

Do Weak Institutions Prolong Crises?

On the Identification, Characteristics, and Duration of Declines During Economic Slumps

Richard Bluhm

Denis de Crombrughe

Adam Szirmai



WORLD BANK GROUP

Development Economics

Knowledge and Strategy Team

January 2020

Abstract

This paper studies periods of prolonged contractions in output per capita in a sample of 145 countries from 1950 to 2014. Economic slumps are defined as abrupt interruptions of a period of growth by several regime switches. Slumps start with a sharp contraction along with a trend break, which is followed by another switch when growth stabilizes again. The paper then analyzes the correlates of these slumps, focusing on the length and depth of the contraction, from the beginning of the slump to its trough. The results establish three new stylized facts: (i) weak political

institutions predate crises whereas political reforms tend to follow them, (ii) the length and depth of economic declines are robustly correlated with executive constraints and ethnic heterogeneity, and (iii) there is a robust interaction between these two variables, suggesting that institutions constraining leaders are important for stabilizing growth. This is particularly relevant for Sub-Saharan Africa, where politics are often ethnic and decision makers are comparatively unconstrained.

This paper is a product of the Knowledge and Strategy Team, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at bluhm@mak.uni-hannover.de.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Do Weak Institutions Prolong Crises?

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

Richard Bluhm

Denis de Crombrughe

Adam Szirmai

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

JEL codes: O43, O11, C41, F43

Richard Bluhm (corresponding author) is an assistant professor at Leibniz University Hannover, Germany; email: bluhm@mak.uni-hannover.de. Denis de Crombrughe is an associate professor at Maastricht University, the Netherlands; email: d.decrombrughe@maastrichtuniversity.nl. Adam Szirmai is a professorial fellow at UNU-MERIT and professor at Maastricht University, the Netherlands; email: szirmai@merit.unu.edu. This paper has been presented at ABCDE, AFD Paris, DIAL, EEA-ESEM, ETH Zurich, Duke Fuqua, Hertie School/DIE Workshop, Ifo Workshop on Political Economy, Nederlandse Economendag, NIPF-DEA, Silvaplana Workshop in Political Economy, and Warwick Workshop on Economic Growth. We have greatly benefited from comments by Oded Galor, Martin Gassebner, Melanie Krause, Nicolas Meisel, Thomas Roca, Ajay Shah, Jan-Egbert Sturm, Kaj Thomsson, and Bart Verspagen. We would also like to thank Aart Kraay and three anonymous referees for their valuable suggestions and help in improving the paper. We gratefully acknowledge financial support from the Agence Française de Développement (AFD).

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; several African nations went from ‘up and coming’ in the 1950s to long periods of stagnation between 1973 and 2000; and Latin America, long characterized by political turmoil and economic volatility, experienced a lost decade in the 1980s. Many economic slumps in developing countries over this period were considerably deeper and longer than even the ‘Great Depression’ of the 1930s. There is also a long list of relatively short-lived advanced economy crises. What can we learn from such abrupt reversals of growth, and why do declines last so much longer in some countries than in others?

The instability of growth is of great concern in economics. It has been suggested that it not only affects output in the short run (e.g. [Ramey and Ramey, 1995](#)), but that curbing volatility also plays 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](#), [Szirmai and Foster-McGregor, 2017](#)). 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](#)). Unstable growth has serious welfare implications ([Pritchett et al., 2016](#)). However, the implications of negative regime switches are only beginning to be explored. Long-lasting slumps can nullify decades of positive growth, with little hope 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 an economic slump: a positive growth regime is interrupted by a sharp downward shift and a trend break, followed by a second trend break when growth resumes (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 that the crisis itself may 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 seem able to explain 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 that 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) decompose 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. They 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 play 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, our analysis suggests that controlling inflation could reduce the duration of declines. 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 episodes 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 classify smaller fluctuations in tranquil times as important.² An unappreciated weakness of standard

¹See, 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.

²By contrast, uniform economic criteria would treat all series as if they were generated by the same

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. Their 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 2014.

The restricted structural change approach has another notable advantage: it can take into account meaningful restrictions on the growth process before and after the desired episode. 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 upward break (‘upbreak’) first, it is likely to miss the downward break (‘downbreak’) which follows too soon. The appropriately restricted approach would 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 approach might register

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.

³Since 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’.

only the second downturn, 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. [Supplementary Appendix S3](#) (available at *The World Bank Economic Review* website) discusses the results obtained from identifying economic slumps by ‘inverting’ the episodes of sustained growth from [Berg et al. \(2012\)](#), who apply a variant of the Bai-Perron approach, and compares them to those identified by our two-break model.

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 at time t , 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 there is an instantaneous drop at the start of the slump. The other coefficients are left unrestricted, allowing the model to 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 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 seems to offer few additional benefits while adding layers of complexity.⁵

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 hypothesis at the desired significance level, we

⁴An 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 Madagascar in Figure 1d would then not qualify as a slump.

⁵We 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 restrictions are an alternative but then require more ex post classification in return (see the comparison in Supplementary Appendix S3).

record the break pair $(\hat{t}_{b1}, \hat{t}_{b2})$ 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% respectively, so as to arbitrate between committing type I and type II errors. Using the 20% threshold increases the risk of detecting more false positives and, hence, primarily serves as a robustness check. The process starts again on each subsample until the bootstrap test fails to reject the null hypothesis of no breaks or the sample gets too small ($T \leq 20$).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

Before we can define the duration of the decline phase, we need to establish when a slump has finished. A slump starts with the first break year but does not necessarily end with the second break at t_{b1} . 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 not totally over until the level of GDP per capita has at least 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 $t_c > \hat{t}_{b1}$ where $y_{t_c} \geq y_{\hat{t}_{b1}}$.

Within a slump, we identify the trough separating the decline from the recovery phase, and then focus exclusively 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

⁶Supplementary Appendix S3 provides a formal description of the break search algorithm and the bootstrap.

(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 occurs without a change in the growth regime.

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 \hat{t}_{b1} . To cover all cases, we estimate the trough to have occurred at time:

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

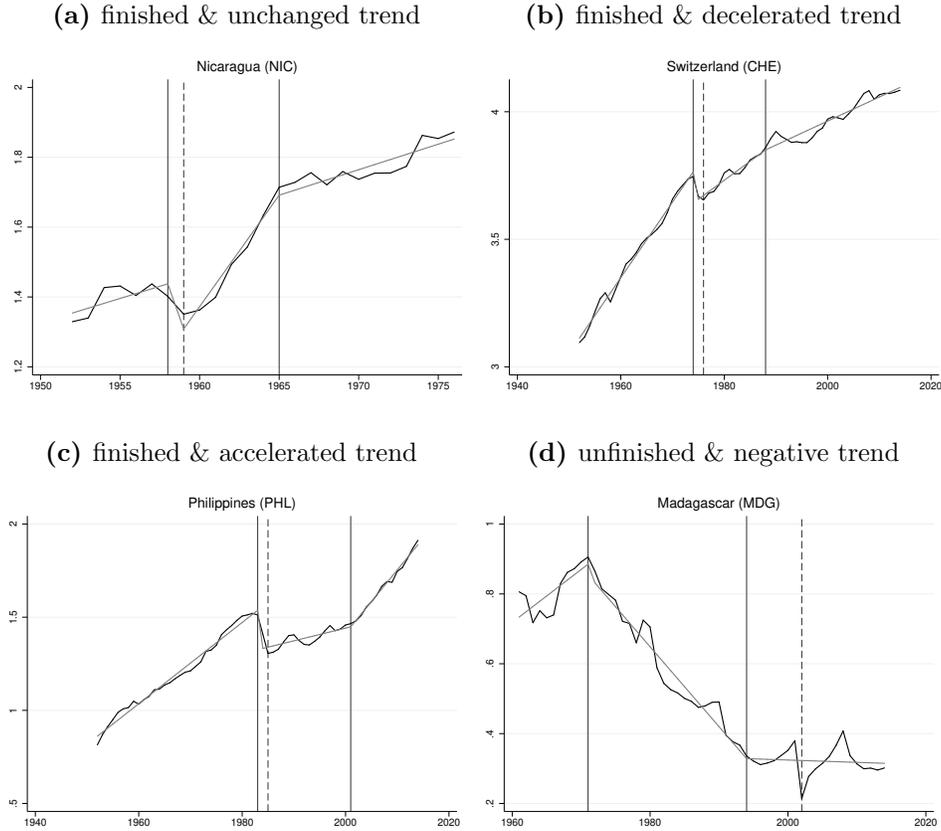
We denote the duration of the contraction phase—lasting from the initial break (\hat{t}_{b1}) to the estimated trough (\hat{t}_{min})—by $\tilde{t}_D = \hat{t}_{min} - \hat{t}_{b1}$.

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 \hat{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 \tilde{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](#) illustrates the diversity of slumps identified by this method. Panel (a) shows a finished slump in Nicaragua where trend growth is nearly unchanged after the second break.

⁷If 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.

Figure 1 – Four types of slumps



Notes: Models refitted using the estimated breaks \hat{t}_{b1} and \hat{t}_{b2} but without the optimal $AR(p)$ terms to emphasize the trend breaks. The bold vertical lines are at \hat{t}_{b1} and \hat{t}_{b2} , respectively. The dashed vertical line indicates \hat{t}_{min} .
Source: Authors' analysis based on data from the Penn World Tables version 9.0 (series *rgdpna/pop*).

The slump begins in 1958 and the trough is found in 1959. This is a relatively short-lived crisis before the ousting of the Somoza dictatorship ushers in a major slump (and civil war), lasting from 1977 to 1993. Panel (b) shows a finished slump in Switzerland where trend growth decelerates after the second break. In 1974, the Swiss economy is strongly affected by the oil crisis of the mid-1970s, leading to a 8.79% drop in GDP per capita within two years. After a strong recovery, Switzerland enters a low growth regime typical for the high-income economies in Western Europe of the 1980s and 1990s. Panel (c) shows a finished slump in the Philippines occurring in the early 1980s with an accelerated trend after the second break. The Philippines were hit by a combination of external shocks: the second oil crisis, declining export prices, and rising world interest rates, as well as internal shocks caused by a debt

overhang, bail outs and capital flight. The first break in 1983 coincides with the assassination of the opposition leader Benigno S. Aquino, the trough occurs in 1985, and the second break takes place in 2001—a few years before the formal end of the slump. Although the duration of the decline phase is only two years, GDP per capita contracted by 18.63% in this short period of time. Last but not least, panel (d) shows an unfinished slump with a continuing decline in Madagascar. Madagascar grew rapidly for over a decade following independence from France in 1960 but then experienced a dramatic and ongoing collapse from 1971 onward. This roughly coincides with the transition of power from Philibert Tsiranana—Madagascar’s first president—to Didier Ratsiraka who fully seized power in 1975 by means of a coup. By 2002, the average Malagasy was about 50% poorer than in 1971. Madagascar’s GDP per capita did not recover its pre-slump level over a period of 31 years. At the end of the observed period, the decline is still ongoing and the provisional trough coincides with the censoring cutoff in 2014.

3 Characteristics of slumps

We apply the sequential algorithm to the entire Penn World Table (v9.0, series *rgdpna/pop*) yielding 57 or 77 slumps between 1950 and 2014, depending on how permissive we are towards type I errors.⁸ [Supplementary Appendix S2](#) lists all episodes and [Supplementary Appendix S4](#) provides summary statistics.

We detect many 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 and another peak occurs between 1979 and 1981 during the second oil shock and debt crisis of the early 1980s. The period between the 1970s and early 1980s is generally marked by heightened volatility, as has been documented in a number of studies ([Easterly](#)

⁸We run the algorithm on all the PWT (v9.0) countries with a population of at least one million and at least 20 years of data. However, 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, requiring that an actual contraction occurs within the range of the two estimated breaks, otherwise there is no slump.

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

	Mean Depth	Median Depth	Mean Duration	Median Duration	Number of Spells	Censored Spells	Number of Countries
<i>Panel (a) Structural breaks estimated with size = 0.10</i>							
<i>Income Level</i>							
High Income (OECD)	-10.9	-11.1	2.4	2	7	1	29
High Income (Other)	-37.5	-46.0	10.0	11	4	2	12
Upper Middle Income	-26.3	-20.8	8.4	3	13	2	33
Lower Middle Income	-23.6	-19.0	6.3	2	17	2	39
Low Income	-31.9	-32.1	14.0	4	16	4	34
<i>Geographical Region</i>							
Africa	-30.9	-28.1	15.6	8	22	6	44
<i>Northern</i>	-16.1	-18.9	17.1	9	3	0	5
<i>Sub-Saharan</i>	-33.3	-36.2	4.5	2	19	6	39
Americas	-20.7	-18.7	4.9	3	12	1	23
Asia	-28.4	-22.6	4.4	1	14	3	43
Europe	-17.0	-14.5	2.9	2	9	1	35
<i>Total</i>	-26.0	-21.0	10.3	3	57	11	145
<i>Panel (b) Structural breaks estimated with size = 0.20</i>							
<i>Income Level</i>							
High Income (OECD)	-7.5	-4.9	1.9	2	13	1	29
High Income (Other)	-37.5	-46.0	10.0	11	4	2	12
Upper Middle Income	-23.3	-18.9	7.3	2	17	2	33
Lower Middle Income	-23.5	-20.0	7.0	3	22	3	39
Low Income	-31.0	-36.2	14.1	5	21	5	34
<i>Geographical Region</i>							
Africa	-29.2	-28.0	13.9	8	29	7	44
<i>Northern</i>	-16.1	-18.9	14.9	5	3	0	5
<i>Sub-Saharan</i>	-30.7	-32.1	5.7	2	26	7	39
Americas	-22.5	-20.0	7.7	4	14	2	23
Asia	-24.5	-16.6	4.0	1	18	3	43
Europe	-13.1	-9.9	3.0	2	16	1	35
<i>Total</i>	-23.5	-18.9	9.4	3	77	13	145

Notes: 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.

Source: Authors' analysis based on data from the Penn World Tables version 9.0 (series *rgdpna/pop*).

et al., 1993, Rodrik, 1999, Pritchett, 2000, Jones and Olken, 2008). Several collapses follow the post-communist transitions of 1989–1990. We find no slumps starting in the period of the early and mid-2000s, but a handful of slumps begin in 2008.⁹ Note that the global recession of 2008–2009 is still a bit too close to the end of the sample to permit reliable

⁹Supplementary Appendix S4 reports the annual frequency of starting dates for both samples.

break estimation (although we later vary all potentially important parameters influencing the detection likelihood).¹⁰

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. The four longest declines all occur in sub-Saharan Africa, even though the mean duration is only 4.5 years due to a high frequency of short but deep crises. Declines in North African countries were only half as deep as in sub-Saharan Africa, but the median slump lasted almost a decade.¹¹ Declines in Asia have been very deep as well but are generally much shorter.

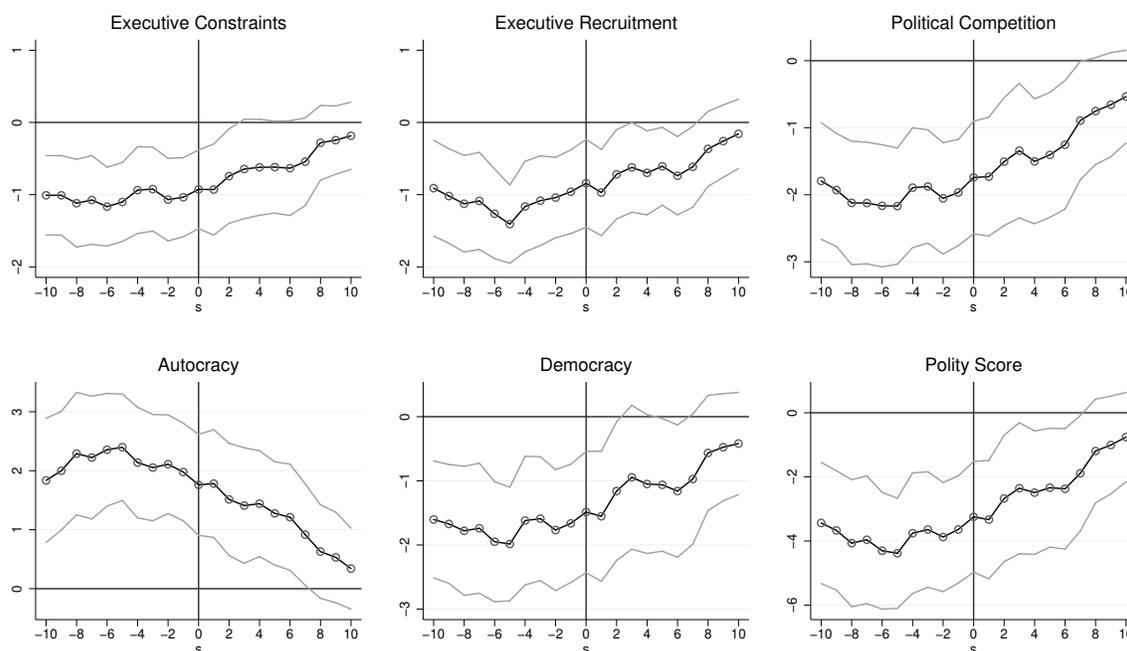
The movement of other variables around the starting date of a slump is another interesting feature of their characteristics. We are especially interested in any evidence of institutional change occurring before, during or after a slump. To study this question descriptively, we employ an event methodology 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.

¹⁰If a slump were to start in 2008, a minimum inter-break period of four years means that the next break could only be detected in 2012. Depending on the trimming parameter and the length of the current time series the algorithm may trim off one to three years at the end of the series, so that detecting a sequence of two breaks might not be feasible without a longer time series.

¹¹Note that we observe only a handful of slumps (or less) in Northern Africa which is why we treat the African continent as a single group of countries in the remainder of the paper.

We run the following regression for each measure of political institutions: $x_{it} = \sum_{s=-10}^{10} \delta_{t, \hat{t}_{b1}+s} \beta_s + \mu_i + \epsilon_{it}$ where $\delta_{t, \hat{t}_{b1}+s}$ is the Kronecker delta which is equal to one if $t = \hat{t}_{b1} + s$ 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 \hat{t}_{b1} . The first year of the slump is $s = 1$ corresponding to $t = \hat{t}_{b1} + 1$. 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.¹²

Figure 2 – Institutions and politics



Source: Authors' analysis based on data from the Polity IV project and structural breaks estimated using the Penn World Tables version 9.0 (series *rgdpna/pop*).

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 gradually return to normal levels thereafter. Figure 2 shows that, ten years before a slump, countries tend to

¹²In this case, ‘normal’ refers to all observations other than the 21 years around the downbreak.

score about one point less on indexes of constraints on the executive or executive recruitment, about two points less (or more) on indexes of political competition, autocracy or democracy, and three to four points less on the combined Polity score.¹³ These sizable differences vanish in the decade after the beginning of a slump. Eight years after the downbreak, we can no longer reject the null hypothesis 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. Another implication is that we have to rule out feedback from crises to institutions in the empirical analysis of slump duration.

Supplementary Appendix S6 reports the results of similar tests for an array of macroeconomic policy indicators. The general pattern is as expected: the size of the government increases during a slump, while prices and exchange rates react predictably. Other variables are trending over the entire period, but almost none exhibit such stark changes at the start of a slump as the pattern of institutional change documented here.

4 Explaining the duration of declines

Empirical strategy

We model the duration of declines in an ‘accelerated failure time’ (AFT) framework, where ‘failure time’ refers to time until the turnaround at \hat{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

¹³These indexes have different ranges: ‘Executive Constraints’ ranges 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.

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 elapsed within a decline spell ('spell time'). We specify the following regression equation for crisis durations in AFT form:

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

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, $\mathbf{x}_0 = (x_{0,1}, x_{0,2}, \dots, x_{0,k})'$ 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 $\mathbf{z}_t = (z_{t,1}, z_{t,2}, \dots, z_{t,m})'$ is an $m \times 1$ vector of time-varying controls assumed to be 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.¹⁴ 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 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 may respond endogenously to crises. We are less concerned about ethnic diversity, since it is time-constant in our data and historically predetermined.¹⁵ 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).¹⁶ Therefore, to prevent feedback, we set potentially endogenous covariates at their last pre-slump value, corresponding to $t_0 = \hat{t}_{b1}$. In effect, we rely on the temporal ordering to identify these relationships. The estimates should nonetheless be interpreted as partial correlations rather than causal effects.¹⁷ Unobserved

¹⁴For details, see Table S7.5 in Supplementary Appendix S7.

¹⁵The ethnic configuration of a country rarely changes as a short-run response to an economic decline, which is not to deny some tragic exceptions.

¹⁶Countries 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.

¹⁷Supplementary Appendix S5 shows simple split sample plots where we show 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.

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, the log of initial GDP, and the US real 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 US real 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 parameterize 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. As mentioned earlier, this raises the possibility of detecting ‘false slumps’ which

Table 2 – Correlates of slump duration

	<i>Additive</i>			<i>Interacted</i>		
	<i>Fixed effects on . . .</i>					
	None (1)	Region (2)	Region + Time (3)	None (4)	Region (5)	Region + Time (6)
<i>Panel (a) Structural breaks estimated with size = 0.10</i>						
Executive Constraints (\widetilde{INS}_0)	-0.150** (0.062)	-0.174*** (0.063)	-0.227*** (0.072)	-0.225*** (0.072)	-0.255*** (0.077)	-0.282*** (0.084)
Fractionalization ($\widetilde{ELF15}$)	0.011** (0.005)	0.016*** (0.005)	0.016*** (0.005)	0.014*** (0.005)	0.020*** (0.005)	0.019*** (0.005)
$\widetilde{INS}_0 \times \widetilde{ELF15}$				-0.004** (0.002)	-0.004*** (0.001)	-0.003* (0.002)
US Real Interest Rate	-0.139** (0.061)	-0.134*** (0.052)	-0.122** (0.056)	-0.118* (0.060)	-0.118** (0.051)	-0.111** (0.054)
Initial log GDP	-0.149** (0.076)	-0.017 (0.091)	-0.056 (0.078)	-0.157** (0.072)	-0.029 (0.089)	-0.061 (0.078)
Exits	43	43	43	43	43	43
Spells	54	54	54	54	54	54
Years of Decline	383	383	383	383	383	383
Log- \mathcal{L}	-76.214	-73.641	-68.075	-74.780	-71.978	-66.941
Pseudo-R ²	0.098	0.129	0.195	0.115	0.149	0.208
<i>Panel (b) Structural breaks estimated with size = 0.20</i>						
Executive Constraints (\widetilde{INS}_0)	-0.179*** (0.055)	-0.177*** (0.059)	-0.167*** (0.055)	-0.237*** (0.058)	-0.239*** (0.062)	-0.215*** (0.060)
Fractionalization ($\widetilde{ELF15}$)	0.007 (0.004)	0.011** (0.005)	0.011** (0.005)	0.009** (0.005)	0.014*** (0.005)	0.013*** (0.005)
$\widetilde{INS}_0 \times \widetilde{ELF15}$				-0.004** (0.002)	-0.003** (0.001)	-0.003* (0.002)
US Real Interest Rate	-0.073 (0.046)	-0.067* (0.041)	-0.075 (0.054)	-0.059 (0.044)	-0.059 (0.040)	-0.071 (0.052)
Initial log GDP	-0.130* (0.067)	-0.021 (0.075)	-0.065 (0.070)	-0.135** (0.064)	-0.037 (0.073)	-0.075 (0.069)
Exits	61	61	61	61	61	61
Spells	74	74	74	74	74	74
Years of Decline	489	489	489	489	489	489
Log- \mathcal{L}	-105.691	-102.837	-97.312	-104.042	-101.201	-96.122
Pseudo-R ²	0.083	0.108	0.156	0.097	0.122	0.166

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 54 (out of 57) or 74 (out of 77) spells. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Source: Authors' analysis based on data from the Penn World Tables version 9.0 (series *rgdpna/pop*). Additional data come from the Polity IV project, Desmet et al. (2012), and the Federal Reserve Economic Data (FRED) database.

is why we consider it to be a robustness check and always focus our interpretation on the first set of results (unless indicated otherwise). The coefficients of executive constraints and fractionalization are significant in all estimations, apart from column (1) in panel (b) where the appropriate fixed effects are omitted and the coefficient on fractionalization turns marginally insignificant. An improvement in executive constraints by one standard deviation (2.28 points) is associated with a reduction in the duration of the decline phase of about

29–40%. A difference in fractionalization of one standard deviation (33.05 percentage points) corresponds to a difference in duration of about 30–41%. Neither the inclusion of region fixed effects nor the addition of initial decade fixed effects fundamentally change these results.

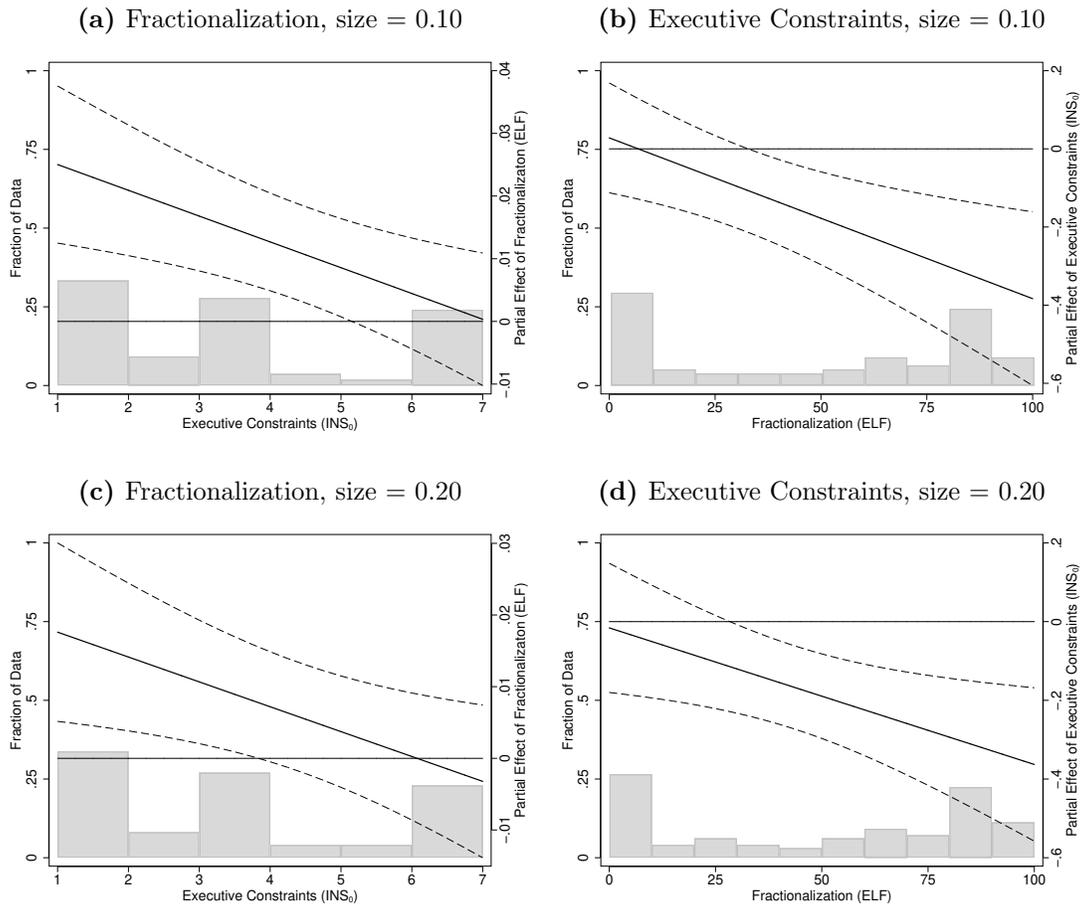
A key point of our paper is that these first few models miss a potentially important interaction effect. If greater ethnic diversity challenges the ability of political actors to take coordinated action, more cohesive institutions may help to overcome this vulnerability by internalizing these disputes and limiting the downside risks for the involved groups (see e.g. 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.¹⁸ The demeaned variables are denoted \widetilde{INS}_0 and \widetilde{ELF} . The interaction effect is negative (as expected), significant at conventional levels in each regression, and remarkably stable. Since the earlier specifications are nested in these columns, testing the significance of

¹⁸This 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.

the interaction basically amounts to testing that the interaction model fits the data better. Likelihood ratio tests also prefer the interaction model (at the 10% significance level) and the pseudo- R^2 s improve in every column.

Figure 3 – Average partial effects



Note(s): Panels (a) and (b) illustrate the results for a nominal size of 10 percent. The break underlying panels (c) and (d) are estimated with a nominal size of 20 percent. The average partial effects shown in panels (a) to (d) are based on column (5) 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 percent confidence limits. *Source:* Authors' analysis based on data from the Penn World Tables version 9.0 (series *rgdpna/pop*). Additional data come from the Polity IV project, Desmet et al. (2012), and the Federal Reserve Economic Data (FRED) database.

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 coefficients. For example, when the executive constraints index is at unity (‘unlimited authority’), then a one percentage point higher fractionalization corresponds to a 2.5% longer duration of contractions (and a standard deviation rise in fractionalization when constraints are weak more than 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 about zero (0.41) to near-total heterogeneity (96.5).

The predictions also cover a wide range of the observed duration. At the average score of executive constraints, a country with the highest (lowest) degree of ethnic heterogeneity is expected to decline for about 15.7 years (2.1 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 75th percentile of ethnic heterogeneity (= 91.9), a country with the highest (lowest) score of executive constraints is expected to decline for about 2.5 years (20 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 -31% at perfect heterogeneity. The estimates for the larger sample are very similar.

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 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 growth rate during the contraction. Hence, we may define $\bar{g}_i \equiv (y_{i,\hat{t}_{min}} - y_{i0}) / (\hat{t}_{min} - t_{b1}) \equiv (y_{i,\hat{t}_{min}} - y_{i0}) / \tilde{t}_D$ as the average rate of decline

and $\tilde{t}_D \times \bar{g}_i \equiv y_{i,\hat{t}_{min}} - y_{i0}$ as the overall depth of the decline. Note that \bar{g}_i is negative by construction. We scale both outcomes by one hundred for readability.

Table 3 – Average growth rate during decline and total depth

	<i>Dependent variable</i>					
	\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)
<i>Panel (a) Structural breaks estimated with size = 0.10</i>						
Executive Constraints (\widetilde{INS}_0)	0.227 (0.190)	0.627** (0.238)	0.732*** (0.217)	4.151*** (1.486)	5.617*** (1.351)	6.275*** (1.581)
Fractionalization ($\widetilde{ELF15}$)	0.007 (0.013)	-0.017 (0.011)	-0.017 (0.016)	-0.153** (0.074)	-0.291*** (0.092)	-0.309*** (0.100)
$\widetilde{INS}_0 \times \widetilde{ELF15}$	0.002 (0.005)	0.003 (0.004)	0.003 (0.005)	0.048 (0.033)	0.060** (0.029)	0.053 (0.032)
US Real Interest Rate	0.082 (0.135)	0.015 (0.140)	0.220 (0.268)	2.773** (1.207)	2.613** (1.122)	0.805 (2.204)
Initial log GDP	0.158 (0.261)	0.413 (0.310)	0.436* (0.240)	2.076 (1.482)	2.179 (1.796)	2.571 (1.766)
Region FE	NO	YES	YES	NO	YES	YES
Initial Decade FE	NO	NO	YES	NO	NO	YES
Spells	54	54	54	54	54	54
Log- \mathcal{L}	-142.247	-137.284	-125.939	-243.640	-241.930	-236.354
Adjusted R ²	-0.064	0.056	0.302	0.232	0.231	0.296
<i>Panel (b) Structural breaks estimated with size = 0.20</i>						
Executive Constraints (\widetilde{INS}_0)	0.302* (0.157)	0.513** (0.197)	0.481** (0.188)	4.309*** (1.230)	5.159*** (1.234)	5.016*** (1.302)
Fractionalization ($\widetilde{ELF15}$)	0.002 (0.010)	-0.021** (0.010)	-0.021* (0.012)	-0.106* (0.062)	-0.241*** (0.081)	-0.251*** (0.086)
$\widetilde{INS}_0 \times \widetilde{ELF15}$	0.007* (0.004)	0.009** (0.004)	0.010** (0.004)	0.061* (0.032)	0.074** (0.030)	0.065** (0.032)
US Real Interest Rate	0.017 (0.110)	0.015 (0.112)	0.124 (0.225)	1.806* (0.960)	1.924** (0.931)	0.698 (1.855)
Initial log GDP	0.146 (0.218)	0.270 (0.268)	0.407 (0.247)	1.650 (1.303)	1.668 (1.531)	2.296 (1.591)
Region FE	NO	YES	YES	NO	YES	YES
Initial Decade FE	NO	NO	YES	NO	NO	YES
Spells	74	74	74	74	74	74
Log- \mathcal{L}	-189.696	-186.990	-176.881	-331.143	-328.817	-324.852
Adjusted R ²	0.001	0.029	0.200	0.222	0.235	0.256

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

Source: Authors' analysis based on data from the Penn World Tables version 9.0 (series *rgdpna/pop*). Additional data come from the Polity IV project, [Desmet et al. \(2012\)](#), and the Federal Reserve Economic Data (FRED) database.

We find only weak evidence 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 interacted model without and with region fixed effects. Column

(3) adds initial decade effects. Though the coefficient of institutions may be statistically significant, it is economically small (e.g. about 0.22–0.63 percentage points of \bar{g}_i in the case of a unit change in executive constraints away from the mean) and explains almost none 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 (4) to (6) illustrate that we now recover the previously estimated effects, albeit with some loss of significance. (The coefficients have reversed signs since \bar{g}_i is negative.) Moreover, the interaction models explain 23–30% of the variation in depth, highlighting the relevance of the estimated effects.

Taken together, the estimates lead 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 is achieved increases with the number of groups involved in the decision-making process and the veto rights 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. A weakly constrained executive might actually reduce the duration of the decline phase in such models, if 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 one example, Zambia's decline from 1968 until 1994 can be understood through the lens of this coordination perspective and fits well with the stylized facts presented here. Zambia is very ethnically heterogeneous, had weak constraints on the political executive just after independence, and experienced one of the longest declines in our data set. The crisis began with an exogenous terms of trade shock when copper prices fell rapidly in the late 1960s (Fraser, 2010). Declining incomes sparked tensions between the Bemba-speaking Copperbelt and the government, so that violent riots had become a frequent sight (Posner, 2005). Then president Kenneth Kaunde reacted with nationalizations and the introduction of a single party government. Kaunde ethnically balanced his cabinet, which he hoped would curb tribalism (Larmer, 2008) but had the opposite effect (Posner, 2005). The crisis dragged on until mine workers and a former trade unionist brought down Kaunde's government, re-introduced multi-party democracy, which enabled broad non-tribal coalitions, and ushered in an era of structural reforms. Copper prices eventually recovered and aided the recovery but the average Zambian was poorer in 2014 than at the start of the crisis. [Supplementary Appendix S1](#) discusses the political economy of this case in more detail.

As mentioned in the introduction, the argument that impersonal rules and limits on executive power can help to mitigate the risk of deeper economic contractions also 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, such as the Bemba miners in Zambia, with whom the ruling elite needs to coordinate (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 [Supplementary Appendix S7](#).

[Table 4](#) takes the estimated breaks and data as given and focuses on unobserved heterogeneity and influential observations. Column (1) adds country-level random effects (frailties) and income-level fixed effects on top of region fixed effects. Column (2) drops former Soviet countries. Both perturbations do not affect our findings. Column (3) takes a more radical step and drops all of Africa. The estimated effect sizes are smaller and we lose precision on the interaction term. This loss of significance is hardly surprising. The African continent is home to some of the most ethnically diverse countries with the weakest institutions. Africa drives our results but also provides much of the variation needed to estimate the overall effect.¹⁹ Column (4) deletes the censored spells. Now the estimated effect of political institutions is even smaller, and the interaction term becomes insignificant. A look into the underlying data reveals what is going on. Countries with completed spells

¹⁹We have also taken a less radical approach and interacted our three variables of interest with an Africa dummy. In every case, the added interactions are individually and jointly insignificant at the 5% level. We take this as further evidence that our results can be generalized to settings outside of Africa, such as Asia, where institutions are stronger and linguistic diversity is comparable. We thank an anonymous referee for pointing us towards this test.

have an average score of 3.1 on the index of executive constraints and an average degree of heterogeneity close to 55 out of 100. Countries with ongoing slumps score 1.54 in terms of executive constraints and have an average level of heterogeneity of about 77. 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 of censoring. The whole point of duration analysis is that we can use censored observations without introducing bias. Note that if the slumps are defined in terms of GDP rather than GDP per capita, there are considerably fewer censored spells and our results hold in both samples without the censored data points.

Table 4 – Robustness: heterogeneity, outliers and recurrent spells

	<i>Current modification</i>					
	Country RE (1)	Without Ex-Soviet (2)	Without Africa (3)	Without Censored (4)	Single Spells (5)	Endogenous Institutions (6)
<i>Panel (a) Structural breaks estimated with size = 0.10</i>						
Executive Constraints (\widetilde{INS}_0)	-0.292*** (0.100)	-0.274*** (0.078)	-0.093* (0.054)	-0.035 (0.051)	-0.256** (0.105)	-0.213** (0.087)
Fractionalization (\widetilde{ELF}_{15})	0.019*** (0.007)	0.021*** (0.005)	0.007** (0.004)	0.009** (0.004)	0.020*** (0.006)	0.018*** (0.006)
$\widetilde{INS}_0 \times \widetilde{ELF}_{15}$	-0.005** (0.002)	-0.004*** (0.001)	-0.002* (0.001)	-0.002 (0.001)	-0.005*** (0.002)	-0.005*** (0.002)
Region FE	YES	YES	YES	YES	YES	YES
Income Level FE	YES	NO	NO	NO	NO	NO
Exits	43	40	30	43	38	38
Spells	54	51	35	43	46	55
<i>Panel (b) Structural breaks estimated with size = 0.20</i>						
Executive Constraints (\widetilde{INS}_0)	-0.212*** (0.080)	-0.250*** (0.063)	-0.104** (0.053)	-0.085* (0.044)	-0.250*** (0.081)	-0.221*** (0.073)
Fractionalization (\widetilde{ELF}_{15})	0.012** (0.006)	0.014*** (0.005)	0.003 (0.005)	0.008** (0.003)	0.012* (0.006)	0.012** (0.005)
$\widetilde{INS}_0 \times \widetilde{ELF}_{15}$	-0.004** (0.002)	-0.004** (0.001)	-0.002 (0.002)	-0.002** (0.001)	-0.004** (0.002)	-0.005*** (0.002)
Region FE	YES	YES	YES	YES	YES	YES
Income Level FE	YES	NO	NO	NO	NO	NO
Exits	61	57	42	61	52	54
Spells	74	70	48	61	61	75

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

Source: Authors' analysis based on data from the Penn World Tables version 9.0 (series *rgdpna/pop*). Additional data come from the Polity IV project, [Desmet et al. \(2012\)](#), and the Federal Reserve Economic Data (FRED) database.

Until now, we have treated spells recurring in the same country as interchangeable.

Column (5) of [Table 4](#) investigates whether this particular form of conditional independence is a reasonable assumption. It turns out that our findings are robust to excluding all spells other than the first, ruling out dependency among recurring slumps.

Column (6) allows for institutions to change endogenously during the decline phase (meaning we replace \widetilde{INS}_0 by \widetilde{INS}_t). We showed earlier that constraints on the executive and a number of other institutional indicators improve in the aftermath of a crisis. So far, we have excluded this endogenous adjustment for two reasons: (i) it could bias the estimation in favor of our hypothesis (if institutions both improve as a response and then shorten the remaining decline phase), and (ii) elites might find it difficult to credibly promise to bind their hands and not renege later on ([Acemoglu and Robinson, 2006](#)). In any case, these concerns are not borne out empirically. The results remain quantitatively and qualitatively similar to our baseline.²⁰

A serious issue 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 (PWT) 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.²¹ 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.²² We take these concerns seriously and repeat the restricted structural break search using two very different data sets (in addition to the PWT 9). Specifically, we use the version 7 vintage of the PWT—before the computation of the chained series changed

²⁰Another way of tackling this issue is to truncate the data and censor all spells lasting longer than a few, say five, years. Our results are robust to this type of truncation (at various points). We thank an anonymous referee for suggesting this test to us.

²¹We are indebted to an anonymous referee for pointing this out to us.

²²The 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* in the robustness checks.

fundamentally—and national accounts data from the World Development Indicators in local currencies. Each time we consider GDP per capita and GDP. 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

	<i>GDP per capita series</i>			<i>GDP series</i>		
	PWT7 <i>rgdpch</i> (1)	PWT9 <i>rgdpna/pop</i> (2)	WDI <i>gdplcu/pop</i> (3)	PWT7 <i>rgdpch × pop</i> (4)	PWT9 <i>rgdpna</i> (5)	WDI <i>gdplcu</i> (6)
<i>Panel (a) Structural breaks estimated with size = 0.10</i>						
Executive Constraints (\widetilde{INS}_0)	-0.209*** (0.064)	-0.255*** (0.077)	-0.170** (0.069)	-0.181*** (0.046)	-0.163*** (0.045)	-0.161*** (0.051)
Fractionalization (\widetilde{ELF}_{15})	0.018*** (0.004)	0.020*** (0.005)	0.009* (0.005)	0.011*** (0.003)	0.009*** (0.002)	0.008*** (0.002)
$\widetilde{INS}_0 \times \widetilde{ELF}_{15}$	-0.003** (0.001)	-0.004*** (0.001)	-0.005** (0.002)	-0.003*** (0.001)	-0.004*** (0.001)	-0.003*** (0.001)
Exits	48	43	47	59	63	53
Spells	58	54	59	62	64	56
<i>Panel (b) Structural breaks estimated with size = 0.20</i>						
Executive Constraints (\widetilde{INS}_0)	-0.197*** (0.048)	-0.239*** (0.062)	-0.175*** (0.060)	-0.122*** (0.036)	-0.109** (0.046)	-0.142*** (0.042)
Fractionalization (\widetilde{ELF}_{15})	0.014*** (0.004)	0.014*** (0.005)	0.009** (0.004)	0.008*** (0.003)	0.006** (0.003)	0.007*** (0.002)
$\widetilde{INS}_0 \times \widetilde{ELF}_{15}$	-0.004*** (0.001)	-0.003** (0.001)	-0.004*** (0.002)	-0.004*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)
Exits	67	61	61	72	76	75
Spells	79	74	75	75	77	78

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

Source: Authors' analysis based on data from the Penn World Tables version 7.0, Penn World Tables version 9.0, and the World Development Indicators. Additional data come from the Polity IV project, [Desmet et al. \(2012\)](#), and the Federal Reserve Economic Data (FRED) database.

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). All results are qualitatively (and often even numerically) similar. 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.

Table 6 – Robustness: policy variables

	<i>Policy variable</i>					
	Policy Volatility (1)	Government Size (2)	Inflation (ln[1 + δ]) (3)	RER Underval (4)	Export Share (5)	Trade Openness (d.f.) (6)
<i>Panel (a) Structural breaks estimated with size = 0.10</i>						
Executive Constraints (\widetilde{INS}_0)	-0.255*** (0.075)	-0.254*** (0.073)	-0.235*** (0.073)	-0.253*** (0.075)	-0.239*** (0.079)	-0.260*** (0.078)
Fractionalization ($\widetilde{ELF}15$)	0.020*** (0.005)	0.020*** (0.005)	0.019*** (0.005)	0.020*** (0.005)	0.018*** (0.005)	0.018*** (0.005)
$\widetilde{INS}_0 \times \widetilde{ELF}15$	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.002)	-0.004*** (0.001)	-0.004*** (0.001)
Policy Variable	-0.011 (0.027)	-0.000 (0.012)	0.005*** (0.002)	0.142 (0.388)	0.019** (0.009)	0.006 (0.006)
Exits	43	43	32	43	43	43
Spells	54	54	39	54	54	54
<i>Panel (b) Structural breaks estimated with size = 0.20</i>						
Executive Constraints (\widetilde{INS}_0)	-0.243*** (0.062)	-0.241*** (0.062)	-0.221*** (0.060)	-0.238*** (0.062)	-0.238*** (0.062)	-0.250*** (0.067)
Fractionalization ($\widetilde{ELF}15$)	0.013*** (0.005)	0.014*** (0.005)	0.014** (0.005)	0.013*** (0.005)	0.012** (0.005)	0.012** (0.005)
$\widetilde{INS}_0 \times \widetilde{ELF}15$	-0.003** (0.001)	-0.003** (0.001)	-0.003** (0.002)	-0.004** (0.002)	-0.003** (0.001)	-0.004** (0.001)
Policy Variable	-0.013 (0.022)	0.003 (0.010)	0.005*** (0.001)	0.069 (0.295)	0.014** (0.007)	0.005 (0.004)
Exits	61	61	45	61	61	61
Spells	74	74	54	74	74	74

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

Source: Authors' analysis based on data from the Penn World Tables version 9.0 (series *rgdpna/pop*). The policy variables are computed using data from the PWT version 9.0, the World Development Indicators and the International Financial Statistics. Additional data come from the Polity IV project, Desmet et al. (2012), and the Federal Reserve Economic Data (FRED) database.

The relationships reported here survive many other sensible 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 [Supplementary Appendix S7](#).

Finally, we examine if accounting for macroeconomic policies changes the nature of our conclusions with respect to political institutions and heterogeneity.²³ 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 \(2013\)](#), 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. Among the added indicators, only inflation and the export share turn out to be significant predictors of the duration of declines. Seriously analyzing the role of policy variables exceeds the scope of this paper, but it seems safe to conclude that such effects are not competing with our main results.

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 pattern. We show that a restricted structural change approach directly incorporating this pattern 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

²³We are indebted to an anonymous referee for pointing us in this direction.

suggests that there are, at least, two important qualifications to be made.

First, our event study illustrates that weak political institutions precede 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 an indication 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 has emphasized 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). 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. Working Paper 2015–003, UNU-MERIT.
- Broadberry, S. and J. J. Wallis (2017). 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). OLS and IV estimation of regression models including endogenous interaction terms. Working Paper 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.
- Fraser, A. (2010). Introduction: Boom and bust on the Zambian Copperbelt. In A. Fraser and M. Larmer (Eds.), *Zambia, Mining, and Neoliberalism: Boom and Bust on the Globalized Copperbelt*, pp. 1–30. New York: Palgrave Macmillan US.
- 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.
- Larmer, M. (2008). Enemies within? Opposition to the Zambian one-party state, 1972–1980. In J.-B. Gewald, M. Hinfelaar, and G. Macola (Eds.), *One Zambia, Many Histories*, pp. 98–125. Leiden, The Netherlands: Brill.

- 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.
- Posner, D. N. (2005). *Institutions and ethnic politics in Africa*. Cambridge University Press.
- 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.
- Szirmai, A. and N. Foster-McGregor (2017). Understanding the ability to sustain growth. Working Paper 173, Groningen Growth and Development Centre.

Supplementary Appendix

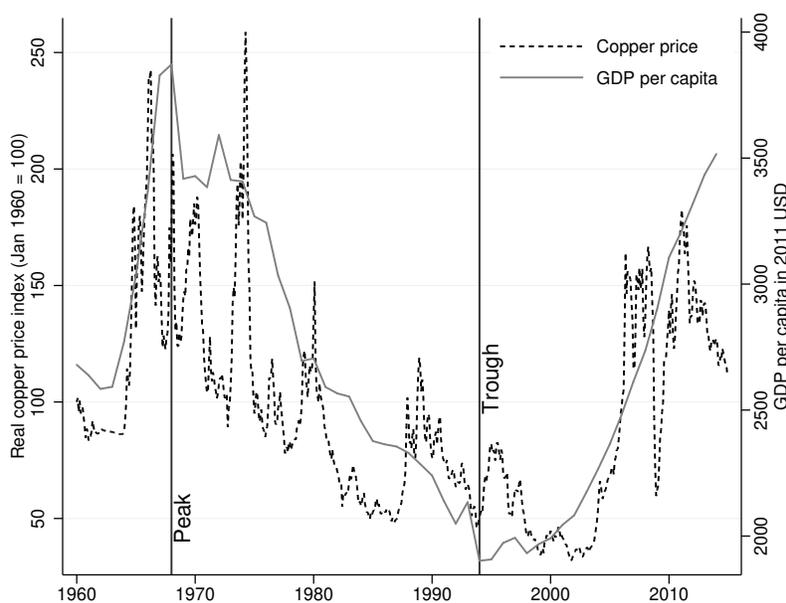
S1 The case of Zambia	ii
S2 List of breaks	v
S3 Estimation of structural breaks	ix
S4 Variables and summary statistics	xviii
S5 Split sample plots	xx
S6 Additional event plots	xxi
S7 Additional regression results	xxiii

S1 The case of Zambia

Zambia’s decline from 1968 until 1994 is one of the longest in our data set. It encompasses a succession of crises which paved the way for a spectacular reversal of fortunes. At independence, Zambia had a large industrial base and was considered “one of the richest and most promising of the new African states” (Ferguson, 1999, p. 2). By the late 1980s, it had fallen far behind other “promising” countries such as Kenya and most countries in West Africa. While many countries in sub-Saharan Africa experienced growth collapses in the 1970s and decline through the 1980s, Zambia’s crisis started a few years earlier and was exceptionally deep, wiping out about half of GDP per capita according to our own measure of crisis depth.

Zambia is one of the most diverse countries in Africa. Although there are “only” eight major language groups, Zambia is home to a total of 46 different languages¹ and more than 70 tribes (Posner, 2005). In other words, the probability that two randomly drawn individuals speak a different language is 85% (i.e. ELF = 0.85). Although the Bemba speakers are a large group, Zambia is only moderately polarized and its ethnic groups have coexisted peacefully for most of its history.

Figure S1.1 – Real GDP per capita and copper prices



Notes: Annual real GDP per capita is based on the series *rgdpcma/pop* from the Penn World Table Version 9. The index of monthly real copper prices is based on the nominal copper prices from World Bank Commodity Price Data (The Pink Sheet) and deflated by the monthly US CPI from FRED [CPALTT01USM661S]. 01/1960 corresponds to 100.

In 1964, president Kenneth Kaunda’s United National Independence Party (UNIP) had won the first independent multi-party elections in a landslide victory and was presiding over an early post-independence boom. The newly formed country was home to about a third of the world’s known copper reserves and prices, mining output and productivity had been

¹See <https://www.ethnologue.com/country/ZM/languages>.

rising since the early 1930s. At independence, copper mining accounted for 45% of GDP at factor prices (Szirmai et al., 2002).

By mid-1968, signs appeared that the economy was weakening as the result of an adverse commodity shock. Copper prices fell sharply in April 1968 (see Figure S1.1), real wages started to decline from 1969 onward (Fraser, 2010). Kaunda tried to control the flows of the declining mineral rents by nationalizations. In 1969, the Zambian state took a 51% majority stake in the mining companies. This was initially welcomed by the international observers who expected that profits would be more secure with government participation, but was then followed by more ambitious Zambianization programs (Fraser, 2010). By the late 1980s, parastatal companies accounted for some 90% of all commercial and industrial activities and about 35% of GDP (Szirmai et al., 2002). The reforms were ill-timed. After rising to an all-time high during the oil crisis in 1973-1974, copper prices entered a secular decline until shortly after the turn of the millennium.

Economic nationalization was quickly followed by political centralization. The 1968 multi-party election was already marked by ethnic strife and increasing tribalism was a widespread concern. “In the five years leading up to the elimination of multi-party politics, riots, arson, beatings, and other forms of violence had become regular features of political competition in several areas of the country” (Posner, 2005, pp. 84–85). Fraser (2010, p. 9) makes that more explicit when he writes “as the recession hit, Kaunda lived in fear of a political rebellion emerging on the Copperbelt and closed down political competition both within the party and nationally.” The industrial base in the Copperbelt suffered greatly from the recession due to its near exclusive orientation towards mining and related industries (Szirmai et al., 2002).

Kaunda tried to actively avert ethnic conflict by tribally balancing his cabinet. This policy, however, solidified ethnic competition within the party (Larmer, 2008). The Bemba-speakers of the Copperbelt and Northern Province felt particularly disadvantaged. Larmer (2008, p. 104) argues that “Kaunda’s decision to ethnically balance the Central Committee was perceived by Bembas as unfairly marginalising their disproportionately high membership in the party [...] and increasing the potential for a breakaway party.” This is precisely what happened: the United Progressive Party (UPP) was launched in 1971, which was quickly followed by violent clashes with UNIP followers in the Copperbelt and repeated arrests of party leaders.

Zambia’s First Republic officially ended in 1972 when Kaunda banned opposition parties and introduced single-party government. Interestingly, this constitutional change masks a remarkable level of institutional continuity: “Zambia’s multi-party First and Third Republics were characterized by relatively illiberal forms of democracy, and its lengthy one-party Second Republic was marked by a relatively mild form of authoritarianism” (Posner, 2005, p. 85). This is reflected in our data. From 1968 onward, the Polity IV database records the second worst category of constraints on the executive, just above ‘unlimited authority.’ By 1972, this deteriorated only by one point to the lowest category of executive constraints.

Economic mismanagement and rent-seeking took their toll on the Zambian economy after the oil price shock of the 1970s. The state first reacted with increased international borrowing, but the continued decline of copper prices after the repeated oil price shock in 1979 meant that borrowing costs sky-rocketed. In the face of declining rents, “ZCCM [Zambia Consolidated Copper Mines] was treated as a ‘cash cow,’ milked without corresponding

investment in machinery and prospecting ventures. No new mines were opened after 1979” (Fraser, 2010, p. 9). In the 1980s, Zambia was one of the most indebted nations in the world, with a peak debt-to-GDP ratio of about 234% in 1986.² The IMF and the World Bank stepped in repeatedly with debt-relief programs in exchange for structural adjustment.

The deep recession spurred mounting unrest, raising the pressure to liberalize and return to multi-party democracy. Mine workers and the labor union eventually brought down Kaunda’s UNIP government. Kaunda withdrew Zambia from a structural adjustment program when food riots broke out in the Copperbelt and paved the way for free elections. Frederick Chiluba, a former trade unionist, was heading the Multi-Party Democracy (MMD) movement which won a landslide victory over UNIP in 1991. Kaunda stepped down, marking Africa’s first peaceful transition from a one-party state in the 1990s. The Polity IV project recognizes this as a regime change and henceforth codes Zambia as having ‘substantial limitations on executive authority.’ While Chiluba emphasized the need for ‘dual’ political and economic reform, he quickly introduced Zambians to another iteration of limited democracy by restricting the press, arresting and declaring two states of emergency (Larmer, 2008, Fraser, 2010).

Chiluba and the MMD government took sweeping action to stabilize the economy. Within a few years the exchange rate was floated, capital flows and trade were liberalized, agricultural marketing was liberalized, public sector reforms were undertaken, and the mining sector was privatized. This ushered in a new area of international support and large aid inflows (Rakner, 2003). In the 1990s, privatization was first hailed as widely successful by the World Bank and other observers (Rakner, 2003), then re-assessed as a failure when big companies such as Anglo-American found themselves unable to turn some of Zambia’s largest mines to profit (Fraser, 2010).

Figure S1.1 shows that the economy finally started to recover after the trough in 1994. By the mid-1990s, GDP growth turned positive but was uneven across sectors and in terms of its welfare implications. Rural poverty rates were falling while urban poverty remained high amid fast population growth (McCulloch et al., 2000). Although the reform process stalled and the recovery was sluggish, none of the key reforms were reversed in this decade. In 2001, the copper mines were reporting their first year of positive productivity growth since the 1970s. Robust growth thus returned just before rising copper prices from 2004 onward fueled a new export boom. Although Zambia is still heavily dependent on copper exports, recent volatility in world markets did not translate into another recession.

²See www.imf.org/external/datamapper/datasets/DEBT/1.

S2 List of breaks

Table S2.1 – Global parameters

Data:	PWT 9.0 (<i>rgdpna</i>)	Max AR (p_{max}):	4
Sample start (T_0):	1950	Bootstrap replications:	1000
Sample end (T):	2014	Bootstrap errors:	parametric
Trimming (τ):	0.05	Bootstrap type:	recursive
Min. $\hat{t}_{b1} - \hat{t}_{b2}$ lag (h):	4	Bootstrap significance (α^s):	0.1 (0.2)

Table S2.2 – Estimated slumps: smaller sample of 57 episodes^a

Code	T_0	\hat{t}_{b1}	\hat{t}_{min}	\hat{t}_{b2}	T	Sup- W	Crit. W	p-value	Drop (%)	\tilde{t}_D	Csrd.
ALB	1970	1990	1992	2010	2014	31.8	9.8	0.000	-32.34	2	0
BDI	1960	1992	2005	2005	2014	10.9	9.6	0.060	-36.16	13	0
BFA	1959	1982	1984	1994	2014	9.8	7.7	0.037	-11.43	2	0
CAF	1960	1970	2014	2010	2014	6.6	5.6	0.059	-58.63	44	1
CHE	1950	1974	1976	1988	2014	16.4	9.4	0.002	-8.79	2	0
CHL	1951	1953	1954	1959	1971	9.6	7.6	0.043	-7.27	1	0
CHL	1951	1972	1975	1976	1980	14.0	10.5	0.030	-20.97	3	0
CHL	1951	1981	1983	1992	2014	11.4	9.5	0.034	-18.45	2	0
CHN	1952	1960	1961	1991	2014	11.7	10.1	0.048	-29.31	1	0
CHN	1992	1997	1998	2007	2014	13.9	12.8	0.074	-0.23	1	0
CIV	1960	1979	2011	2011	2014	15.9	9.1	0.002	-44.30	32	1
CRI	1950	1980	1982	2001	2014	10.4	8.9	0.035	-14.12	2	0
DZA	1960	1984	1994	1994	2014	14.2	7.5	0.001	-18.87	10	0
EGY	1950	1972	1973	1984	2014	8.9	8.8	0.098	-1.32	1	0
EST	1990	2008	2009	2012	2014	12.6	9.6	0.041	-14.47	1	0
ETH	1950	1969	1992	2000	2014	15.2	9.9	0.005	-38.15	23	0
GAB	1960	1976	2009	1980	2014	12.8	9.2	0.019	-56.59	33	1
GBR	1950	1979	1981	2005	2014	9.6	8.1	0.025	-3.02	2	0
GNB	1960	1969	1971	1973	2014	9.1	6.8	0.024	-28.04	2	0
GNB	1974	1979	1983	1991	2014	8.5	6.8	0.043	-22.37	4	0
GNB	1992	1997	1998	2004	2014	20.0	8.0	0.003	-18.68	1	1
GTM	1950	1980	1986	1984	2014	14.4	9.8	0.009	-19.04	6	0
HUN	1970	1990	1993	2004	2014	12.3	9.8	0.032	-14.91	3	0
IDN	1960	1997	1999	2001	2014	29.3	7.9	0.000	-14.51	2	0
IND	1950	1978	1979	2001	2014	13.5	11.6	0.031	-7.19	1	0
IRN	1955	1969	1970	1973	1976	42.2	12.1	0.001	-40.93	1	0
IRN	1955	1977	1988	1988	2014	9.3	9.1	0.084	-59.26	11	1
IRQ	1970	1990	1991	1998	2014	15.5	7.9	0.003	-67.02	1	0
JPN	1950	1973	1974	1990	2014	15.2	11.3	0.018	-2.50	1	0
MDA	1990	1993	1999	1999	2014	9.4	6.5	0.031	-39.11	6	0
MDG	1960	1971	2002	1994	2014	8.0	7.5	0.074	-49.93	31	1
MEX	1950	1981	1988	1986	2014	12.3	10.4	0.027	-13.35	7	0
MLI	1960	1963	1967	2008	2014	9.3	6.4	0.022	-17.73	4	0
MNG	1970	1989	1993	2000	2014	16.0	11.2	0.020	-26.53	4	0
MOZ	1960	1981	1985	1992	2014	12.0	10.8	0.060	-26.76	4	0
NAM	1960	1981	1990	1998	2014	11.0	10.2	0.061	-20.75	9	0

Continued on next page

Table S2.2 – *Continued from previous page*

Code	T_0	$\hat{t}b_1$	\hat{t}_{min}	$\hat{t}b_2$	T	Sup- W	Crit. W	p-value	Drop (%)	\tilde{t}_D	Csrd.
NIC	1950	1958	1959	1965	1976	10.3	10.0	0.093	-4.99	1	0
NIC	1950	1977	1993	1992	2014	19.8	9.7	0.002	-58.24	16	1
NPL	1960	1982	1983	2007	2014	11.8	11.2	0.081	-5.21	1	0
PER	1950	1987	1992	1991	2014	13.7	9.7	0.012	-31.00	5	0
PHL	1950	1983	1985	2001	2014	23.6	10.0	0.000	-18.63	2	0
POL	1970	1979	1982	1993	2014	14.6	11.8	0.031	-21.79	3	0
QAT	1970	1976	1991	1985	2014	9.9	7.8	0.034	-57.11	15	1
RUS	1990	2008	2009	2012	2014	7.2	6.9	0.091	-7.82	1	0
RWA	1960	1993	1994	1997	2014	17.9	7.9	0.000	-46.21	1	0
SAU	1970	1980	1987	1984	2014	16.7	9.2	0.002	-56.85	7	1
SDN	1970	1976	1984	1984	2014	9.8	8.9	0.064	-28.07	8	0
SLE	1961	1994	1999	1999	2014	15.5	8.6	0.004	-50.35	5	0
SLV	1950	1979	1983	1998	2014	15.2	9.2	0.008	-25.16	4	0
SVN	1990	2008	2013	2013	2014	54.5	11.6	0.001	-11.07	5	1
SWZ	1970	1982	1985	1992	2014	19.0	6.0	0.001	-6.08	3	0
THA	1950	1996	1998	2003	2014	11.9	7.7	0.011	-12.21	2	0
TTO	1950	1961	1962	1975	1981	13.2	11.3	0.052	-0.70	1	0
TTO	1950	1982	1993	2003	2014	12.2	9.9	0.033	-35.15	11	0
TZA	1960	1980	1983	1997	2014	15.6	11.4	0.021	-11.07	3	0
UGA	1950	1972	1986	1985	2014	13.9	9.6	0.010	-38.46	14	0
ZMB	1955	1968	1994	1995	2014	19.4	9.9	0.002	-50.88	26	1

^a Out of a total of 69 episodes identified by the sequential algorithm, 12 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)]: ARM (1), AZE (1), BIH (1), CRI (1), DNK (1), EGY (1), HRV (1), KOR (1), MRT (1), QAT (1), SDN (1), TWN (1). In these tables T_0 and T denote the first and last year of the active (sub-)sample, respectively.

Table S2.3 – Estimated slumps: larger sample of 77 episodes^a

Code	T_0	$\hat{t}b_1$	\hat{t}_{min}	$\hat{t}b_2$	T	Sup- W	Crit. W	p-value	Drop (%)	\tilde{t}_D	Csrd.
ALB	1970	1990	1992	2010	2014	31.8	8.2	0.000	-32.34	2	0
BDI	1960	1992	2005	2005	2014	10.9	8.0	0.060	-36.16	13	0
BFA	1959	1982	1984	1994	2014	9.8	6.3	0.037	-11.43	2	0
BGR	1970	1988	1999	1996	2014	9.8	8.3	0.103	-25.80	11	0
BOL	1950	1981	1986	2005	2014	7.0	6.9	0.193	-20.96	5	0
CAF	1960	1970	2014	2010	2014	6.6	4.7	0.068	-58.63	44	1
CHE	1950	1957	1958	1961	1973	7.4	5.9	0.104	-3.52	1	0
CHE	1950	1974	1976	1988	2014	16.4	8.2	0.002	-8.79	2	0
CHL	1951	1953	1954	1959	1971	9.6	6.0	0.051	-7.27	1	0
CHL	1951	1972	1975	1976	1980	14.0	8.1	0.028	-20.97	3	0
CHL	1951	1981	1983	1992	2014	11.4	8.0	0.025	-18.45	2	0
CHN	1952	1960	1961	1991	2014	11.7	8.5	0.045	-29.31	1	0
CHN	1992	1997	1998	2007	2014	13.9	10.1	0.070	-0.23	1	0
CIV	1960	1979	2011	2011	2014	15.9	7.4	0.000	-44.30	32	1
CMR	1960	1966	1967	1973	1984	11.1	8.7	0.103	-12.75	1	0
CMR	1960	1985	1994	1992	2014	9.7	8.2	0.096	-42.70	9	1
COG	1960	1973	1977	1981	2014	9.1	8.3	0.156	-15.40	4	0
CRI	1950	1980	1982	2001	2014	10.4	7.8	0.054	-14.12	2	0

Continued on next page

Table S2.3 – *Continued from previous page*

Code	T_0	$\hat{t}b_1$	\hat{t}_{min}	$\hat{t}b_2$	T	Sup- W	Crit. W	p-value	Drop (%)	\tilde{t}_D	Csrd.
DZA	1960	1984	1994	1994	2014	14.2	6.5	0.001	-18.87	10	0
EGY	1950	1972	1973	1984	2014	8.9	7.5	0.098	-1.32	1	0
EST	1990	2008	2009	2012	2014	12.6	7.1	0.041	-14.47	1	0
ETH	1950	1969	1992	2000	2014	15.2	8.5	0.005	-38.15	23	0
FRA	1950	1974	1975	2004	2014	9.9	9.7	0.181	-1.65	1	0
GAB	1960	1976	2009	1980	2014	12.8	7.8	0.010	-56.59	33	1
GBR	1950	1979	1981	2005	2014	9.6	6.9	0.031	-3.02	2	0
GHA	1955	1974	1983	2003	2014	9.7	8.6	0.119	-36.47	9	0
GNB	1960	1969	1971	1973	2014	9.1	5.5	0.024	-28.04	2	0
GNB	1974	1979	1983	1991	2014	8.5	5.4	0.043	-22.37	4	0
GNB	1992	1997	1998	2004	2014	20.0	6.3	0.003	-18.68	1	1
GTM	1950	1980	1986	1984	2014	14.4	8.7	0.015	-19.04	6	0
HTI	1960	1980	2010	1994	2014	7.7	7.1	0.152	-45.27	30	1
HUN	1970	1990	1993	2004	2014	12.3	8.3	0.032	-14.91	3	0
IDN	1960	1997	1999	2001	2014	29.3	6.5	0.000	-14.51	2	0
IND	1950	1978	1979	2001	2014	13.5	10.3	0.031	-7.19	1	0
IRL	1950	1955	1957	1973	1989	12.1	7.1	0.017	-4.95	2	0
IRN	1955	1969	1970	1973	1976	42.2	9.6	0.001	-40.93	1	0
IRN	1955	1977	1988	1988	2014	9.3	7.8	0.084	-59.26	11	1
IRQ	1970	1990	1991	1998	2014	15.5	6.7	0.003	-67.02	1	0
ISR	1950	1964	1965	1971	2014	9.7	8.3	0.098	-4.33	1	0
JPN	1950	1973	1974	1990	2014	15.2	9.5	0.018	-2.50	1	0
KEN	1950	1991	2002	2003	2014	7.2	6.7	0.154	-7.98	11	0
LTU	1990	2008	2009	2012	2014	8.2	6.1	0.100	-13.54	1	0
MDA	1990	1993	1999	1999	2014	9.4	4.9	0.031	-39.11	6	0
MDG	1960	1971	2002	1994	2014	8.0	6.1	0.074	-49.93	31	1
MEX	1950	1981	1988	1986	2014	12.3	8.8	0.027	-13.35	7	0
MLI	1960	1963	1967	2008	2014	9.3	5.5	0.022	-17.73	4	0
MNG	1970	1989	1993	2000	2014	16.0	9.3	0.020	-26.53	4	0
MOZ	1960	1981	1985	1992	2014	12.0	9.2	0.060	-26.76	4	0
MYS	1955	1984	1986	1992	2014	9.3	8.2	0.116	-5.46	2	0
MYS	1993	1997	1998	2006	2014	15.1	7.0	0.015	-9.64	1	0
NAM	1960	1981	1990	1998	2014	11.0	8.7	0.061	-20.75	9	0
NIC	1950	1958	1959	1965	1976	10.3	8.1	0.093	-4.99	1	0
NIC	1950	1977	1993	1992	2014	19.8	8.1	0.002	-58.24	16	1
NPL	1960	1982	1983	2007	2014	11.8	9.6	0.081	-5.21	1	0
PER	1950	1987	1992	1991	2014	13.7	8.2	0.012	-31.00	5	0
PHL	1950	1983	1985	2001	2014	23.6	8.3	0.000	-18.63	2	0
POL	1970	1979	1982	1993	2014	14.6	9.9	0.031	-21.79	3	0
PRT	1950	1957	1958	1966	1972	13.4	13.2	0.196	-0.02	1	0
PRT	1950	1973	1975	2000	2014	11.0	10.2	0.129	-6.01	2	0
QAT	1970	1976	1991	1985	2014	9.9	6.7	0.039	-57.11	15	1
RUS	1990	2008	2009	2012	2014	7.2	5.4	0.091	-7.82	1	0
RWA	1960	1993	1994	1997	2014	17.9	6.5	0.000	-46.21	1	0
SAU	1970	1980	1987	1984	2014	16.7	7.9	0.002	-56.85	7	1
SDN	1970	1976	1984	1984	2014	9.8	7.5	0.064	-28.07	8	0
SLE	1961	1974	1977	1990	1993	9.3	7.6	0.101	-6.01	3	0
SLE	1961	1994	1999	1999	2014	15.5	7.3	0.004	-50.35	5	0
SLV	1950	1979	1983	1998	2014	15.2	7.7	0.008	-25.16	4	0
SVN	1990	2008	2013	2013	2014	54.5	9.2	0.001	-11.07	5	1

Continued on next page

Table S2.3 – *Continued from previous page*

Code	T_0	$\hat{t}b_1$	\hat{t}_{min}	$\hat{t}b_2$	T	Sup- W	Crit. W	p-value	Drop (%)	\tilde{t}_D	Csrd.
SWZ	1970	1982	1985	1992	2014	19.0	4.9	0.001	-6.08	3	0
SYR	1960	1983	1989	2010	2014	9.1	7.9	0.111	-23.64	6	0
TCD	1960	1976	1981	1999	2014	7.7	7.3	0.172	-45.13	5	0
THA	1950	1996	1998	2003	2014	11.9	6.5	0.011	-12.21	2	0
TTO	1950	1961	1962	1975	1981	13.2	9.4	0.052	-0.70	1	0
TTO	1950	1982	1993	2003	2014	12.2	8.6	0.033	-35.15	11	0
TZA	1960	1980	1983	1997	2014	15.6	9.7	0.021	-11.07	3	0
UGA	1950	1972	1986	1985	2014	13.9	8.1	0.010	-38.46	14	0
ZMB	1955	1968	1994	1995	2014	19.4	8.3	0.002	-50.88	26	1

^a Out of a total of 93 episodes identified by the sequential algorithm, 16 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)]: ARM (1), AZE (1), BIH (1), CRI (1), DNK (1), EGY (1), HRV (1), IRL (1), KOR (1), LVA (1), MRT (1), QAT (1), SDN (1), SWZ (1), TWN (1), UZB (1). T_0 and T denote the first and last year of the active (sub-)sample, respectively.

S3 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 $[t_{b1}, t_{b2}] \subset [\tau T, (1 - \tau)T]$.³
4. Compute the sup- W test statistic of the null hypothesis 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

³For 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.

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.

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 S7.1 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 $u_t \sim \text{i.i.d. } \mathcal{N}(0, \hat{\sigma}_{\hat{e}}^2)$
3. Construct a bootstrap sample of equal size as the original sample:

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

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. Re-run 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% and 20% throughout the text).

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. They 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 very well for identifying growth spurts, it cannot be easily modified to reliably identify a broad sample of recessions.

An obvious starting point is 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 36 potential slumps in their original data.⁴ Table S3.1 provides a full list of these slumps, including statistics on their duration and depth. For a level playing field, we re-run our own structural break search on the Penn World Tables version 6.3 which includes data covering the same period as their paper, i.e. 1950–2006.⁵ We identify 50 slumps in this data which is substantially larger than the number of inverted episodes.⁶

Our strategy is to contrast aggregate summary statistics and visually compare individual spells, since what constitutes a slump is often best understood with actual figures and well-known episodes. However, there are limits to the conclusions we can draw from such anecdotal evidence. Establishing which structural change method is truly “better” at identifying slumps could be attempted by Monte Carlo simulations comparing different techniques and their ability to detect the stylized types of slumps found in actual data. We consider this beyond the scope of this paper.⁷

The main discrepancies we have encountered are (i) the start dates and economic meaning of the identified episodes do not necessarily correspond to what we define as slumps, (ii) several slumps are found when the country is already in a crisis, and (iii) our method detects considerably more advanced economy crises, which are lacking in the inverted Berg et al. (2012) episodes.

Table S3.2 shows summary statistics for both approaches. Several things stand out: (i) The inverted Berg et al. (2012) method yields no slumps in upper middle income countries

⁴We are grateful to Andrew Berg for providing us with the list of episodes and an anonymous referee for suggesting this direct comparison. The Berg et al. (2012) estimates are based on the Penn World Tables version 6.2 over the period 1950–2004 and data from the IMF World Economic Outlook for 2005–2006. As in our sample, we drop micro-states and countries with less than 20 years of data. We drop 5 additional episodes which occur when a previous slump is still ongoing, that is, GDP per capita in a specific country at a potential second downbreak is below the level of GDP per capita observed at the previous downbreak. This is equivalent to the filters we apply to our data.

⁵The Penn World Tables versions 6.2 and 6.3 should be nearly identical, with the exception that the latter contains data through 2007.

⁶Note that changing the underlying data implies that we are examining a different list of slumps here than in the rest of the paper.

⁷We would expect that our method does better on slumps that correspond to our structure comprising sharp drop, recovery and stabilization, while one-break-at-a-time approaches should do better in scenarios where there is a single downbreak and continued stagnation.

(where we find 16), only four slumps in advanced economies (where we find 12), and no slumps in Europe (where we find 12); *(ii)* most breaks selected by the inversion are very deep, with surprisingly little variation across income groups or regions; and *(iii)* the majority of slumps based on the inverted approach lasts very long, often exceeding a decade, whereas our two-break model suggests much greater variability in duration.

Figure S3.1 directly compares four representative episodes found by both methods (out of 13 episodes which come out very similar in both approaches). The upper graphs show two examples (Benin and Costa Rica) where our approach more accurately detects the starting date of the crisis and also places the second break at a meaningful point. In both cases, the Berg et al. (2012) method also detects a meaningful second break which coincides with the start of the recovery. The lower graphs illustrate two episodes where both methods deliver almost exactly the same result. Hence, for the subset of shared episodes, our approach works at least as well as the inverted episodes from Berg et al. (2012) and often tends to improve the detection of the starting date.

Figure S3.2 takes a closer look at several more inverted episodes. In several cases, the Bai-Perron approach detects breaks when the country already is in a crisis (i.e., it registers the second dip of a double-dip recession).⁸ In Liberia, there is a sharp downturn in 1988 (the year before the first Liberian civil war) but the country has already been in decline in the years before. These cases are difficult to categorize. We typically try to find the earliest starting date of the crisis, but there are arguments to be made in favor of a start in 1988–89. Berg et al. (2012) find many breaks in Saudi Arabia, but the sequence which passes the -2% growth requirement starts much too late. Our approach instead puts the start of the slump to 1973. In other cases, such as Madagascar or Mongolia, the start of the slump is placed before the rapid downturn occurs.

Figure S3.3 shows several crises in OECD countries which are absent in the inverted Berg et al. (2012) episodes. In fact, of the countries currently in the OECD, the inverted approach only finds declines in Chile and Mexico (before OECD membership). Our method detects substantially more slumps in advanced economies, including the Finnish banking crisis of the 1990s and several advanced economy reactions to the oil price shocks of the 1970s. Conversely, there are episodes in the Berg et al. (2012) data which are not included in our list of spells and are plausible candidates. These discrepancies are to be expected, since the two-break model and multiple one-break models test a different number of coefficients. However, many of those episodes are found by our approach once we vary the break search parameters, e.g. by raising the nominal size or changing the interbreak period. In the light of the many robustness checks reported in the paper and this appendix, we think it is safe to conclude that our selection of slumps is not biased in a particular direction.

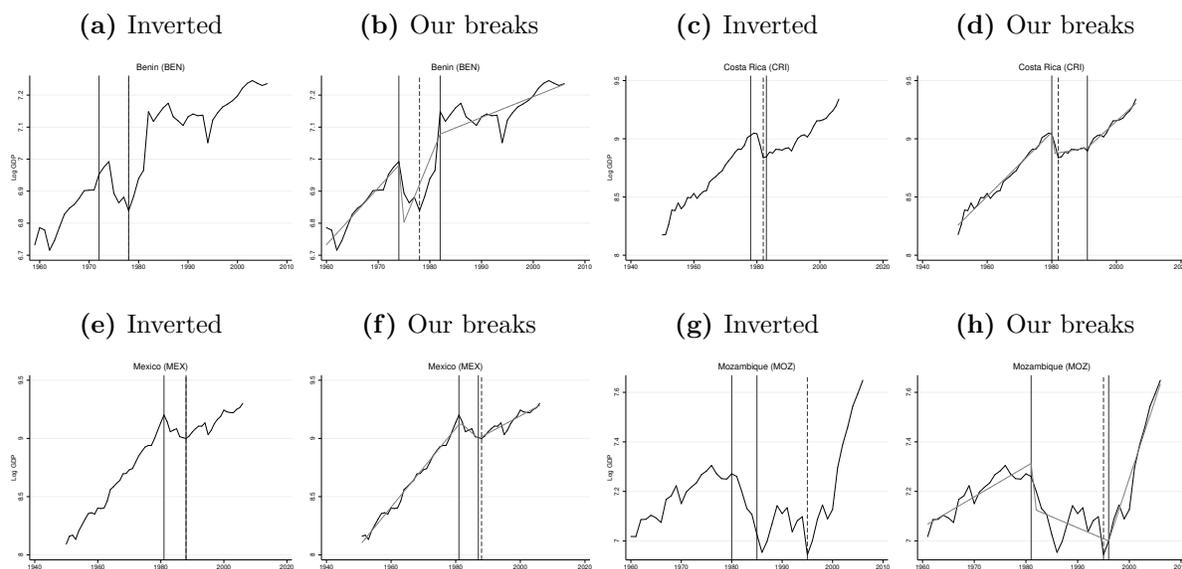
There are other conceptual issues which lead us to prefer the restricted structural change approach. Imposing a specific growth rate after the break treats all time series alike in an absolute sense, although a deep slump in a stable country like Germany might look very different than that of a volatile 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

⁸The breaks occur very early in Cambodia and Kuwait, relative to the start of the respective PWT series. We suspect Berg et al. (2012) may have extended the sample backwards for selected countries.

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 can have the undesirable effect that a strong recovery could mask an initial slump, or a downbreak could occur when a country is already in a crisis.

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. We are convinced that an intelligent filtering of Bai-Perron style breaks could allow a sensible classification of slumps, but it requires adding several criteria and appears to yield fewer slumps than our approach. An interesting extension of this line of work would be to combine both approaches and sequentially test whether (a part of) a GDP series exhibits a single downbreak or a sequence of breaks fitting our restrictions.

Figure S3.1 – Examples of similar episodes



Notes: Illustration of slumps obtained using (i) the inverted definition of Berg et al. (2012) and (ii) our two-break model. The depicted GDP per capita series are from the Penn World Tables version 6.3. Solid vertical lines indicate break points and dashed lines indicate the empirical trough.

Table S3.1 – Filtered Downbreaks from Berg et al. (2012): 36 Episodes

Code	T_0	\hat{t}_{b1}	\hat{t}_{min}	\hat{t}_{b2}	T	Drop (%)	\tilde{t}_D	Csrd.
ARE	1970	1980	1988	1986	2006	-51.80	8	1
BDI	1960	1992	2006	–	2006	-36.25	14	1
BEN	1959	1972	1978	1978	2006	-10.72	6	0
BGD	1959	1969	1974	1974	2006	-11.70	5	0
BOL	1950	1981	1986	1986	2006	-19.65	5	0
CHL	1951	1971	1975	1976	2006	-20.87	4	0
CMR	1960	1985	1995	1994	2006	-34.80	10	1
CRI	1950	1978	1982	1983	2006	-17.38	4	0
ETH	1950	1982	1987	1987	2006	-26.46	5	0
GAB	1960	1976	2006	–	2006	-31.26	30	1
GHA	1955	1972	1977	1977	2006	-36.73	5	1
GTM	1950	1980	1988	1987	2006	-19.86	8	0
HTI	1960	1980	2004	1994	2006	-38.98	24	1
IRN	1955	1976	1981	1981	2006	-57.00	5	1
JAM	1953	1973	1985	1980	2006	-28.73	12	1
JOR	1954	1967	1975	1975	2006	-33.59	8	0
JOR	1976	1986	1992	1991	2006	-37.84	6	1
KHM	1970	1970	1982	1987	2006	-51.63	12	0
KWT	1970	1972	1990	1983	2006	-77.01	18	1
LBR	1970	1988	1995	1994	2006	-88.80	7	1
MDG	1960	1972	1982	1982	2006	-17.05	10	0
MEX	1950	1981	1988	1988	2006	-18.20	7	0
MNG	1970	1988	1993	1993	2006	-29.59	5	0
MOZ	1960	1980	1995	1985	2006	-27.91	15	0
MRT	1960	1977	1984	1982	2006	-17.14	7	0
NIC	1950	1976	1993	1986	2006	-57.46	17	1
PHL	1950	1981	1985	1987	2006	-15.19	4	0
SAU	1970	1992	2002	2001	2006	-24.22	10	1
SLE	1961	1987	1999	1999	2006	-60.85	12	1
SLV	1950	1978	1983	1983	2006	-26.85	5	0
SYR	1960	1982	1989	1990	2006	-22.92	7	0
TGO	1960	1979	2002	1986	2006	-47.67	23	1
TTO	1950	1982	1993	1989	2006	-32.89	11	0
UGA	1950	1976	1988	1981	2006	-35.02	12	0
ZAR	1950	1973	2000	1990	2006	-82.05	27	1
ZWE	1954	2000	2006	–	2006	-55.50	6	1

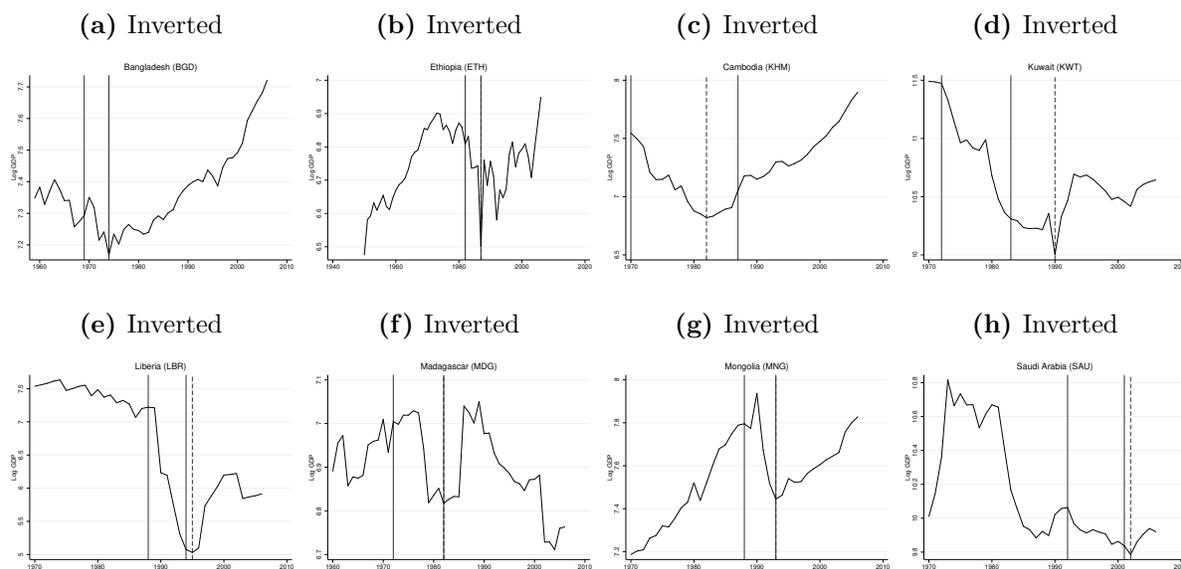
Note(s): There are 55 potential slumps in the Berg et al. (2012) data once we invert their definition of sustained growth. We then apply the same filters used for our data. Of those 55, 14 are in countries with less than one million inhabitants or less than 20 observations of GDP per capita. Of the remaining 41 episodes, 5 occur when a previous slump is still ongoing, that is, GDP per capita in a specific country at a potential second downbreak is below the level of GDP per capita observed at the previous downbreak. The final sample of 36 episodes is listed above. T_0 and T denote the first and last year of the active (sub-)sample, respectively.

Table S3.2 – Depth and duration, by income level and region

	<i>Panel (a) Berg et al. (2012)</i>				<i>Panel (b) Our breaks</i>			
	Mean Depth	Mean Length	Number of Spells	Number of Censored	Mean Depth	Mean Length	Number of Spells	Number of Censored
<i>Income Level</i>								
High Income (OECD)	-46.5	14.5	4	3	-8.7	2.0	12	0
High Income (Other)	-28.9	16.2	6	3	-39.5	19.7	3	2
Upper Middle Income	–	–	–	–	-22.5	7.4	16	3
Lower Middle Income	-29.8	8.7	10	3	-15.6	3.1	10	1
Low Income	-40.3	16.1	16	8	-28.0	10.3	9	2
<i>Geographical Region</i>								
Africa	-40.5	19.0	15	9	-27.5	10.9	13	3
Americas	-28.1	11.6	10	3	-15.5	5.1	14	1
Asia and Oceania	-37.5	10.4	11	5	-25.7	11.2	11	4
Europe	–	–	–	–	-11.2	2.5	12	0
<i>Total</i>	-36.2	16.0	36	17	-19.8	7.3	50	8

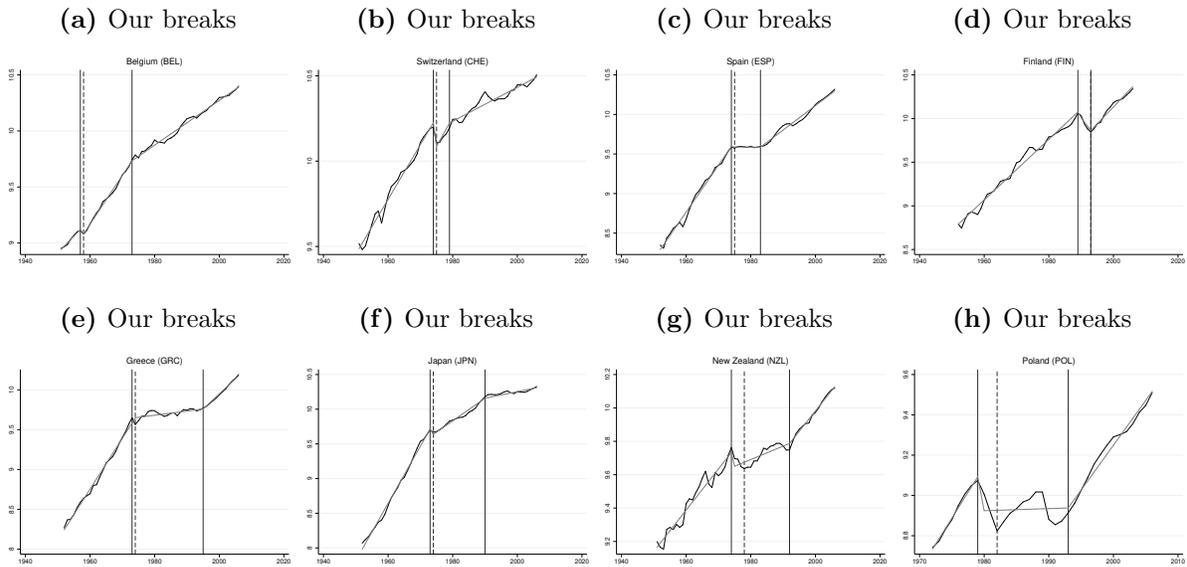
Note(s): Depth is defined as the percent decrease in GDP per capita at the trough relative to GDP per capita before the slump (not log difference). Mean duration is expressed in years. As a result of some spells being censored, both mean duration and depth are underestimated. The number of countries 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.

Figure S3.2 – Examples of questionable episodes



Notes: Illustration of slumps obtained using the inverted definition of Berg et al. (2012). The depicted GDP per capita series are from the Penn World Tables version 6.3. Solid vertical lines indicate break points and dashed lines indicate the empirical trough.

Figure S3.3 – Examples of “missing” slumps in OECD countries



Notes: Illustration of slumps obtained using our two-break model. The depicted GDP per capita series are from the Penn World Tables version 6.3. Solid vertical lines indicate break points and dashed lines indicate the empirical trough.

S4 Variables and summary statistics

Table S4.1 – Sources and summary statistics: pre-slump values, larger sample

VARIABLE	Mean	Std. Dev.	$N \times T$	Source
Polity IV Score	-2.19	7.41	489	Polity IV
Democracy Score	2.74	3.92	489	Polity IV
Autocracy Score	4.93	3.67	489	Polity IV
Executive Recruitment	4.76	2.40	489	Polity IV
Executive Constraints (INS_0)	3.27	2.28	489	Polity IV
Political Competition	3.85	3.52	489	Polity IV
Regime Durability	21.05	24.45	501	Polity IV
Fractionalization ($ELF15$)	53.47	33.05	513	Desmet et al. (2012)
US Real Interest Rate ^a	1.08	2.63	513	FRED
Initial log GDP ^b	10.50	1.94	513	PWT 9.0
Policy Volatility ^c	8.66	6.51	513	PWT 9.0
Government Size	22.10	13.38	513	PWT 9.0
Inflation ($\ln(1 + \delta)$)	17.46	38.62	345	WDI/IFS
RER Underval ^e	0.03	0.45	513	PWT 9.0
Export Share	18.58	19.45	513	PWT 9.0
Trade Openness (de facto)	60.68	38.55	513	PWT 9.0
Investment Price	0.41	0.28	513	PWT 9.0
Trade Openness (de jure)	0.22	0.42	462	Wacziarg and Welch (2008)
Leader Exit	0.04	0.19	513	Goemans et al. (2009) - Archigos 4.1
War/Conflict (any)	0.26	0.44	513	Gleditsch et al. (2002)
Life Expectancy ^f	57.34	11.79	505	World Population Prospects
Education (All) ^g	4.44	3.14	445	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.

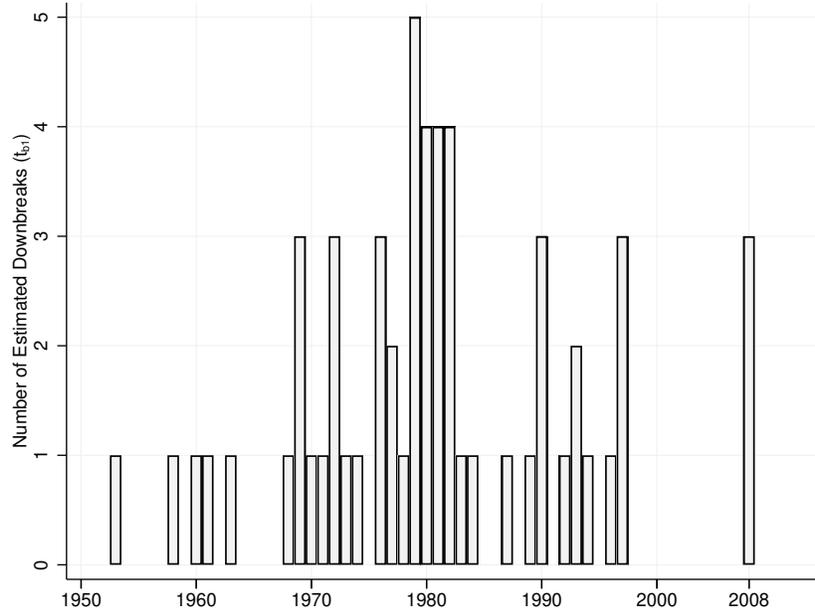
^e Following Rodrik (2008), we compute the index of real exchange rate undervaluation as the difference between $\ln RER_{it} - \ln \widehat{RER}_{it}$, where $\ln RER_{it} = \ln(XRAT_{it}/PPP_{it})$, and $\ln \widehat{RER}_{it} = \alpha + \beta \ln y_{it} + \lambda_t + e_{it}$ from a pooled panel regression. 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 1952 and linearly interpolate to the middle of the next group (1957), and so on. The data is from the 2010 edition of the World Population Prospects (medium-fertility variant).

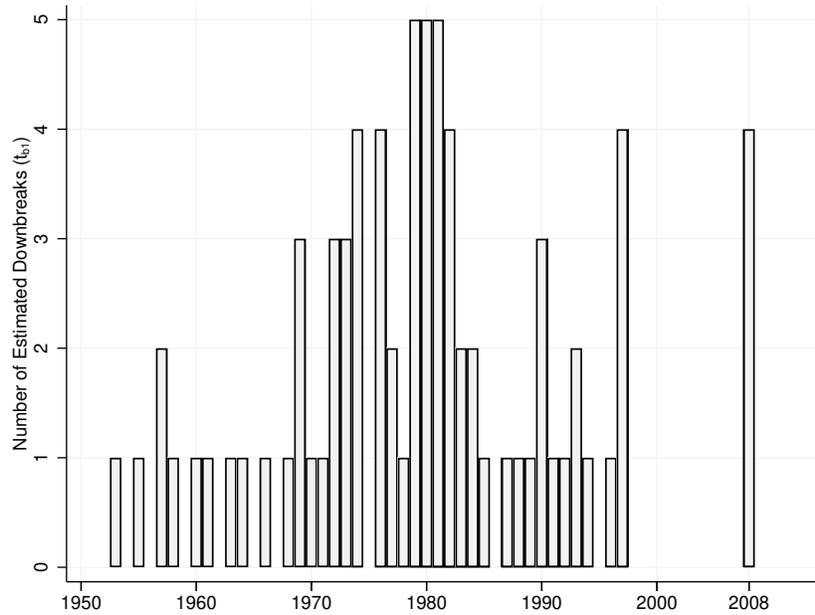
^g Converted into annual data by linear interpolation.

Figure S4.1 – Distribution of starting dates

(a) Structural breaks estimated with size = 0.10



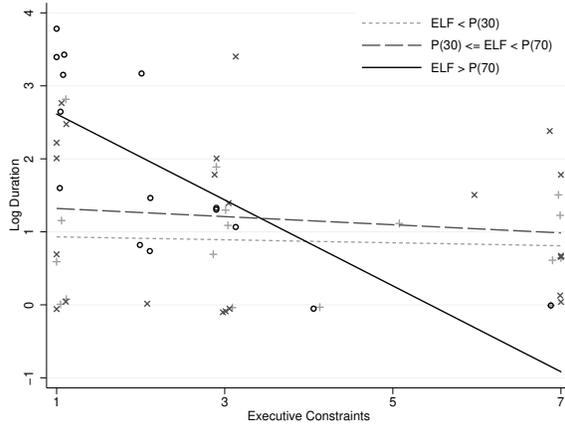
(b) Structural breaks estimated with size = 0.20



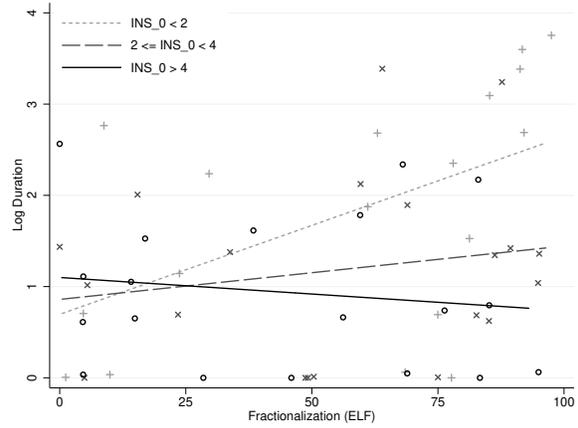
S5 Split sample plots

Figure S5.1 – Linear associations in the raw data, split by categories of ...

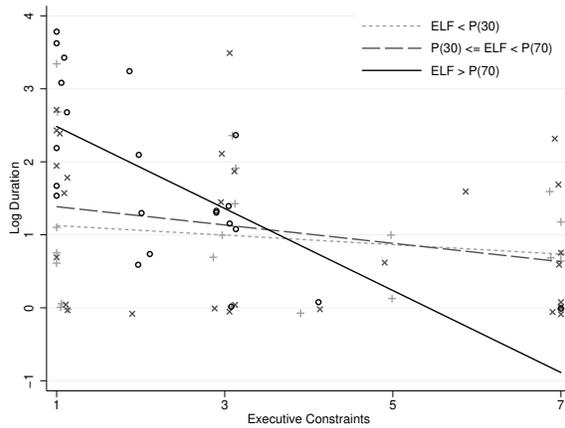
(a) Fractionalization, size = 0.10



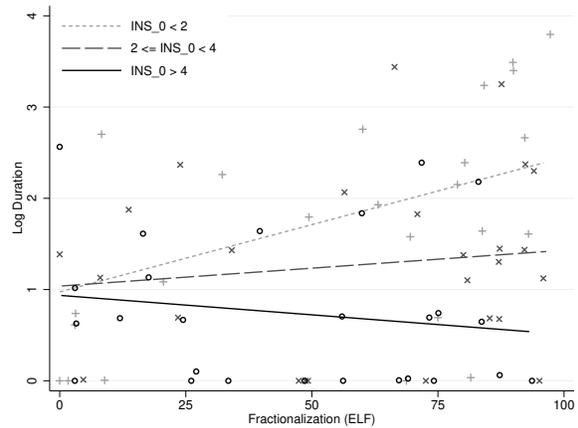
(b) Executive Constraints, size = 0.10



(c) Fractionalization, size = 0.20

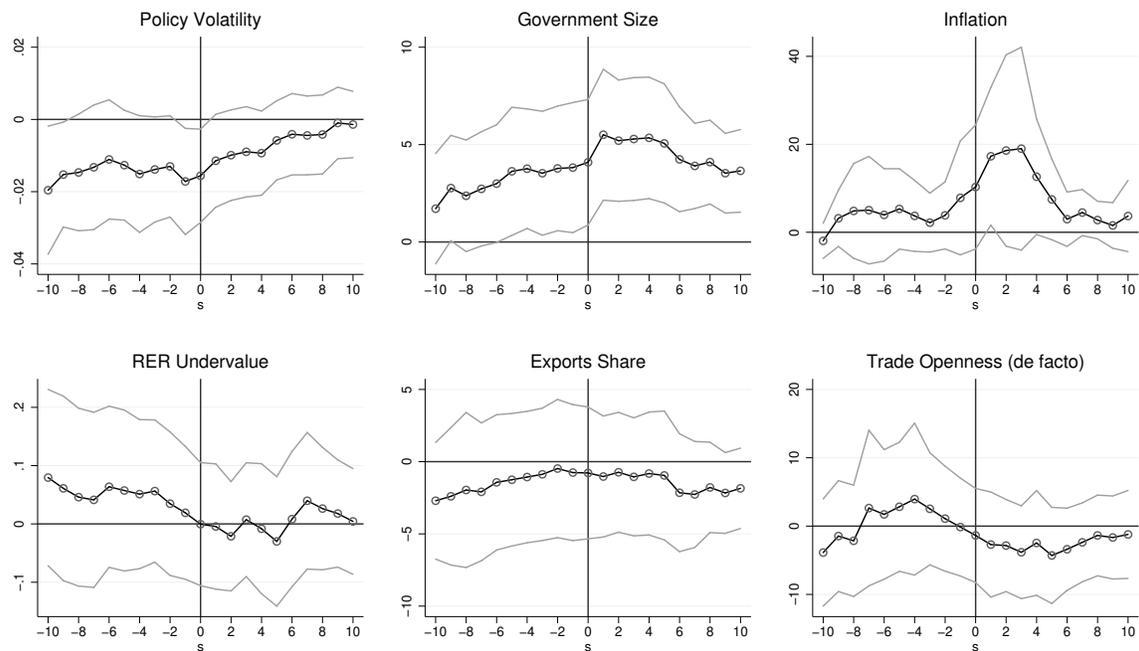


(d) Executive Constraints, size = 0.20



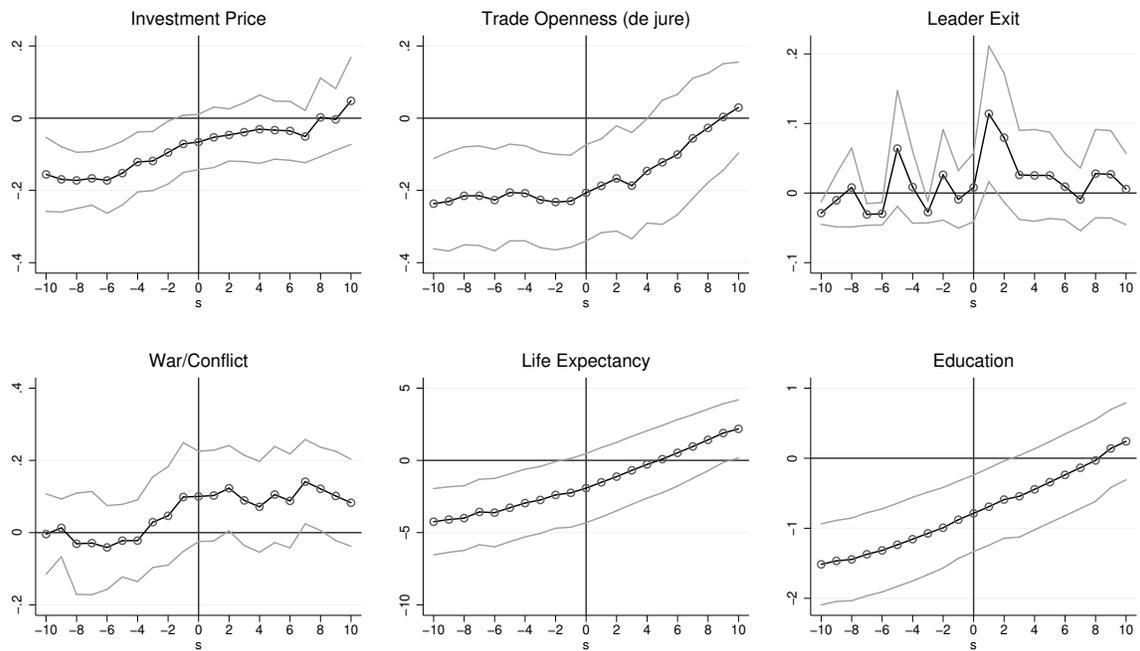
S6 Additional event plots

Figure S6.1 – 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 S6.2 – 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).

S7 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 S7.1 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 built 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 built recursively but the resampled and rescaled residuals are multiplied by Rademacher's two-point distribution.

Table S7.2 varies the minimum period required before the restricted break search algorithm can find a second break.

Table S7.1 – 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)
<i>Panel (a) Structural breaks estimated with size = 0.10</i>						
Executive Constraints (\widetilde{INS}_0)	-0.255*** (0.077)	-0.215*** (0.056)	-0.189*** (0.053)	-0.352*** (0.093)	-0.150** (0.065)	-0.249*** (0.090)
Fractionalization ($\widetilde{ELF15}$)	0.020*** (0.005)	0.017*** (0.004)	0.017*** (0.004)	0.019*** (0.007)	0.009** (0.005)	0.011 (0.007)
$\widetilde{INS}_0 \times \widetilde{ELF15}$	-0.004*** (0.001)	-0.004*** (0.001)	-0.003*** (0.001)	-0.004** (0.002)	-0.003* (0.002)	-0.003 (0.002)
Exits	43	53	43	29	69	22
Spells	54	62	51	37	79	29
<i>Panel (b) Structural breaks estimated with size = 0.20</i>						
Executive Constraints (\widetilde{INS}_0)	-0.239*** (0.062)	-0.198*** (0.065)	-0.273*** (0.065)	-0.338*** (0.081)	-0.128** (0.054)	-0.202** (0.080)
Fractionalization ($\widetilde{ELF15}$)	0.014*** (0.005)	0.015*** (0.004)	0.017*** (0.004)	0.019*** (0.006)	0.006 (0.004)	0.001 (0.006)
$\widetilde{INS}_0 \times \widetilde{ELF15}$	-0.003** (0.001)	-0.004*** (0.001)	-0.005*** (0.001)	-0.005*** (0.002)	-0.003** (0.001)	-0.000 (0.002)
Exits	61	71	61	43	91	41
Spells	74	84	73	53	107	50

Notes: All models include region FEs, the US real interest rate, the log of 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 S7.2 – Different interbreak periods for structural breaks

	<i>Interbreak period (h)</i>					
	<i>h</i> = 2 (1)	<i>h</i> = 3 (2)	<i>h</i> = 4 (3)	<i>h</i> = 5 (4)	<i>h</i> = 6 (5)	<i>h</i> = 7 (6)
<i>Panel (a) Structural breaks estimated with size = 0.10</i>						
Executive Constraints (\widetilde{INS}_0)	-0.309*** (0.081)	-0.267*** (0.076)	-0.255*** (0.077)	-0.284*** (0.079)	-0.262*** (0.071)	-0.306*** (0.087)
Fractionalization (\widetilde{ELF}_{15})	0.021*** (0.005)	0.020*** (0.005)	0.020*** (0.005)	0.020*** (0.005)	0.018*** (0.005)	0.019*** (0.005)
$\widetilde{INS}_0 \times \widetilde{ELF}_{15}$	-0.005*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.002)	-0.003*** (0.001)	-0.004*** (0.001)
Exits	47	46	43	40	45	40
Spells	60	58	54	51	57	50
<i>Panel (b) Structural breaks estimated with size = 0.20</i>						
Executive Constraints (\widetilde{INS}_0)	-0.263*** (0.061)	-0.250*** (0.063)	-0.239*** (0.062)	-0.227*** (0.061)	-0.223*** (0.061)	-0.265*** (0.071)
Fractionalization (\widetilde{ELF}_{15})	0.015*** (0.005)	0.012** (0.005)	0.014*** (0.005)	0.013*** (0.004)	0.013** (0.005)	0.015*** (0.005)
$\widetilde{INS}_0 \times \widetilde{ELF}_{15}$	-0.004** (0.001)	-0.003* (0.002)	-0.003** (0.001)	-0.004*** (0.001)	-0.003* (0.001)	-0.003** (0.001)
Exits	65	60	61	61	57	57
Spells	79	73	74	74	70	68

Notes: All models include region FEs, the US real interest rate, the log of 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 S7.3](#) 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 S7.4](#) 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 S7.3 – Different measures of institutions

	<i>Measure of political institutions (X_0)</i>					
	Executive Recruitment (1)	Political Competition (2)	Autocracy (3)	Democracy (4)	Polity Score (5)	Regime Durability (6)
<i>Panel (a) Structural breaks estimated with size = 0.10</i>						
Political Institutions (\widetilde{X}_0)	-0.135 (0.086)	-0.107 (0.074)	0.131** (0.058)	-0.106** (0.046)	-0.063** (0.026)	-0.002 (0.008)
Fractionalization (\widetilde{ELF})	0.014*** (0.005)	0.016*** (0.005)	0.017*** (0.005)	0.017*** (0.005)	0.017*** (0.005)	0.014** (0.005)
$\widetilde{X}_0 \times \widetilde{ELF}$	0.000 (0.002)	-0.001 (0.001)	0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	0.000 (0.000)
Exits	43	43	43	43	43	43
Spells	54	54	54	54	54	54
<i>Panel (b) Structural breaks estimated with size = 0.20</i>						
Political Institutions (\widetilde{X}_0)	-0.135* (0.070)	-0.105** (0.053)	0.127*** (0.046)	-0.098*** (0.036)	-0.060*** (0.021)	-0.000 (0.006)
Fractionalization (\widetilde{ELF})	0.008* (0.005)	0.010* (0.005)	0.012** (0.005)	0.010** (0.005)	0.011** (0.005)	0.007 (0.005)
$\widetilde{X}_0 \times \widetilde{ELF}$	0.001 (0.002)	-0.001 (0.001)	0.001 (0.001)	-0.001 (0.001)	-0.000 (0.001)	0.000 (0.000)
Exits	61	61	61	61	61	61
Spells	74	74	74	74	74	74

Notes: All models include region FEs, the US real interest rate, the log of 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 S7.4 – Additional policy variables and other determinants

	<i>Added variable</i>					
	Investment Price (1)	Trade Openness (d.j.) (2)	Leader Exit (3)	War/ Conflict (4)	Life Expectancy (5)	Education (6)
<i>Panel (a) Structural breaks estimated with size = 0.10</i>						
Executive Constraints (\widetilde{INS}_0)	-0.255*** (0.077)	-0.221*** (0.072)	-0.263*** (0.077)	-0.273*** (0.085)	-0.292*** (0.076)	-0.306*** (0.090)
Fractionalization ($\widetilde{ELF}15$)	0.020*** (0.005)	0.023*** (0.005)	0.020*** (0.005)	0.021*** (0.005)	0.022*** (0.004)	0.023*** (0.005)
$\widetilde{INS}_0 \times \widetilde{ELF}15$	-0.004*** (0.001)	-0.003** (0.001)	-0.004*** (0.001)	-0.004*** (0.002)	-0.002** (0.001)	-0.004*** (0.001)
Added Variable	0.002 (0.919)	-0.229 (0.320)	-0.816 (0.666)	-0.251 (0.363)	0.054*** (0.016)	0.117* (0.070)
Exits	43	39	43	43	41	40
Spells	54	47	54	54	51	49
<i>Panel (b) Structural breaks estimated with size = 0.20</i>						
Executive Constraints (\widetilde{INS}_0)	-0.243*** (0.062)	-0.192*** (0.059)	-0.260*** (0.064)	-0.244*** (0.065)	-0.279*** (0.067)	-0.298*** (0.077)
Fractionalization ($\widetilde{ELF}15$)	0.014*** (0.005)	0.014*** (0.005)	0.014*** (0.005)	0.014*** (0.005)	0.015*** (0.005)	0.016*** (0.005)
$\widetilde{INS}_0 \times \widetilde{ELF}15$	-0.003** (0.001)	-0.002 (0.001)	-0.004*** (0.001)	-0.004** (0.002)	-0.002* (0.001)	-0.004*** (0.001)
Added Variable	0.334 (0.539)	-0.524* (0.275)	-0.958** (0.444)	-0.151 (0.291)	0.042*** (0.014)	0.083 (0.060)
Exits	61	56	61	61	58	57
Spells	74	66	74	74	70	68

Notes: All models include region FEs, the US real interest rate, the log of 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 S7.5 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 positive in panel (a), so that the hazard is monotonically decreasing, and negative in panel (b), 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, flat, or decreasing hazard rates. The estimated shape parameter suggests that the baseline hazard is flat over time. In contrast, the Gompertz model in column (5) suggests a shape that is monotonically decreasing. Among these parametric models, the log-likelihood is highest and the AIC is second-lowest for the log-normal distribution, our preferred model. In column (6), we specify a semi-parametric Cox model which does not restrict the shape of the baseline hazard. In this instance, we lose significance on the interaction term, but note that this is not the case when we use GDP to determine the slumps, or GDP per capita from any of the other national accounts series considered earlier.

Table S7.5 – Varying the functional form

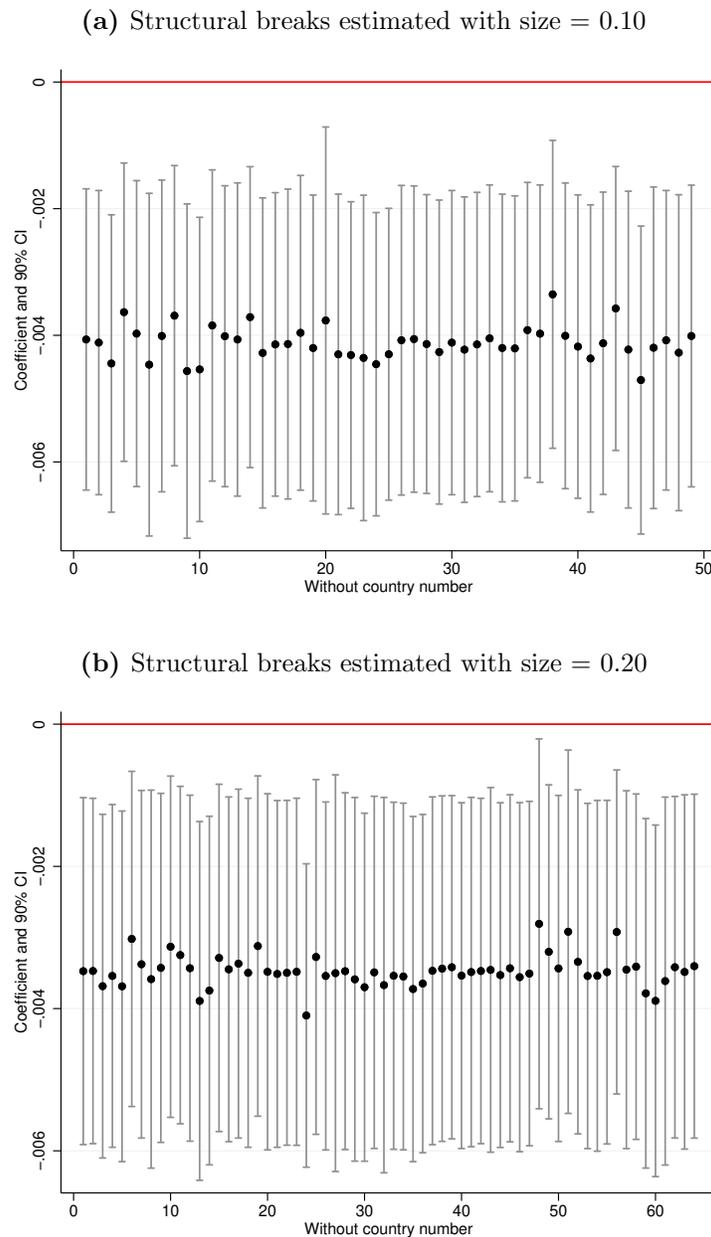
	Coefficients		Hazard Ratios ($\mathbb{H}_0 : \text{HR} = 1$)			
	Survival model					
	Log-normal (1)	Log-logistic (2)	Exponential (3)	Weibull (4)	Gompertz (5)	Cox PH (6)
<i>Panel (a) Structural breaks estimated with size = 0.10</i>						
Executive Constraints (\widetilde{INS}_0)	-0.255*** (0.077)	-0.233** (0.092)	1.374*** (0.115)	1.379*** (0.118)	1.299*** (0.093)	1.202*** (0.078)
Fractionalization ($\widetilde{ELF15}$)	0.020*** (0.005)	0.015*** (0.005)	0.978*** (0.006)	0.978*** (0.007)	0.983*** (0.005)	0.991* (0.005)
$\widetilde{INS}_0 \times \widetilde{ELF15}$	-0.004*** (0.001)	-0.004** (0.002)	1.004** (0.002)	1.004** (0.002)	1.003* (0.002)	1.002 (0.002)
Shape Parameter	0.063 (0.113)	0.385*** (0.118)		1.015 (0.105)	0.928** (0.032)	
Exits	43	43	43	43	43	43
Spells	54	54	54	54	54	54
Years of Decline	383	383	383	383	383	383
Log- \mathcal{L}	-71.978	-75.363	-76.961	-76.954	-74.475	-136.044
AIC	165.955	164.726	173.922	175.909	170.951	282.087
<i>Panel (b) Structural breaks estimated with size = 0.20</i>						
Executive Constraints (\widetilde{INS}_0)	-0.239*** (0.062)	-0.241*** (0.065)	1.352*** (0.102)	1.356*** (0.106)	1.278*** (0.080)	1.229*** (0.072)
Fractionalization ($\widetilde{ELF15}$)	0.014*** (0.005)	0.011** (0.005)	0.986** (0.007)	0.986* (0.007)	0.988** (0.005)	0.995 (0.005)
$\widetilde{INS}_0 \times \widetilde{ELF15}$	-0.003** (0.001)	-0.004** (0.002)	1.003 (0.002)	1.003 (0.002)	1.002 (0.002)	1.002 (0.002)
Shape Parameter	0.045 (0.101)	-0.442*** (0.115)		1.013 (0.081)	0.927*** (0.024)	
Exits	61	61	61	61	61	61
Spells	74	74	74	74	74	74
Years of Decline	489	489	489	489	489	489
Log- \mathcal{L}	-101.201	-104.426	-108.430	-108.423	-104.301	-211.731
AIC	224.402	222.853	236.860	238.846	230.603	433.462

Notes: All models include region FEs, the US real interest rate, the log of 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 S7.1 shows 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 90% confidence interval.

Figure S7.1 – Leave one out graph of the interaction coefficient



Notes: Illustration of the robustness of the interaction effect to dropping individual countries. All underlying models include region FEs, the US real interest rate, the log of 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. Working Paper 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.
- Ferguson, J. (1999). *Expectations of modernity: Myths and meanings of urban life on the Zambian Copperbelt*. Berkeley, CA: University of California Press.
- Fraser, A. (2010). Introduction: Boom and bust on the Zambian Copperbelt. In A. Fraser and M. Larmer (Eds.), *Zambia, Mining, and Neoliberalism: Boom and Bust on the Globalized Copperbelt*, pp. 1–30. New York: Palgrave Macmillan US.
- 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.
- Larmer, M. (2008). Enemies within? Opposition to the Zambian one-party state, 1972–1980. In J.-B. Gewald, M. Hinfelaar, and G. Macola (Eds.), *One Zambia, Many Histories*, pp. 98–125. Leiden, The Netherlands: Brill.
- 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.
- McCulloch, N., B. Baulch, and M. Cherel-Robson (2000). Poverty, inequality and growth in Zambia during the 1990s. Working Paper 114, Institute of Development Studies.
- Papell, D. H. and R. Prodan (2014). Long-run time series tests of constant steady-state

- growth. *Economic Modelling* 42, 464–474.
- Posner, D. N. (2005). *Institutions and ethnic politics in Africa*. Cambridge University Press.
- Rakner, L. (2003). *Political and economic liberalisation in Zambia 1991-2001*. Stockholm: Nordic Africa Institute.
- Rodrik, D. (2008). The real exchange rate and economic growth. *Brookings Papers on Economic Activity* 2008(2), 365–412.
- Szirmai, A., F. Yamfwa, and C. Lwamba (2002). *Zambian manufacturing performance in comparative perspective*. GGDC research memorandum No. 200253, Groningen Growth and Development Centre, University of Groningen.
- Wacziarg, R. and K. H. Welch (2008). Trade liberalization and growth: New evidence. *World Bank Economic Review* 22(2), 187–231.