2008	12.8	12.4	0.038	-2.24	1	0
BEN	1959	1975	1978	1982	2008	6.5	8.5	0.177	-11.72	3	0
BEN	1983	1986	1994	2007	2008	10.5	9.4	0.033	-11.82	8	0
BGR	1970	1988	1997	1997	2008	16.3	13.1	0.014	-23.79	9	0
BHR	1970	1980	1987	1986	2008	14.4	11.1	0.009	-44.12	7	1
BRA	1950	1980	1983	2003	2008	12.5	11.8	0.040	-14.60	3	0
BWA	1960	1964	1965	1972	2008	8.7	11.1	0.192	-11.82	1	0
CAF	1960	1978	2005	2005	2008	8.3	8.2	0.047	-46.38	27	1
CHE	1950	1974	1975	1978	2008	10.7	11.2	0.071	-7.87	1	0
CHL	1951	1953	1954	1972	1973	12.0	7.9	0.007	-9.06	1	0
CHL	1951	1974	1975	1979	1980	13.3	11.5	0.020	-16.50	1	0
CHL	1951	1981	1983	1995	2008	12.6	11.1	0.020	-21.22	2	0
CHN	1952	1960	1962	1977	2008	13.9	13.1	0.027	-23.71	2	0
CIV	1960	1979	1981	1985	2008	8.2	10.8	0.181	-12.64	2	0
CIV	1986	1989	2008	1996	2008	7.2	10.9	0.162	-24.08	19	1
CMR	1960	1966	1967	1973	1985	9.2	14.2	0.170	-12.84	1	0
CMR	1960	1986	1995	1990	2008	12.0	12.1	0.054	-40.46	9	1
COG	1960	1974	1977	1982	2008	11.9	12.3	0.063	-21.35	3	0
CRI	1950	1955	1956	1963	1979	11.4	11.0	0.047	-4.39	1	0
CRI	1950	1980	1982	2002	2008	17.2	10.8	0.002	-17.47	2	0
CUB	1970	1988	1993	1995	2008	11.4	12.6	0.085	-34.70	5	0
CYP	1950	1957	1960	1964	1972	10.6	12.8	0.089	-16.75	3	0
CYP	1950	1973	1975	1977	2008	15.5	9.6	0.001	-31.40	2	0
DNK	1950	1954	1955	1965	2008	12.9	11.8	0.030	-1.56	1	0
DZA	1960	1984	1994	1996	2008	10.9	8.1	0.010	-14.09	10	0
EGY	1950	1971	1974	1982	2008	8.1	9.8	0.112	-10.37	3	0
EGY	1983	1986	1987	1992	2008	16.1	13.0	0.018	-6.94	1	0
ETH	1950	1972	1992	1993	2008	11.5	9.8	0.019	-30.68	20	0
FIN	1950	1989	1993	2006	2008	10.6	11.8	0.084	-16.34	4	0
GAB	1960	1976	1987	1997	2008	10.6	11.3	0.068	-50.56	11	1
GMB	1960	1982	1998	2002	2008	16.4	11.3	0.002	-25.33	16	0
GRC	1951	1973	1974	1994	2008	17.9	12.1	0.000	-6.92	1	0
GTM	1950	1980	1988	1984	2008	15.1	12.3	0.014	-19.14	8	0

Continued on next page

Table A-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
HUN	1970	1990	1992	2004	2008	15.6	13.7	0.030	-10.56	2	0
IDN	1960	1997	1999	2001	2008	13.5	11.3	0.016	-17.49	2	0
IND	1950	1978	1979	2001	2008	11.3	13.6	0.145	-4.30	1	0
IRL	1950	1955	1957	1979	1980	10.1	10.1	0.049	-4.81	2	0
IRL	1950	1981	1983	2006	2008	8.1	10.6	0.143	-3.34	2	0
IRN	1955	1976	1981	1980	2008	15.9	11.0	0.002	-56.78	5	1
IRQ	1970	1990	2003	1994	2008	9.1	10.1	0.075	-66.43	13	1
JPN	1950	1973	1974	1990	2008	13.5	13.4	0.050	-2.85	1	0
LAO	1970	1996	1997	2004	2008	8.3	10.3	0.135	-3.91	1	0
MAR	1950	1954	1957	1962	2008	8.8	10.0	0.104	-21.94	3	0
MEX	1950	1981	1988	1995	2008	11.9	10.6	0.021	-17.03	7	0
MLI	1960	1963	1974	2005	2008	5.9	7.8	0.155	-19.29	11	0
MNG	1970	1990	1993	2003	2008	46.5	12.4	0.000	-41.81	3	0
MOZ	1960	1981	1986	1995	2008	12.6	12.2	0.037	-24.99	5	0
MYS	1955	1984	1986	1993	2008	9.1	11.3	0.119	-7.47	2	0
NLD	1950	1957	1958	1973	1977	13.3	9.8	0.009	-5.18	1	0
NLD	1950	1978	1982	1983	2008	8.1	9.9	0.155	-1.27	4	0
NPL	1960	1979	1980	2000	2008	10.6	9.9	0.035	-5.33	1	0
NZL	1950	1974	1978	1992	2008	9.9	10.5	0.067	-9.03	4	0
OMN	1970	1979	1980	1985	2008	12.4	8.9	0.008	-21.61	1	0
PAK	1950	1996	1997	2006	2008	8.9	10.9	0.153	-11.19	1	0
PAN	1950	1982	1989	2003	2008	8.7	10.2	0.123	-8.75	7	0
PER	1950	1958	1959	1966	1976	11.9	9.0	0.014	-6.91	1	0
PER	1950	1977	1992	1992	2008	11.0	10.6	0.039	-29.30	15	0
PHL	1950	1983	1985	2003	2008	12.8	10.4	0.012	-16.78	2	0
POL	1970	1979	1982	1993	2008	13.8	11.7	0.018	-22.55	3	0
PRI	1950	1972	1974	2003	2008	10.0	11.1	0.099	-7.72	2	0
PRT	1950	1957	1958	1967	1972	16.2	21.9	0.116	-0.24	1	0
PRT	1950	1973	1975	1999	2008	9.1	11.9	0.188	-11.46	2	0
PRY	1980	1989	2002	2002	2008	8.8	8.7	0.046	-14.24	13	1
RWA	1960	1993	1994	1997	2008	18.0	7.8	0.000	-45.38	1	0
SAU	1986	1992	1999	2002	2008	14.6	13.7	0.039	-18.75	7	0
SLE	1961	1974	1977	1989	1994	9.6	12.3	0.132	-6.74	3	0
SLE	1961	1995	1999	2006	2008	14.2	10.5	0.009	-41.65	4	1
SLV	1950	1978	1983	1987	2008	18.2	10.8	0.000	-25.82	5	0
TGO	1960	1979	2008	1989	2008	9.6	9.8	0.055	-53.60	29	1
THA	1950	1996	1998	2003	2008	10.7	7.7	0.007	-14.17	2	0
TTO	1950	1961	1963	1969	1981	16.8	14.4	0.028	-0.78	2	0
TTO	1950	1982	1993	2006	2008	12.4	13.3	0.075	-28.96	11	0
TUN	1961	1981	1988	1988	2008	8.5	10.1	0.099	-5.89	7	0
UGA	1950	1977	1986	1987	2008	11.6	11.1	0.044	-30.27	9	0
USA	1950	1957	1958	1966	2008	8.7	9.2	0.069	-2.51	1	0
VEN	1950	1979	2003	2003	2008	8.6	10.3	0.130	-35.49	24	1
ZAF	1950	1957	1958	1970	1979	20.5	14.1	0.012	-1.22	1	0
ZAF	1950	1980	1992	1992	2008	9.3	12.1	0.174	-17.16	12	0
ZMB	1955	1968	2001	2000	2008	15.0	11.1	0.010	-68.99	33	1

[a] Out of a total of 100 episodes identified by the sequential algorithm, 14 are not actually slumps. (They satisfy the coefficient restrictions with the AR(p) terms included but not without, so that there is no observed decline in GDP per capita; see footnote 8 in the main text). These discarded episodes are [country code (spell number)]: AUT (1), AUT (2), CHN (1), FIN (1), FRA (1), HKG (1), IDN (1), IRN (1), MRT (1), PRY (1), TZA (1).

[b] The smaller sample of 59 slumps based on a nominal size of 10% is a strict subset of the breaks listed above. However, it is not exactly equal to simply removing all episodes with a p-value exceeding 0.1 in the table above. The reason is that a second or third slump which is significant at the stricter level may have only been found after another slump was identified in the same country at the 20% threshold. Eliminating these five "orphaned" episodes yields the correct list.

[c] These episodes have been estimated sequentially (by splitting the sample before the first break and after the second break) but are reported according to their sequence in calendar time. To see this, consider the case of Cyprus. The first estimation runs from 1950 until 2008 and finds two breaks in 1973 and 1977. The second estimation then runs from 1950 until 1972 and finds two breaks in 1957 and 1964 but these precede the earlier breaks, so they are reported first.

B Estimation of structural breaks

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:

$$y_t = \alpha + \beta t + \gamma_0 \mathbf{1}(t > t_{b1}) + \gamma_1 (t - t_{b1}) \mathbf{1}(t > t_{b1}) + \gamma_2 (t - t_{b2}) \mathbf{1}(t > t_{b2}) + \sum_{i=1}^{p} \delta_i y_{t-i} + \epsilon_t$$

   where y_t is the log of GDP per capita in year t, t_{bi} 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 $t_{b2} \geq t_{b1} + h$ 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 $t_{b1} \in [\tau T, (1-\tau)T - h]$ and $t_{b2} \in [\tau T + h, (1-\tau)T]$. We set $\tau = 0.05$. Let Λ_τ denote the set of all possible episodes $[t_{b1}, t_{b2}] \subset [\tau T, (1-\tau)T]$.[1]

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:

$$\sup_{[t_{b1}, t_{b2}] \in \Lambda_\tau} W(t_{b1}, t_{b2}) = \sup_{[t_{b1}, t_{b2}] \in \Lambda_\tau} \left(\frac{T-K}{3} \right) \frac{SSR^r - SSR^u}{SSR^u}$$

   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 (as described on the next page).

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.

[1] 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.

Bootstrapping the sup-Wald statistic

There have been several suggestions on how to best bootstrap structural change tests. 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, but we reestimate our preferred specification based on breaks obtained with different techniques (see Table F-1 of this appendix).

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. $\mathcal{N}(0, \hat{\sigma}^2_{\hat{e}})$

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 $\sup_{[t_{b1}^*, t_{b2}^*]\in\Lambda_\tau} W^*_j$, 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.

$$\hat{p}^* = \frac{1}{B}\sum_{j=1}^{B} \mathbf{1}\left(\sup_{[t_{b1}^*, t_{b2}^*]\in\Lambda_\tau} W^*_j > \sup_{[t_{b1}, t_{b2}]\in\Lambda_\tau} W(t_{b1}, t_{b2}) \right)$$

   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).

Comparison with Berg et al. (2012)

Berg et al. (2012) employ a similar approach to ours in their paper on sustained growth. They use an unrestricted variant of the Bai and Perron (2003) structural break algorithm and then classify these episodes ex post. Berg et al. (2012) define a positive growth episode as beginning with an upbreak followed by a period of at least 2% average growth and ending with a statistical downbreak (and less than 2% average growth thereafter) or the end of the sample. This procedure identifies 140 upbreaks and 140 downbreaks, which together make up 104 positive growth spells after applying the growth rate filter. While this approach works well for identifying growth spurts, it cannot be easily modified to reliably identify recessions. Here we briefly outline some of the problems we encountered in trying to make these adjustments.

An obvious starting point would be to turn around their definition of a positive growth spell. If we require that a slump begins with a downbreak followed by a period of at least −2% average growth and ends with any upbreak or the end of the sample, then we identify 62 potential slumps.[2] This number is surprisingly close to our smaller sample, even though their data ends two years earlier in 2006. *Prima facie* this seems to suggest that we might similarly use their approach instead of our restricted structural change approach. The main problems are that the start dates and economic meaning of the identified episodes do not necessarily correspond to what we define as slumps.

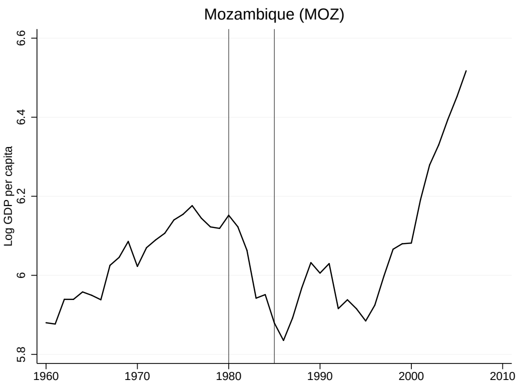
Figure B-1 directly compares three episodes we classified using this method to our break estimates. The first episode in Mozambique matches the definition of a slump very well, even though the second break occurs before the economy is on a continuous recovery path. Our approach identifies the same starting year but also places the second break at a meaningful point. The other examples are less favorable for the unrestricted break search approach. In the Republic of Congo, there is a noticeable slump but the starting date is imprecisely estimated and an earlier, more abrupt crash seems to be missed. Our method puts the first break in 1974, the trough in 1977 and the second break in 1982. Ethiopia is a great example why parameter restrictions matter: it is already declining when the downbreak occurs. Berg et al. (2012) identify a downbreak in 1982 followed by an upbreak in 1987, whereas we identify the start in 1972, the trough in 1992, and the second break in 1993. We have experimented with other definitions requiring only downbreaks and mildly negative growth, but then we quickly catch more slowdowns in growth which are not really recessions. If we remove the questionable episodes then the sample sizes quickly become too small, suggesting we are missing important slumps.

There are other conceptual issues which lead us to prefer the restricted structural change approach. Imposing a specific growth rate after the break treats all time series alike in an absolute sense, although a deep slump in a stable country like Germany might look very different than that of a volatile developing country. A key advantage of our approach is that it superimposes the desired structure onto the time series but does not require auxiliary economic criteria. This matters since break years that do not conform to the structure are dismissed, in favor of the next potential break year. Being agnostic about the break structure implies that any type of break is recorded first and needs to be classified afterwards. This has the undesirable effect that a strong recovery might mask a short initial slump, or a downbreak might occur when a country is already experiencing an economic decline.

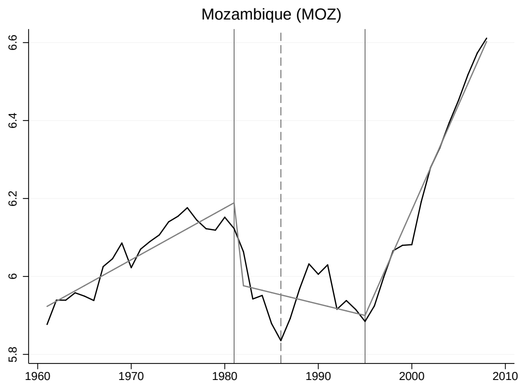
[2] We are grateful to Andrew Berg for providing us with the data and an anonymous referee for suggesting this direct comparison.

Figure B-1 – Comparing unrestricted and restricted structural change approaches

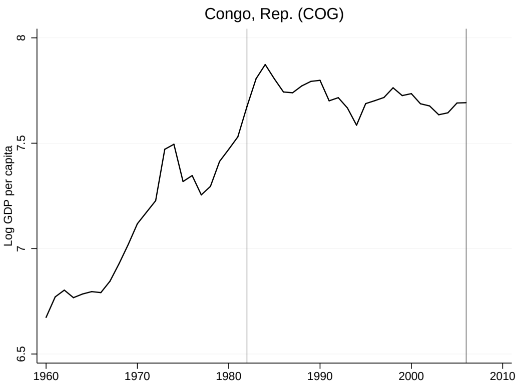
(a) Inverted Berg et al. (2012)



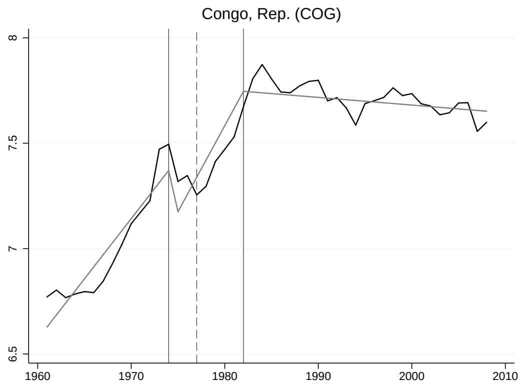
(b) Our breaks



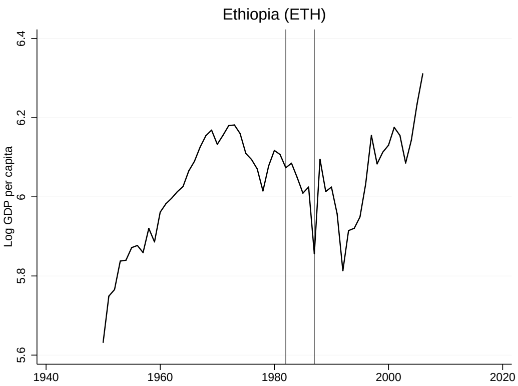
(c) Inverted Berg et al. (2012)



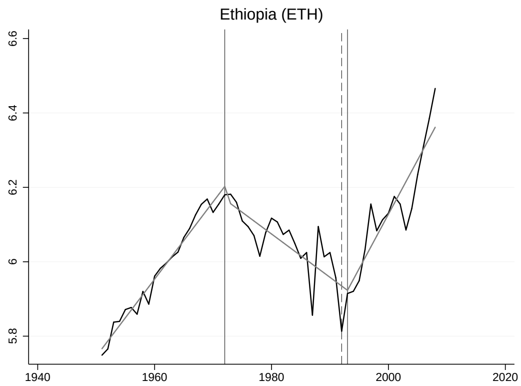
(d) Our breaks



(e) Inverted Berg et al. (2012)



(f) Our breaks



C Variables and summary statistics

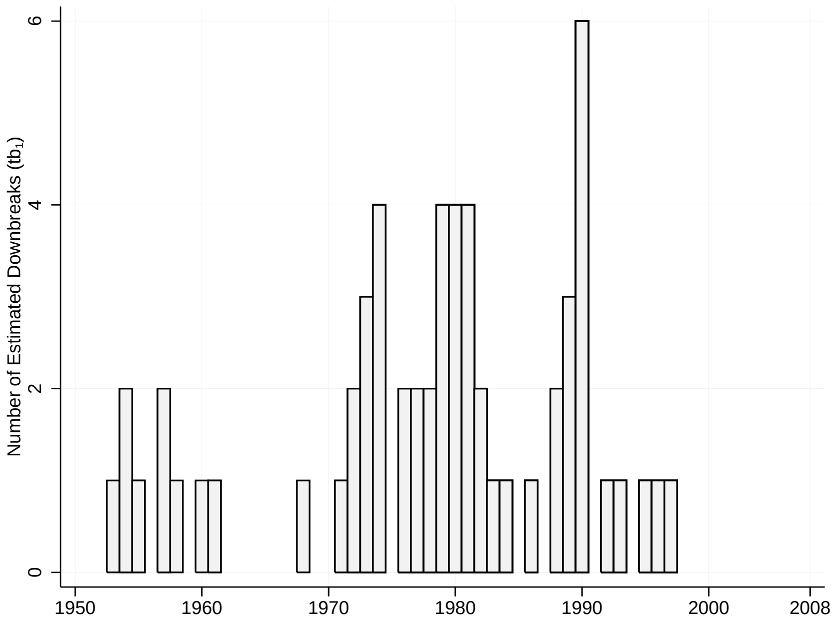
Table C-1 – Sources and summary statistics: pre-slump values, larger sample

VARIABLE	Mean	Std. Dev.	$N \times T$	Source
Polity IV Score	-1.12	7.82	465	Polity IV
Democracy Score	3.42	4.29	465	Polity IV
Autocracy Score	4.54	3.69	465	Polity IV
Executive Recruitment	5.19	2.40	465	Polity IV
Executive Constraints (INS_0)	3.60	2.51	465	Polity IV
Political Competition	4.23	3.79	465	Polity IV
Regime Durability	25.05	27.98	471	Polity IV
Fractionalization ($ELF15$)	47.87	33.88	482	Desmet et al. (2012)
Real US Interest Rate[a]	0.94	2.70	482	FRED
Initial log GDP[b]	16.31	1.73	482	PWT 7.0
Policy Volatility[c]	7.79	6.10	482	PWT 7.0
Government Size	10.12	6.63	482	PWT 7.0
Investment Share	24.30	12.74	482	PWT 7.0
Inflation ($\ln[1+\delta]$)[d]	11.92	22.60	326	WDI/IFS
RER Underval[e]	-0.07	0.45	482	PWT 7.0
Trade Openness (de facto)	62.21	36.65	482	PWT 7.0
Investment Price	82.89	57.80	482	PWT 7.0
Trade Openness (de jure)	0.26	0.44	439	Wacziarg and Welch (2008)
Leader Exit	0.47	0.50	471	Goemans et al. (2009)
War/Conflict	0.16	0.37	482	Gleditsch et al. (2002)
Life Expectancy[f]	60.27	11.77	482	World Population Prospects
Education[g]	4.44	3.01	459	Barro and Lee (2013)

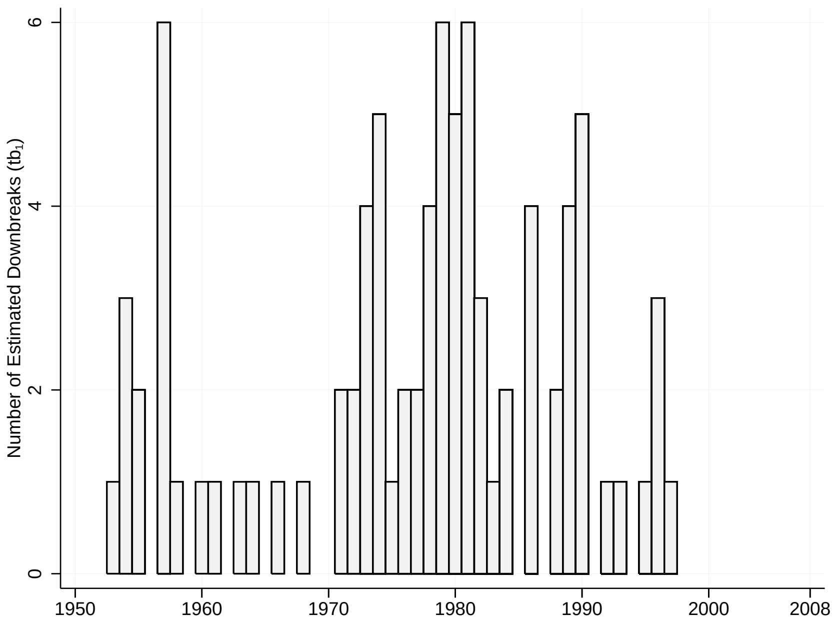
[a] Deflated three months treasury bill rate.
[b] Initial refers to the first observed GDP value in the Penn World Tables.
[c] Following Fatás and Mihov (2013), we first estimate country-level regressions of the form $\Delta \ln G_{it} = \alpha_i + \beta_i \Delta \ln Y_{it} + \varepsilon_{it}$ where G_{it} is real government consumption and Y_{it} is real GDP, and then compute (rolling) standard deviations of the estimated ε_{it} for each country. This procedure nets out cyclical variation in government spending.
[d] We prefer inflation data from the WDI and supplement missing values with data from the IFS.
[e] Following Rodrik (2008), we compute the index of real exchange rate undervaluation as the difference between $\ln RER_{it} - \ln \widehat{RER}_{it}$, where $\ln RER_{it} = \ln(XRAT_{it}/PPP_{it})$ and $\ln \widehat{RER}_{it} = \alpha + \beta \ln y_{it} + \lambda_t + e_{it}$ from a pooled panel regression. Here $XRAT_{it}$ is the exchange rate and PPP_{it} the purchasing power parity, both expressed in LCU per US dollar, y_{it} is real GDP per capita, and the λ_t are time effects.
[f] 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).
[g] Converted into annual data by linear interpolation.

Figure C-1 – Distribution of starting dates

(a) Structural breaks estimated with size = 0.10



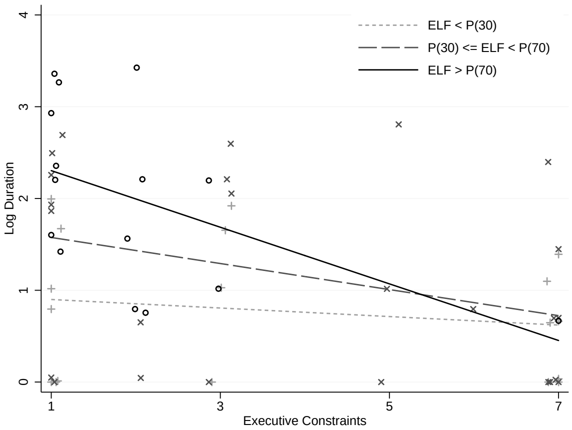
(b) Structural breaks estimated with size = 0.20



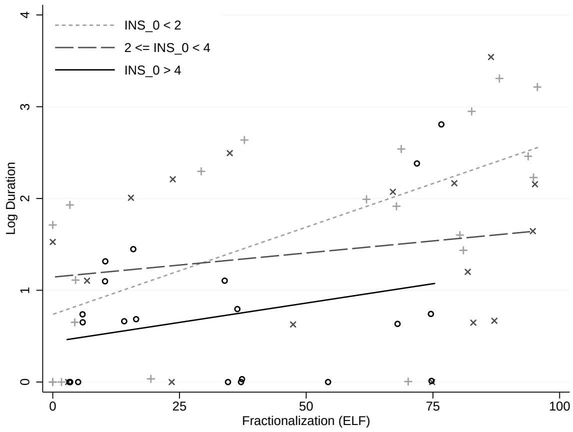
D Split sample plots

Figure D-1 – Linear associations in the raw data, split by categories of ...

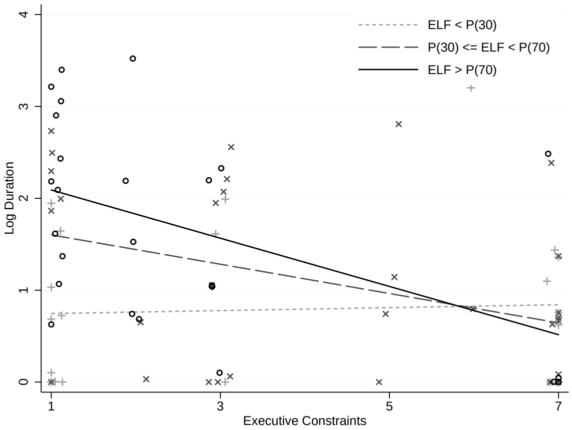
(a) Fractionalization, size = 0.10



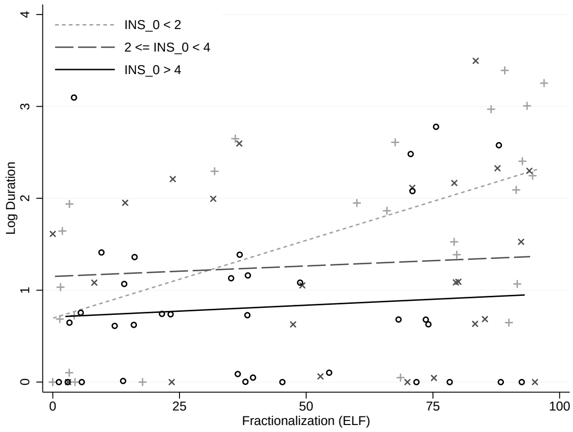
(b) Executive Constraints, size = 0.10



(c) Fractionalization, size = 0.20

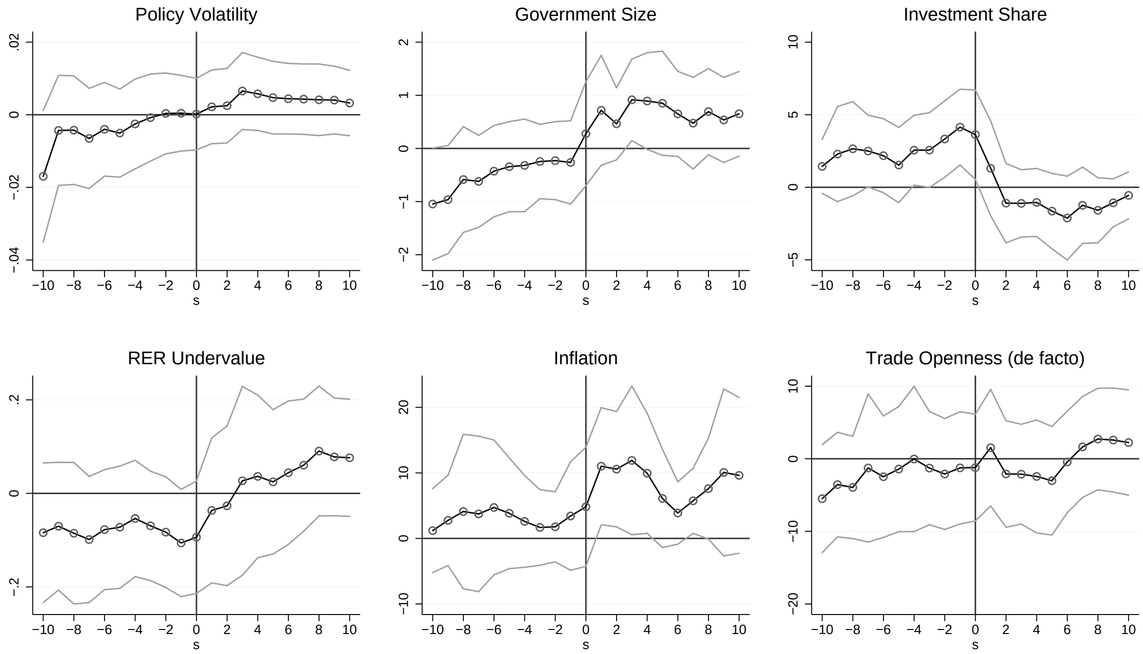


(d) Executive Constraints, size = 0.20



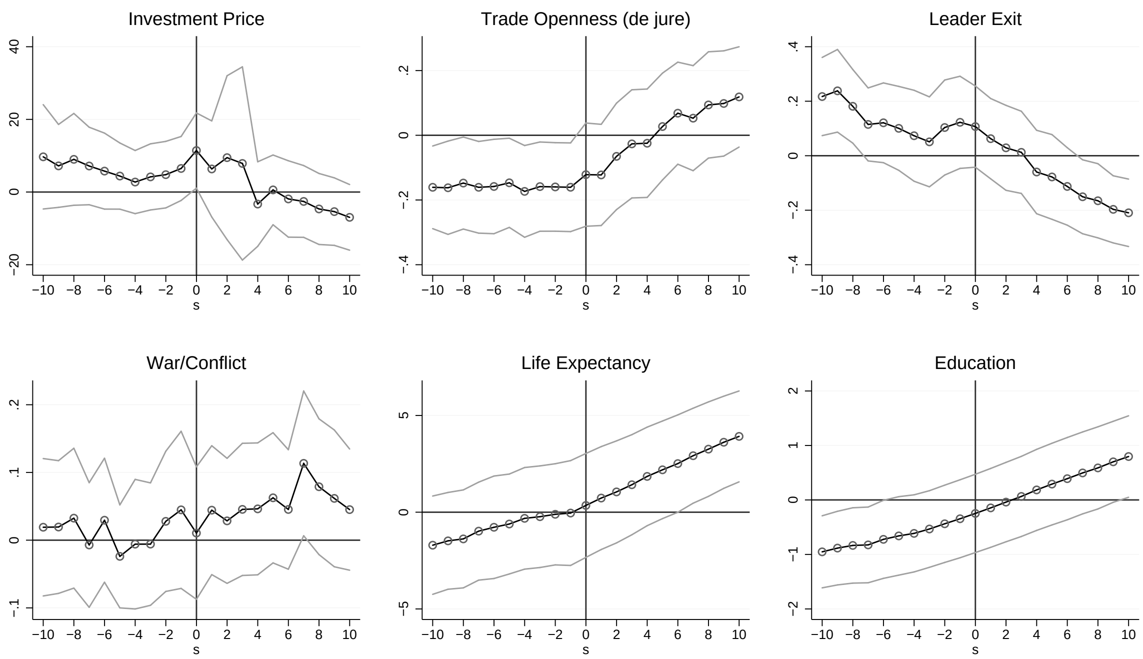
E Additional event plots

Figure E-1 – Additional variables I, size = 0.10



Notes: Illustration of the behavior of different covariates around the start of the slump according to the method described in the main text. Results for breaks with size = 0.20 are very similar (not reported).

Figure E-2 – Additional variables II, size = 0.10



Notes: Illustration of the behavior of different covariates around the start of the slump according to the method described in the main text. Results for breaks with size = 0.20 are very similar (not reported).

F Additional regression results

Varying break search parameters

We reestimate the breaks and our preferred specification several times in order to verify that our main results are robust to a departure from our preferred parameters.

Table F-1 reports the results from varying the bootstrapping technique. Column (1) repeats the estimates from the parametric bootstrap used throughout the paper for reference. Columns (2) and (3) use the same parametric bootstrap but with t-distributions and various degrees of freedom in order to account for excess mass in the tails. The remaining columns use different non-parametric bootstraps which do not require distributional assumptions. Column (4) resamples from the observed (rescaled) residuals. Column (5) uses Hansen's heteroskedastic fixed design bootstrap where the sample is not build recursively and is based on the optimal model with breaks, not the null model. Column (6) uses a Wild bootstrap where the bootstrap samples are again build recursively but the resampled and rescaled residuals are multiplied by Rademacher's two-point distribution.

Table F-2 varies the minimum period required before the restricted break search algorithm can find a second break.

Finally, Table F-3 extends the time series until 2014 and reports results for GDP per capita and GDP from the PWT version 9 and the World Development Indicators.

Table F-1 – Different bootstrap techniques for structural breaks

	Parametric			*Semi-parametric*		
	Normal (1)	$t(n-k)$ (2)	$t(5)$ (3)	Residual (4)	Hansen FD (5)	Wild (6)
	Panel a) Structural breaks estimated with size = 0.10					
Executive Constraints ($\widetilde{INS_0}$)	-0.217***	-0.205***	-0.188***	-0.270***	-0.221***	-0.201***
	(0.064)	(0.058)	(0.066)	(0.073)	(0.044)	(0.067)
Fractionalization ($\widetilde{ELF}$)	0.018***	0.016***	0.015***	0.018***	0.017***	0.002
	(0.004)	(0.004)	(0.005)	(0.006)	(0.004)	(0.006)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003**	-0.004***	-0.002*	-0.003*	-0.003***	-0.004*
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Exits	48	51	46	33	67	30
Spells	58	62	54	42	81	37
	Panel b) Structural breaks estimated with size = 0.20					
Executive Constraints ($\widetilde{INS_0}$)	-0.190***	-0.184***	-0.176***	-0.207***	-0.157***	-0.249***
	(0.046)	(0.046)	(0.046)	(0.049)	(0.043)	(0.063)
Fractionalization ($\widetilde{ELF}$)	0.013***	0.013***	0.012***	0.017***	0.011***	0.014***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003***	-0.003**	-0.003**	-0.004***	-0.003**	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Exits	69	78	71	54	88	45
Spells	81	91	82	65	103	55

Notes: All models include region FEs, the real US interest rate, initial GDP, and a constant. The standard errors are clustered at the country level to account for repeated spells. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table F-2 – Different interbreak periods for structural breaks

	Interbreak period (h)					
	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$	$h = 7$
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel a) Structural breaks estimated with size = 0.10					
Executive Constraints ($\widetilde{INS_0}$)	-0.218***	-0.227***	-0.217***	-0.204***	-0.217***	-0.233***
	(0.055)	(0.068)	(0.064)	(0.065)	(0.067)	(0.067)
Fractionalization ($\widetilde{ELF}$)	0.018***	0.019***	0.018***	0.018***	0.017***	0.017***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.004***	-0.003**	-0.003**	-0.003**	-0.003**	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Exits	54	46	48	48	46	47
Spells	65	56	58	57	55	56
	Panel b) Structural breaks estimated with size = 0.20					
Executive Constraints ($\widetilde{INS_0}$)	-0.175***	-0.204***	-0.190***	-0.187***	-0.173***	-0.178***
	(0.043)	(0.047)	(0.046)	(0.046)	(0.046)	(0.045)
Fractionalization ($\widetilde{ELF}$)	0.014***	0.016***	0.013***	0.014***	0.012***	0.013***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003***	-0.004***	-0.003***	-0.003***	-0.003**	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Exits	79	70	69	70	70	70
Spells	92	82	81	82	82	82

Notes: All models include region FEs, the real US interest rate, initial GDP, and a constant. The standard errors are clustered at the country level to account for repeated spells. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table F-3 – Longer time series, breaks until 2014

	PWT9		*WDI*	
	rgdpna/pop	*rgdpna*	*gdplcu/pop*	*gdplcu*
	(1)	(2)	(3)	(4)
	Panel a) Structural breaks estimated with size = 0.10			
Executive Constraints ($\widetilde{INS_0}$)	-0.315***	-0.219***	-0.215***	-0.129***
	(0.086)	(0.045)	(0.072)	(0.047)
Fractionalization ($\widetilde{ELF}$)	0.018***	0.010***	0.013***	0.008***
	(0.006)	(0.002)	(0.005)	(0.003)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.004**	-0.004***	-0.005***	-0.004***
	(0.002)	(0.001)	(0.002)	(0.001)
Exits	44	60	44	46
Spells	56	61	56	47
	Panel b) Structural breaks estimated with size = 0.20			
Executive Constraints ($\widetilde{INS_0}$)	-0.268***	-0.111**	-0.152***	-0.068
	(0.065)	(0.044)	(0.055)	(0.044)
Fractionalization ($\widetilde{ELF}$)	0.014**	0.006**	0.013***	0.008***
	(0.005)	(0.003)	(0.004)	(0.002)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.004**	-0.003***	-0.003**	-0.002*
	(0.002)	(0.001)	(0.001)	(0.001)
Exits	57	75	58	64
Spells	70	76	71	65

Notes: All models include region FEs, the real US interest rate, initial GDP, and a constant. The standard errors are clustered at the country level to account for repeated spells. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Right hand side variations

In Table F-4 we vary the measure of political institutions. Broader measures of institutions are correlated with the length of declines but we find no evidence of an interaction with ethnic heterogeneity. The Polity IV database follows a clear hierarchy in coding institutional features. Executive constraints, executive recruitment and political competition make up the democracy and autocracy scores. These in turn are summarized in the (revised combined) Polity score. Regime durability simply tracks how long a particular regime has been in place. Columns (1) and (2) show that there is a much weaker relationship between these other two institutional dimensions and the duration of declines. The remaining columns add that this also severely weakens the evidence in favor of an interaction effect in the aggregate indicators. We interpret this as indirect evidence that constraints on the executive are a particularly relevant institutional feature, which is also in line with our preferred interpretation. Executive recruitment and the degree of political competition do not create the kind of commitment problem described here and elsewhere. Bluhm and Thomsson (2015) show that the findings are robust to choosing different measures of group heterogeneity and inequality.

Table F-5 reports the results from adding additional policy variables and other determinants. As before, none of these effects are competing with our main results. We have also added variables in groups. Our specifications are robust to including different sets of variables at the same time (e.g. policy, volatility, government size, investment share, and investment price). There are no substantive changes, so that the results are not reported here but are available on request. Note that we do not have enough degrees of freedom to include our variables of interest, the region fixed effects, and all other variables at once.

Table F-4 – Different measures of institutions

	Measure of political institutions (X_0)					
	Executive Recruitment (1)	Political Competition (2)	Autocracy (3)	Democracy (4)	Polity Score (5)	Regime Durability (6)
	Panel a) Structural breaks estimated with size = 0.10					
Political Institutions ($\widetilde{X_0}$)	-0.077	-0.097**	0.093**	-0.094***	-0.048**	0.005
	(0.067)	(0.048)	(0.045)	(0.036)	(0.020)	(0.008)
Fractionalization ($\widetilde{ELF}$)	0.015***	0.016***	0.016***	0.016***	0.016***	0.017***
	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
$\widetilde{X_0} \times \widetilde{ELF}$	0.001	-0.001	0.001	-0.001	-0.000	0.000
	(0.002)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Exits	48	48	48	48	48	48
Spells	58	58	58	58	58	58
	Panel b) Structural breaks estimated with size = 0.20					
Political Institutions ($\widetilde{X_0}$)	-0.095**	-0.086*	0.102***	-0.086***	-0.049***	0.004
	(0.047)	(0.049)	(0.035)	(0.028)	(0.016)	(0.007)
Fractionalization ($\widetilde{ELF}$)	0.011***	0.011***	0.012***	0.012***	0.012***	0.011**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$\widetilde{X_0} \times \widetilde{ELF}$	-0.001	-0.001	0.002	-0.001*	-0.001*	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Exits	69	69	69	69	69	70
Spells	81	81	81	81	81	82

Notes: All models include region FEs, the real US interest rate, initial GDP, and a constant. The standard errors are clustered at the country level to account for repeated spells. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table F-5 – Additional policy variables and other determinants

	Added variable					
	Investment Price (1)	Trade Openness (d.j.) (2)	Leader Exit (3)	War/ Conflict (4)	Life Expectancy (5)	Education (6)
	Panel (a) Structural breaks estimated with size = 0.10					
Executive Constraints ($\widetilde{INS_0}$)	-0.215***	-0.212***	-0.204***	-0.233***	-0.262***	-0.272***
	(0.065)	(0.075)	(0.077)	(0.065)	(0.066)	(0.075)
Fractionalization ($\widetilde{ELF}$)	0.018***	0.018***	0.019***	0.019***	0.019***	0.020***
	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003**	-0.004**	-0.003**	-0.003**	-0.003***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Added Variable	0.002	-0.224	0.043	-0.385	0.043***	0.103
	(0.002)	(0.295)	(0.298)	(0.451)	(0.010)	(0.072)
Exits	48	43	47	48	48	46
Spells	58	52	57	58	58	56
	Panel b) Structural breaks estimated with size = 0.20					
Executive Constraints ($\widetilde{INS_0}$)	-0.179***	-0.178***	-0.179***	-0.189***	-0.208***	-0.216***
	(0.047)	(0.050)	(0.051)	(0.046)	(0.046)	(0.053)
Fractionalization ($\widetilde{ELF}$)	0.013***	0.013***	0.014***	0.013***	0.015***	0.014***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003**	-0.004***	-0.003***	-0.004***	-0.003***	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Added Variable	0.004**	-0.316	0.116	0.158	0.040***	0.096
	(0.002)	(0.257)	(0.261)	(0.375)	(0.011)	(0.062)
Exits	69	63	68	69	69	67
Spells	81	74	80	81	81	79

Notes: All models include region FEs, the real US interest rate, initial GDP, and a constant. The standard errors are clustered at the country level to account for repeated spells. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Functional form

Table F-6 investigates whether the log-normal is a good choice for the hazard function. Column (1) reports the reference results obtained with the log-normal. Column (2) uses a log-logistic hazard instead. The estimated shape parameter is negative, implying that the hazard is first increasing then decreasing as in the log-normal model. Column (3) is the exponential or constant hazard model. Column (4) uses a Weibull parameterization which allows for monotonically increasing or decreasing hazard rates. The estimated shape parameter suggests that the baseline hazard is increasing over time. In contrast, the Gompertz model in column (5) suggests a shape that is monotonically decreasing. 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.

Table F-6 – Varying the functional form

	Coefficients		*Hazard Ratios* ($\mathbb{H}_0$: HR = 1)			
	Survival model					
	Log-normal (1)	Log-logistic (2)	Exponential (3)	Weibull (4)	Gompertz (5)	Cox PH (6)
	Panel a) Structural breaks estimated with size = 0.10					
Executive Constraints ($\widetilde{INS_0}$)	-0.217***	-0.253***	1.350***	1.468***	1.388***	1.348***
	(0.064)	(0.061)	(0.092)	(0.139)	(0.111)	(0.104)
Fractionalization ($\widetilde{ELF}$)	0.018***	0.016***	0.977***	0.971***	0.974***	0.982***
	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003**	-0.004***	1.004**	1.005**	1.005**	1.004**
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Shape Parameter	-0.181**	-0.621***		1.415***	1.040	
	(0.073)	(0.096)		(0.121)	(0.036)	
Exits	48	48	48	48	48	48
Spells	58	58	58	58	58	58
Years of Decline	353	353	353	353	353	353
Log-$\mathcal{L}$	-67.026	-72.393	-71.616	-67.704	-71.104	-144.483
AIC	156.052	158.786	163.232	157.407	164.207	298.967
	Panel b) Structural breaks estimated with size = 0.20					
Executive Constraints ($\widetilde{INS_0}$)	-0.190***	-0.209***	1.231***	1.258***	1.220***	1.212***
	(0.046)	(0.051)	(0.060)	(0.069)	(0.057)	(0.062)
Fractionalization ($\widetilde{ELF}$)	0.013***	0.011***	0.984***	0.983***	0.985***	0.990***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
$\widetilde{INS_0} \times \widetilde{ELF}$	-0.003***	-0.003***	1.005***	1.006***	1.005***	1.004**
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Shape Parameter	-0.078	-0.564***		1.179**	0.976	
	(0.077)	(0.087)		(0.086)	(0.026)	
Exits	69	69	69	69	69	69
Spells	81	81	81	81	81	81
Years of Decline	465	465	465	465	465	465
Log-$\mathcal{L}$	-103.848	-108.301	-110.155	-108.698	-109.794	-239.955
AIC	229.697	230.602	240.310	239.397	241.588	489.909

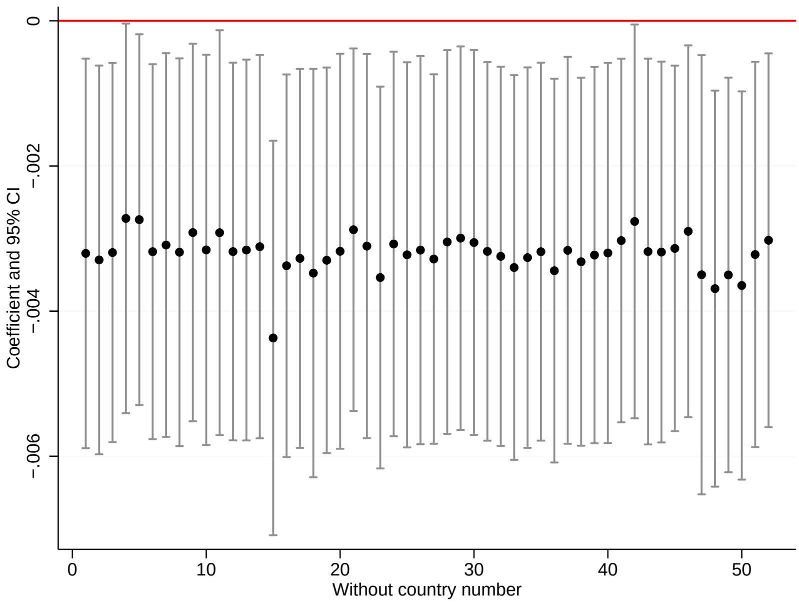
Notes: All models include region FEs, the real US interest rate, initial GDP, and a constant. The standard errors are clustered at the country level to account for repeated spells. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Outliers

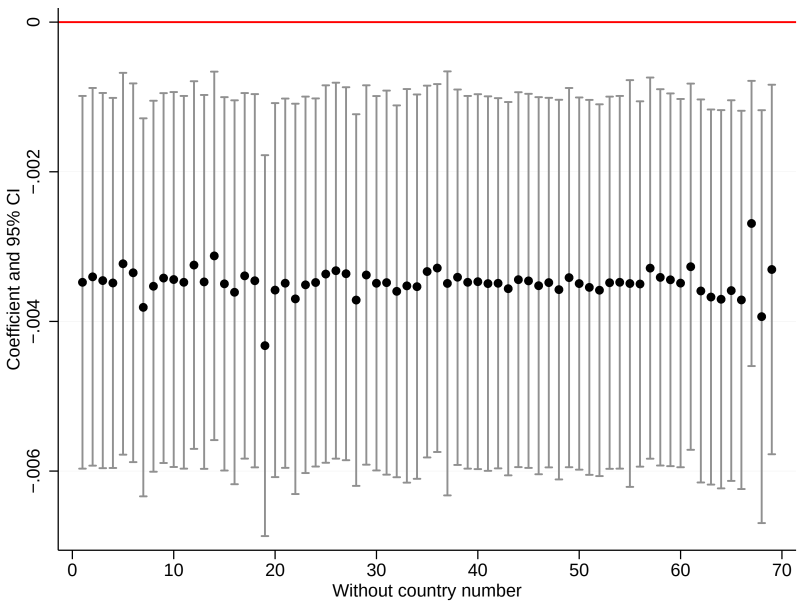
Figure F-1 establishes that no single country is driving our results. It takes our preferred specification with region fixed effects and solely focuses on the interaction coefficient. Each time we reestimate the model leaving out one country and then plot the estimated coefficient including a 95% confidence interval.

Figure F-1 – Leave one out graph of the interaction coefficient

(a) Structural breaks estimated with size = 0.10



(b) Structural breaks estimated with size = 0.20



Notes: Illustration of the robustness of the interaction effect to dropping individual countries. All underlying models include region FEs, the real US interest rate, initial GDP, and a constant. The standard errors are clustered at the country level to account for repeated spells.

Additional references

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.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*(1), 1–22.

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.

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

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.

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.

Fatás, A. and I. Mihov (2013). Policy volatility, institutions, and economic growth. *Review of Economics and Statistics 95*(2), 362–376.

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.

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

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.

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

Rodrik, D. (2008). The real exchange rate and economic growth. *Brookings Papers on Economic Activity 2008*(2), 365–412.

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