Do Weak Institutions Prolong the Fall?

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

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

February 2013

Draft Version
Do Not Quote or Circulate Without Permission

Abstract

This paper analyzes and discusses the identification, characteristics, and duration of the decline phase during economic slumps. To identify slumps and their associated declines, we employ a restricted structural change approach and then apply this method to a large sample of countries. We find three major results. First, slumps occur frequently and in many cases the decline phase lasts very long, particularly in Sub-Saharan Africa, the Middle East and North Africa, and Latin America and the Caribbean. Second, we find strong evidence of institutional underdevelopment before a slump hits and a trend towards positive institutional reform thereafter. Third, the duration of declines decreases with stronger institutions but increases with greater degrees of ethnic cleavages. We provide additional evidence of a non-linear effect suggesting that institutions can potentially help to overcome even the most negative effects of high fractionalization.

Keywords: growth, crises, institutions, structural breaks, duration analysis

JEL Classification: O41, O11, C41, F43

[*]We gratefully acknowledge financial support from the Agence Française de Développement (AFD). The findings, interpretations and conclusions expressed in this paper are solely that of the authors and do not necessarily represent policies or views of the Maastricht Graduate School of Governance, UNU-MERIT, AFD and/or other affiliated institutions. All remaining errors are those of the authors.

[†]UNU-MERIT, Maastricht University, Maastricht Graduate School of Governance, e-mail: *richard.bluhm@maastrichtuniversity.nl*

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

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

1 Introduction

The last seventy years of growth have been far from steady. For every inspiring story about "growth miracles" we can easily find daunting counterparts in the form of "miraculous collapses". For example, the East Asian miracle was interrupted by the Asian financial crisis, China's take-off in 1978 was preceded by decades of disastrous economic policies, Latin America was frequently rocked by political turmoil and economic volatility, and several African nations went from "up and coming" in the 1950s to requiring outside assistance within a few years. Moreover, in the post-war period, there is a long list of relatively short developed country crises including the first global recession in 1957, the global oil crises in 1973–74, and the Nordic banking crisis of the 1990s. What can we learn from such abrupt changes in growth? Do some countries deal better with negative growth shocks than others? Does vulnerability to crises contribute to long-run divergence?

Analyzing this instability of growth is not new. A large and growing literature on trend breaks has established that most growth performances are not steady but instead marked by switching between very different growth regimes. In this view, growth is no longer defined by a single average trend but consists of many qualitatively different episodes, such as crises, recoveries, stagnation, slows downs, and accelerations. This non-linear perspective provides better insights into the underlying dynamics and has established new stylized facts. For example, in developed and developing countries alike, growth is relative easy to ignite (Hausmann et al., 2005) but much harder to sustain (Berg et al., 2012). However, the negative implications of unsteady growth paths are just beginning to be explored. Long-lasting slumps can nullify decades of positive growth and there is no guarantee that lost potential output after a slump is ever fully recouped (Cerra and Saxena, 2008; Reddy and Minoiu, 2009). It thus becomes important to ask, why do some declines last so much longer than others?

A potential answer to this question is that the duration of declines during slumps is driven by the prevailing structure and quality of institutions. Seminal contributions to the institutions and growth literature link stronger institutions to higher *levels* of GDP per capita (Acemoglu et al., 2001, 2002) and other papers show how strong institutions, democracy and political stability bring about reduced output volatility (Acemoglu et al., 2003; Mobarak, 2005; Klomp and de Haan, 2009). However, there is still a lack of evidence linking institutions to more specific short and medium-run growth dynamics.

Each type of growth episode has several characteristics, and institutions – together with many other factors – may play a different role in each one of them. Splitting growth into different episodes implies that we can analyze the switching among growth episodes, the rate of growth within an episode, the duration of an episode, and even the typical sequence of episodes that makes up a growth path. Out of this plethora of possibilities, this paper focuses on three points. First, how can we identify episodes of large economic slumps empirically? Second, what happens when slumps occur? Specifically, is there any evidence of institutional change? And, third, what determines the duration of the decline phase? In particular, do weak institutions prolong the fall?

The notion that weak institutions prolong the decline derives from a large body of political economy theory putting social conflict and the ability of resilient institutions to manage such conflict at the center of development theory. Some of these theories argue that weakly institutionalized societies, or limited access orders, are prone to collapses and that the declining rents resulting from economic crises further strain the institutional set-up and the prevailing political arrangements (e.g. North et al., 2009). Weak institutions thus bring with them an increased vulnerability to crises and potentially much longer declines. Similar mechanisms are suggested in the literature on institutions and macroeconomic volatility. Acemoglu et al. (2003), for example, argue that institutions determine "whether there will be significant swings in the political and social environment leading to crises, and whether politicians will be induced to pursue unsustainable policies in order to remain in power in the face of deep social cleavages." (p. 54). So even if better policy responses are available, a combination of coordination failures, rent seeking, power struggles and dormant social conflict may lead to longer declines in weakly institutionalized environments. Hence, the interplay of institutions and conflict plays out at a "deeper" level than more proximate responses to crises.

Our findings broadly support this theoretical perspective. First, we find very robust evidence of institutional underdevelopment before a slump hits the economy and a clear trend towards broad institutional reforms during slumps, as well as in their immediate aftermath. Second, longer decline phases are robustly linked to weak institutions and particularly strongly to a measure of ethnic cleavages (ethno-linguistic fractionalization). Ethnic cleavages are especially important for understanding declines in Sub-Saharan Africa. Third, we find that institutions and fractionalization interact negatively. Our models predict longer declines in more fractionalized societies with weaker institutions.

The remainder of this paper is structured as follows. Section 2 motivates and outlines the restricted structural change approach used to identify slumps and defines the duration of the decline phase. Section 3 provides descriptive statistics of the estimated slumps and very briefly discusses the data used in the ensuing analysis. Section 4 investigates the characteristics of slumps and the evolution of covariates before, during and after a slump occurs. Section 5 analyzes the duration of the decline phase and provides a substantive interpretation of the main results. Section 6 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 large and growing empirical literature sets out to investigate the characteristics of different types of growth episodes. Many of these papers employ either simple or more complex tests of structural stability to define and identify their episode of interest. For example, Hausmann et al. (2005) use economic criteria to isolate growth accelerations and then date their beginning with a very simple breakpoint test. Other authors, such as Jones and Olken (2008) and Berg et al. (2012), use versions of the much more sophisticated Bai and Perron (1998, 2003) test for multiple unknown change points to classify different growth episodes. A third set of papers solely relies on economic criteria to identify and date the episode of interest (e.g. Calvo et al., 2006; Hausmann et al., 2008; Reddy and Minoiu, 2009).

Not every change in a certain direction amounts to a regime switch. The main advantage of econometric tests for multiple structural breaks over any set of pre-defined economic criteria is that they allow for an inferential approach to identifying growth accelerations, slow-downs or declines. However, since the particular type of structural change is left unspecified, these tests may not select the theoretically desired type of episode but rather any form of significant change. Furthermore, while break estimators work well for identifying growth spurts, they perform poorly when it comes to identifying recessions or growth collapses.[1] Methods based solely on economic criteria, on the contrary, are transparently designed but cannot discriminate among multiple possible starting points or assess whether a particular episode truly constitutes a departure from the previous regime.

To improve the identification of what we interchangeably refer to as deep recessions, slumps, or growth collapses, Papell and Prodan (2011b) suggest a specific *two-break model with restrictions* and demonstrate that this modified structural change approach consistently identifies well-known slumps, such as the Great Depression in the United States. The key innovation is to no longer allow for unrestricted structural change but to impose features of the desired pattern directly. Their two-break model accounts for three regimes (a pre-slump regime, a contraction/ recovery regime, and a post-slump regime) and places sign restrictions onto the estimated coefficients to ensure the breaks occur in the desired direction. Since this approach is a version of Bai's (1999) sequential likelihood ratio test, the number of slumps, which is not known in advance, can then be estimated by recursively applying the model on ever smaller sub-samples until all breaks in the GDP per capita series have been found.

This approach can easily be modified to allow for other plausible structures, such as three break models (e.g., to estimate a pre-slump regime, a decline, a recovery and a post-slump regime). However, estimating three or more breaks for each slump quickly

[1] The main problem with standard techniques for multiple change-points of unknown type is that they may not date the beginning of the collapse very precisely or, as stated above, identify an entirely different pattern. As Papell and Prodan (2011b) point out, this applies quite generally to tests for a single break (Andrews, 1993), tests for multiple endogenous breakpoints (Bai and Perron, 1998, 2003) and tests for multiple endogenous breaks with *linear* restrictions on the parameters (Perron and Qu, 2006).

becomes computationally expensive[2] and does not necessarily provide better results than the simpler two-break model in terms of identifying well-known episodes.[3] While Papell and Prodan (2011b) focused on testing whether growth after a slump returns to its trend path for a few select countries over the long run, this paper applies their method to a large sample of countries over the entire post-war period from 1950 to 2008.

We define slumps according to three criteria. First, a slump is a *departure from a previously positive trend*. Second, a slump must begin with *negative growth in the first year*. Third, all slumps should be *pronounced regime switches* and not just minor business cycle fluctuations. To capture these three desiderata, we specify the following partial structural change model:

$$y_t = \alpha + \beta t + \gamma_1 \mathbf{1}(t > tb_1) + \gamma_{12}(t - tb_1)\mathbf{1}(t > tb_1) + \gamma_2(t - tb_2)\mathbf{1}(t > tb_2) + \sum_{i=1}^{p} \delta_i y_{t-i} + \epsilon_t \quad (1)$$

where y_t is the log of GDP per capita, β is a time trend, γ_1 is the coefficient on an intercept break occurring together with a trend change (γ_{12}) after the first break at time tb_1, γ_2 is the coefficient for a second trend change occurring after the second break at time tb_2, $\mathbf{1}(\cdot)$ is an indicator function selecting the regime, p is the optimal lag order determined by the Schwarz/ Bayesian information criterion (BIC) to parametrically adjust for the presence of serial correlation, and ϵ_t is a martingale sequence with $\mathrm{E}[\epsilon|\{y_{t-1}, y_{t-2}, \dots\}] = 0$.

The model in equation (1) formalizes the notion that the evolution of GDP per capita around a slump is a simple function of time split into three different growth regimes: (1) a pre-slump regime from the beginning of the time series of a country until time tb_1, (2) a slump/recovery regime lasting from time $tb_1 + 1$ to time tb_2, and (3) a post slump regime from time $tb_2 + 1$ onwards. The true location of the breakpoints is not known in advance but estimated within the model. We impose two restrictions to make sure

[2] If we define $q = T - 2\tau T - h$, where τ is the trimming fraction and h is the distance between breaks, then the two break model estimates $(q^2 + q)2^{-1}$ regressions on the first iteration, while a three break model already requires $\sum_{i=1}^{q}\left((q^2+q)2^{-1}\right)$ regressions, with $q = T - 2\tau T - 2h$ to now allow for three instead of two breaks. Since these models are first estimated sequentially and then invoked again during each bootstrap iteration, the computational burden increases from a few hours to a few days.

[3] Before settling on the two break model, we also estimated a three break model with the additional break serving to directly estimate the location of the trough. While such a model is still computationally feasible, many of the resulting episodes were also found by the two break model and the turning points were generally estimated less precisely.

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 $\gamma_1 < 0$, so that a slumps always starts with a drop in the intercept.[4] The slope shifts during and after the slump are left unrestricted, so that the model can also catch unfinished slumps (e.g., declines from tb_1 onwards, possibly lasting until the end of a country's time series).

We implement the sequential break search algorithm as follows. First, we fit the structural change model specified in equation (1) for all possible combinations of tb_1 and tb_2. We always exclude 5% of the observations at the beginning and end of the sample to avoid finding spurious breaks. Second, we compute the sup-W test statistic, that is, the maximum/supremum of a Wald test of the null hypothesis of no structural change ($\mathbb{H}_0 : \gamma_1 = \gamma_{12} = \gamma_2 = 0$) over all possible combinations of break dates satisfying the two restrictions. Third, we bootstrap the empirical distribution of the sup-W statistic (see below). If the bootstrap test rejects at the 10% significance level, we record the break pair and split the sample into a series running until the first break and a series starting just after the second break. The process starts again on each sub-sample until the bootstrap test fails to reject the null hypothesis of no breaks or the sample gets too small.[5] As Bai (1997) has shown, this procedure converges to the true number of breaks.

A key issue in evaluating the statistical significance of endogenous breakpoints is that the individual Wald tests (or F-tests) over which sup-W statistic is computed are not statistically independent. Assuming that there are in fact breaks present in the series, the closer the estimated break dates get to the true breakpoints, the higher the test statistic will be, and vice versa. For several single and multiple change-point problems, the limiting distribution of the sup-W statistic or similar test statistics taking this dependency into account has been derived (Andrews, 1993; Andrews and Ploberger, 1994; Bai, 1997, 1999; Bai and Perron, 1998; Hansen, 2000). However, asymptotic tests tend to under-reject in finite samples (Prodan, 2008) and an asymptotic distribution for our particular version

[4]In reality the decline of GDP may not be immediate but may take some time. The restriction with regard to the intercept shift implies that we assume that there is an instantaneous drop. However, by not restricting the coinciding trend break we relax this assumption and also allow for longer lasting declines.

[5]We stop when $T \leq 20$ to avoid finding spurious breaks.

of restricted structural change is not available. To circumvent both issues, we construct a bootstrap Monte Carlo test as follows. We first estimate the optimal AR(p) model under the null hypothesis of no structural change. Then, we randomly draw new errors from a standard normal distribution with variance equal to that of the residuals estimated by the optimal model under the null, so that $\hat{e}_t^* = u_t$ and $u_t \sim \mathcal{N}(0, \sigma_{\hat{e}}^2)$. Next, we recursively construct a bootstrap sample of equal size to the original series based on the estimated parameters together with the new error series ($\hat{e}_t^*$). Using this bootstrap series ($\hat{y}_t^*$), we then re-run the break search algorithm and compute the sup-W statistic in exactly the same manner as before. This process is repeated 1000 times. At the 10% significance level, the critical value for each estimated sup-W statistic is then located at the 90$^{\text{th}}$ percentile of a vector of containing all recorded bootstrapped sup-W statistics sorted in ascending order. Appendix A gives a more formal description of the break search algorithm and the bootstrap procedure.

The structural break method applied in this paper assumes that GDP per capita is a regime-wise trend stationary process. This is not a trivial requirement. Ever since the issue was first raised by Nelson and Plosser (1982), a vibrant literature debates the question whether most GDP series have a unit root or can be considered trend stationary. Originally, the conflicting views evolved around a clear divide. If an output series is non-stationary, i.e., it has a unit root, then any shock to the economy is permanent. If the series is trend stationary, then shocks are temporary and GDP ultimately recovers to continue on the trend path (that is, after some time GDP is on the same path as if the shock never occurred).[6] More recently, however, the debate has shifted. A process that is subject to structural breaks presents an intermediate case. Broken trend stationarity only implies that within each regime growth is assumed to follow a trend, but from one regime to the next the trend path may change due to (semi-) permanent shocks such as big recessions, growth accelerations or growth slow-downs. In fact, there is mounting evidence

[6] This is easily illustrated. A unit root process, such as a random walk with drift, can be written as $y_t = y_{t-1} + \mu + \epsilon_t$, whereas a trending process is represented by $y_t = \beta t + \epsilon_t$. A random shock ϵ_t is incorporated permanently in the unit root process but not in the trending process. However, this simple distinction is embedded in the larger question of the degree of fractional integration and existence of long-memory in GDP series, see Silverberg and Verspagen (2003) for an empirical test of long-memory and a discussion of the theoretical mechanisms needed to generate different types of series.

that once trend breaks are incorporated, many of the GDP series previously thought to have unit roots may in fact be broken trend stationary (e.g. Zivot and Andrews, 1992; Ben-David and Papell, 1995; Papell and Prodan, 2011a). Broken trends blur the conceptual distinction, as a unit root process can be thought of as process with a trend that changes every single year.[7][8] While we do not attempt to characterize all type of breaks an economy can experience or formally test for unit roots, our approach is very flexible and allows for multiple growth regimes occurring before, during and after an unknown number of slumps. In other words, while we assume that there is some structure in the growth process, we do not suggest that this structure is necessarily generated by neoclassical steady-state growth or any other particular growth model for that matter.

Defining the duration of declines

In order to define the duration of the decline phase of a slump we still need to empirically identify the location of the trough. Dating the trough is relatively straightforward and only depends on whether the slump is finished or still continuing. To this end, we first define the end of a slump to have occurred with certainty at the first year a where $y_a \geq y_{\widehat{tb}_1}$. In other words, a slump is over if the level of GDP per capita preceding the slump has been reached again. The slump is still continuing if output remains below the pre-slump level.[9] Knowing whether or not at least one such year a exists is enough to date the trough and to determine if the episode should be censored. Given this endpoint, the trough is simply the year with the lowest level of GDP per capita in the slump episode. An episode is censored if GDP per capita in the last year for which we have data does not yet exceed the pre-slump level. In such a case, even if GDP per capita seems to be recovering we do not know whether the slump has ended or may continue. More formally, given the set of possible end years $A = \{a \mid a \in (\widehat{tb}_1, T] \text{ and } y_a \geq y_{\widehat{tb}_1}\}$, we estimate the

[7] Or any other interval at which the series is observed.

[8] For the same reason, it is easy to weaken the evidence in favor of a unit root and strengthen the evidence in favor of a broken trend stationary process as long enough breakpoints are permitted. See, for example, Papell and Prodan (2011a) who test for unit root against the alternative of broken trend stationarity and reject the unit root hypothesis in the majority of cases. However, most of these tests are asymmetric as they only allow for structural breaks under the alternative.

[9] This also implies that we exclude episodes estimated by the sequential algorithm if these begin before the previous slump is certain to have ended.

trough to occur at time:

$$t_{min} = \begin{cases} \operatorname{argmin}_{j\in(\widehat{tb}_1,a_0]} y_j, & \exists j \in A \\ \operatorname{argmin}_{j\in(\widehat{tb}_1,T]} y_j, & \nexists j \in A \end{cases} \tag{2}$$

where $a_0 = \min A$ corresponds to the (certain) end of the slump. If the set A is empty, then every unfinished slump still has a local minimum, but the episode is censored.

Given equation (2), the duration of the decline phase lasting from the beginning of the slump to the observed trough is simply:

$$\tilde{t}_D = \hat{t}_{min} - \widehat{tb}_1 \tag{3}$$

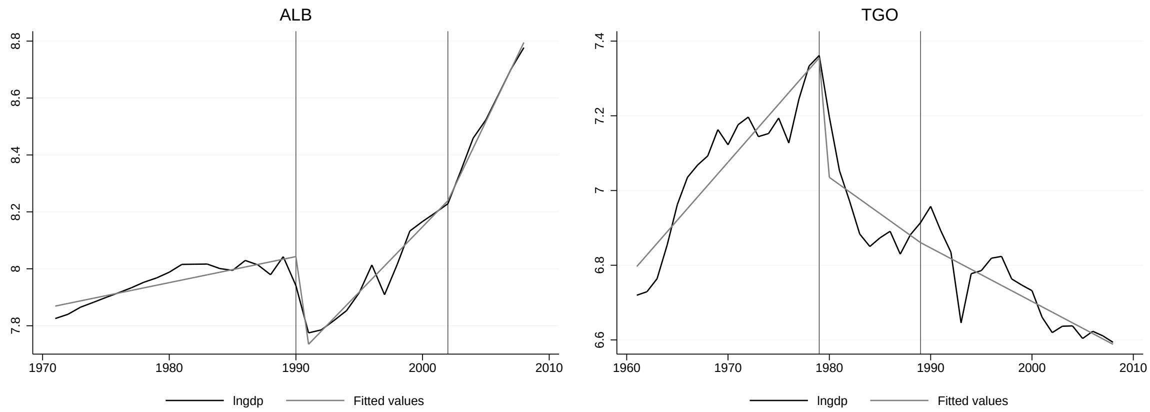
These definitions imply that in very few cases we date the trough to be after the estimated second break, which is purely a consequence of allowing for unfinished episodes and, more generally, the mechanics of the sequential two-break model. If the slump is still ongoing, the second break may have been placed at an arbitrary point maximizing the Wald test but not corresponding to the start of a new growth regime. The true trough may lie in the future.[10] Furthermore, if the slump is not finished, the actual trough could even occur beyond the end of the sample in 2008. Treating these spells as censored implies that in the later analysis we only incorporate the information that certain exit from the slump has not yet occurred at duration $\tilde{t}_D$.

Figure 1 illustrates the difference between a finished and an unfinished/ censored decline phase. Panel (a) shows a finished slump in Albania (ALB) occurring at the time of the post-communist transition. The estimated first break occurs in 1990, the trough is located in 1991 and the second break occurs in 2002. This episode neatly fits the design of the two-break model. Both the intercept drop and the two trend changes are clearly visible. After the initial contraction, GDP per capita fully recovers to pre-slump levels, thus marking the definitive end of the slump. While the duration of the decline phase is only one year, output contracted considerably. GDP per capita in 1991 was 15.32%

[10]For example, a double-dip decline of which we only observe the first part would fit this scenario.

Figure 1 – Two types of slumps

(a) finished **(b)** unfinished & censored



Note(s): Models refitted using the estimated trend breaks $\widehat{tb}_1$ and $\widehat{tb}_2$ but without the corresponding AR(p) terms to emphasize the trend breaks.

lower than in 1990. On the other hand, panel (b) shows an unfinished and censored slump in Togo (TGO). Togo first grew rapidly for about a decade following independence from France in 1960 but then experienced a dramatic collapse under the 38-year reign of General Gnassingbé Eyadéma. The first break clearly occurs in 1979, but the two-break method places the second break at an arbitrary point to accommodate the lasting decline and GDP per capita never recovers to pre-slump levels within the sample period. Hence, the decline is ongoing and the empirical trough coincides with the censoring cutoff in 2008. This is also one of steepest and longest declines in the sample. Togo's GDP per capita in 2008 was 53.6% less than in 1979.

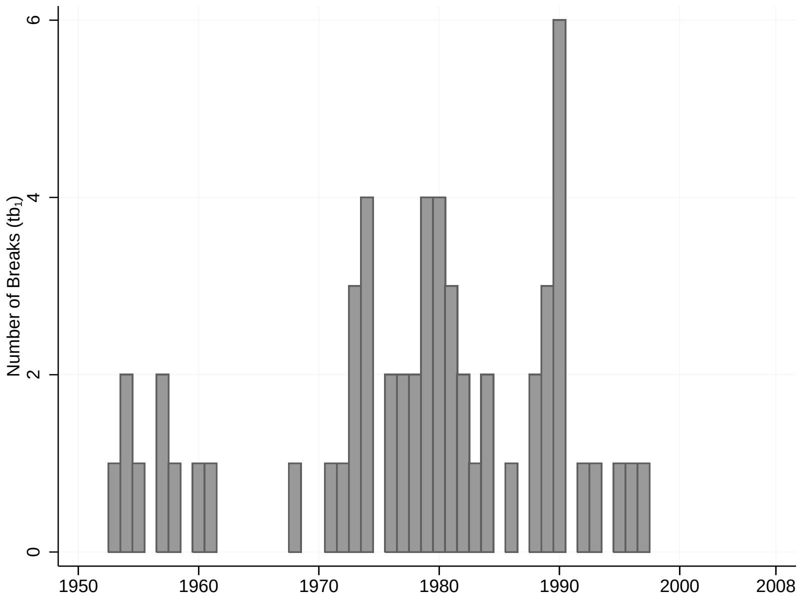
3 Descriptive statistics and data

Applying the sequential algorithm to the entire Penn World Table (v7.0), we find a total of 58 slumps between 1950 and 2008.[11] In this sample, the mean duration from the year

[11] We make two adjustments to the data. First, we only run the algorithm on countries with a population of at least one million to exclude small countries and island economies. Second, we discard episodes that are solely driven by positive breaks in the two slope coefficient(s) but have not already been excluded by the negative intercept drop requirement due to the presence of the AR(p) parameters. A simple rule is applied to these cases. We define a valid episode as an interval of two break dates $\widehat{tb}_1, \widehat{tb}_2 \in [\tau T, (1-\tau)T]$ satisfying the following condition $\exists j \in (\widehat{tb}_1, \widehat{tb}_2]$ such that $\min y_j < y_{\widehat{tb}_1}$, where τ is the trimming fraction and T is the length of the estimation sample. This rule only implies that there must be an actual decline within the range of the two estimated breaks, otherwise it is not a slump.

of the first break until the trough is about 7.7 years[12] and the median duration is 3 years. 10 out of the 58 slumps are censored and thus unfinished, implying that for these spells the location of the trough is not yet definitive. Table 14 in Appendix B lists all episodes and some of their individual characteristics.

Figure 2 – Distribution of Downbreaks



We observe several well-known growth collapses and recessions. In the case of the Finnish banking crisis of the 1990s, we estimate that the last year of the previous growth regime is 1989, the first year of the slump is 1990 and the trough occurs in 1993.[13] Chile's tumultuous economic history shows up in several big slumps. Most notably, Pinochet's coup, the subsequent reform programs, and chronic runaway inflation manifested themselves once in a sudden recession in 1975 and again in a deep but short slump in 1982 and 1983. We also identify several post-communist transitions, but for most former communist countries there is no or too few pre-1990 GDP data available. Some results are more surprising. For example, Poland's economic downturn already occurs during the 1980s and is not fully recovered by the time the post-Soviet transition takes place.

[12]This refers to the restricted mean, which treats the last observed value as the true duration. The actual mean is likely to be larger.

[13]This corresponds well with other estimates (Jonung and Hagberg, 2005). The Finnish crisis is typically dated to occur between 1991-1993, but GDP actually started to decline in 1990 already.

Given this list of episodes, we are first interested in whether any particular period or decade saw more vulnerability than others, or, more generally, how these slumps are distributed across time. Figure 2 examines the annual distribution of the the first break date ($\widehat{tb}_1$) over the entire sample range, which we also refer to as the downbreak. Most slumps begin between the 1970s and the early 1990s. We can clearly observe three periods of elevated risk coinciding with international crises. Seven downbreaks occur between 1973–1974 during the oil crisis, thirteen declines begin between 1979 and 1981 during the debt crisis of the early 1980s, and nine slumps follow the post-communist transitions of 1989–1990. Due to trimming and a deliberate sample cut-off in 2008 to avoid the Great Recession of the late 2000s, we find no beginnings of slumps in the period of the early 2000s and tranquil mid-2000s. The early 2000s recession and stock market crash did not lead to big slumps. Similarly, there are only few slumps in the 1960s but several more in the 1950s, with three slumps beginning around the time of the first post-WWII global recession of 1957. Generally, the period between the 1970s and early 1980s is marked by heightened volatility during which several star performers of the previous years become mired in deep recession – an instability of growth performances across decades that is well-documented in number of studies (Easterly et al., 1993; Rodrik, 1999; Pritchett, 2000; Jones and Olken, 2008).

Table 1 provides two additional perspectives on the data by summarizing the distributions of depth, duration, and number of spells across income groups and geographical regions. For this purpose, we define the depth of a decline as the percent decrease of GDP per capita at the trough relative to its pre-slump level. We detect considerably deeper slumps in low-income and middle-income countries than in high-income (OECD) countries. The spread of depth and duration is very large. High-income (OECD) countries experience relatively short declines with a comparatively soft landing. The median duration is only one year with a mean depth of about -7.1%. In the middle, there is little substantial variation between non-OECD high-income countries and upper/lower-middle-income countries. In all of these three groups, the mean depth is in the range of -20.8% to -27.4% and the median (mean) duration varies between about

5.4 to 6 (2 to 3) years. Low-income countries experience the most dramatic declines. Both mean and median duration are about 16 years, with an associated average depth of -37.1%. Interestingly, the number of spells itself is distributed relatively evenly across the different income groups, suggesting that developed countries, too, experience their fair share of vulnerability. But the incidence of censored spells increases linearly towards the lowest income category.

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

	Mean Depth	Mean Duration	Median Duration	Number of Spells	Censored Spells	Number of Countries
Income Level (2011)						
High-income (OECD)	-7.11%	2.00	1	12	–	29
High-income (Other)	-20.84%	5.38	2	8	1	12
Upper-middle-income	-21.20%	5.39	2	16	2	30
Lower-middle-income	-27.40%[a]	6.00[b]	3	11	3	34
Low-income	-34.17%[a]	15.75[b]	16	11	4	33
Geographical Region						
East Asia & Pacific	-13.63%	2.30	2	10	–	17
Europe & Central Asia	-13.52%	2.36	1	11	–	32
Latin America & Caribbean	-17.34%	5.27	3	15	1	23
Middle East & North Africa	-33.24%[a]	8.66[b]	9	7	3	17
North America	-2.50%	1.00[c]	1[c]	1	–	2
South Asia	-5.32%	1.00[c]	1[c]	1	–	6
Sub-Saharan Africa	-37.14%[a]	17.73[b]	16	13	6	41
Total	-21.87%[a]	7.69[b]	3	58	10	138[d]

Note(s): Depth is defined as the percent decrease in GDP per capita at the trough relative to GDP per capita before the slump (percent, not log approximation). [a] Restricted mean, last observed value is used to estimate depth. Mean depth is underestimated. [b] Restricted mean, last observed value is used as exit time. Mean duration is underestimated. [c] Only one spell in this country-group, actual values used instead of estimates. [d] After dropping countries with less 1 million inhabitants and fewer than 20 observations of GDP per capita.

The geographical distribution reveals three interesting patterns. First, Sub-Saharan Africa and the Middle East & North Africa are the two regions experiencing both the deepest and longest declines. The depth in these regions is particular striking in comparison to the other regions. The mean depth of declines is -33.2% in Sub-Saharan Africa and 37.1% in the Middle East & North Africa, which is up to 20% deeper than declines in Latin America. The duration is longest in Sub-Saharan Africa, with the median spell lasting 16 years and the mean spell lasting about 17.7 years. Declines are shorter in the Middle East & North Africa, where the mean and median are about 8.6 and 9 years respectively. Both regions also have the most censored/unfinished spells due to their long duration. Second, countries in Latin America & the Caribbean experienced slumps most frequently, but the average decline was only moderately deep and lasted for about 5.3 years. Third, when comparing Europe & Central Asia to the East Asia & Pacific region we find very similar values for mean depth, mean/median durations and the incidence of slumps. This could be due to the European group including several spells of post-communist countries[14], but even without these countries the mean depth of declines in Western Europe only is still comparatively large (-10.9%).

Taken together, Table 1 clearly illustrates that there is a relatively strong association between both the mean duration and mean depth of the decline phase with different income levels. This is particularly encouraging, since we subsequently model these observed differences in duration between high and low income economies with "more fundamental" factors such as institutions and ethnic-heterogeneity. Furthermore, the table provides a preliminary indication that there is substantial regional variation which will have to be taken into account in the ensuing analysis.

For brevity and since most of the additional data sources are well-known, we do not separately discuss the construction and summary statistics of the covariates used in the following sections. We include five major categories of variables: 1) a variety of measures for different aspects of *institutions, politics and social conflict*, 2) macroeconomic indicators of *prices and exchange rates*, 3) indicators for volumes and structure of *trade and exports*, 4) a set of variables for *financial globalization, external finances and financial development*, and 5) several *other growth determinants* (such as life expectancy or years of schooling). Table 15 in Appendix C provides an exhaustive list of all variables names, data sources and a basic set of summary statistics for the data used throughout the paper. Not all data is necessarily satisfactory but, in some cases, simply the best available. For example, we rely on the Polity IV database as our primary proxy for institutional development as a whole because of a lack of other time-series data capturing the characteristics of economic institutions. Further, we draw on two different data sources for income inequality as cross-country inequality data is notoriously flawed. We describe these and other noteworthy features or limitations of the

[14] In our sample, this refers to Albania, Bulgaria, Hungary, and Poland. Other transition economies are either not in the Penn World Tables or our models did not estimate any slumps.

data when introducing each variable and discussing the results.

4 The anatomy of slumps

Before turning to the determinants of the length of the decline phase, it is particularly interesting to investigate how a select set of key explanatory variables behaves around the time the slump occurs. The main question for this section is whether some covariates have systematically different means before the break and after, or if their means adjust strongly throughout the slump. Revealing the factors correlated with downbreaks serves two purposes. First, it highlights the characteristics of the slumps in our sample and, second, it provides a preliminary indication of which variables may play a larger role in the subsequent duration analysis. While this holds for all of the variables we examine, we aim in particular to evaluate if there is any evidence of systematic change in the institutional indicators.

To answer these questions, we employ an event methodology suggested recently by Gourinchas and Obstfeld (2012) in a study of currency and banking crises. The basic idea of the method is using dummy variables indicating a predefined distance to the start of the slump as a means of computing changes in the relative mean of each time-varying covariate. The coefficients of these "distance dummies" then measure if and how much these covariates change around the time the slump hits and their standard errors quantify the associated uncertainty. Versions of this approach have been around for some time and were popularized by the "early warning signals" and "twin crises" literatures (e.g. Eichengreen, Rose and Wyplosz, 1995; Kaminsky and Reinhart, 1999).

For each time varying covariate ($x_{it} \in \mathbf{x}_{it}$), we run the following regression:

$$x_{it} = \sum_{s=-5}^{5} \delta_{t,\widehat{tb}_1+s} \beta_s + \mu_i + \epsilon_{it} \tag{4}$$

where $\delta_{t,\widehat{tb}_1+s}$ is the Kronecker delta which is equal to one if $t = \widehat{tb}_1 + s$ and zero otherwise, β_s are coefficients, μ_i is an unobserved country effect and ϵ_{it} is an idiosyncratic error term. We set $s \in [-5, 5]$, so that the result is an 11-year window around the break date $\widehat{tb}_1$.

The results are best reported visually by plotting the coefficients β_s (including 95% confidence bands) as they represent *the conditional expectation of the covariate x_{it} at time s relative to "normal" or tranquil times.*[15] In other words, the points on the graphs are the country-demeaned expectations of each indicator over 11 years around the downbreak, where $\widehat{tb}_1$ corresponds to the estimated break date and $\widehat{tb}_1 + 1$ is the first year of the slump. We compute standard errors that are robust to arbitrary heteroskedasticity and correlation among *both country and time clusters* using the formulas suggested by Thompson (2011) and Cameron, Gelbach and Miller (2011).[16] All estimates thus account for common year shocks in each of these variables as well as within-country correlation, which would otherwise bias the standard errors.[17]

Figure 3 – Growth and Output

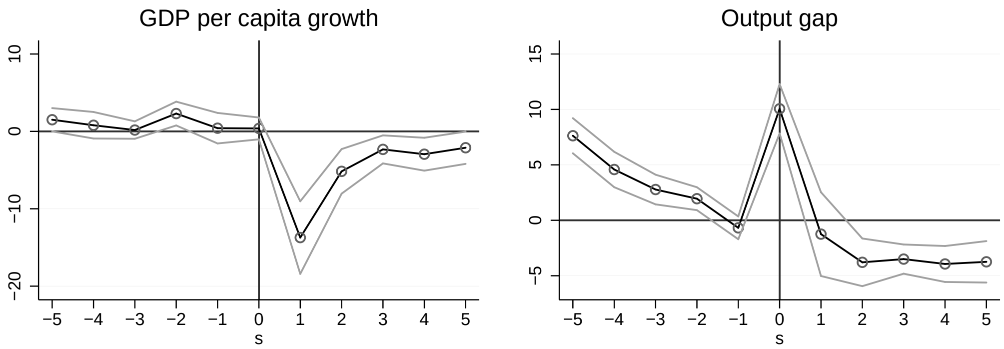


Figure 3 shows how two key economic aggregates behave during the run up to a slump and in the five years after the break date. Taken together, the graphs summarize the growth experience. Before the beginning of a slump, growth is only marginally above that of normal times, suggesting that – on average – the countries in our sample are not experiencing a growth acceleration just before the downbreak occurs. As the slump hits,

[15]This holds only for non-trending variables. Whenever we suspect the presence of time effects, we verify the results with a two-way fixed effects model.

[16]The authors show that two-way clustering can be easily estimated by calculating the variance-covariance (VCE) matrix as follows: $\hat{V}[\hat{\boldsymbol{\beta}}] = \hat{V}^i[\hat{\boldsymbol{\beta}}] + \hat{V}^t[\hat{\boldsymbol{\beta}}] - \hat{V}^{i\cap t}[\hat{\boldsymbol{\beta}}]$, where i denotes country-clustered variances, t denotes time-clustered variances and, in the case of a panel, $i \cap t$ is a White heteroskedasticity robust variance matrix. As noted by Cameron et al. (2011) the resulting $\hat{V}[\hat{\boldsymbol{\beta}}]$ matrix is not always positive semi-definite although its components may be, which occurs often when using fixed effects and clustering over the same unit. We first within transform the data to reduce the size of the VCE matrix and then reconstruct it via a spectral decomposition with all negative eigenvalues set to zero. The corresponding program is available from the authors upon request.

[17]See Petersen (2009) for an overview of when this bias is relevant and how other estimators fail to take it into account.

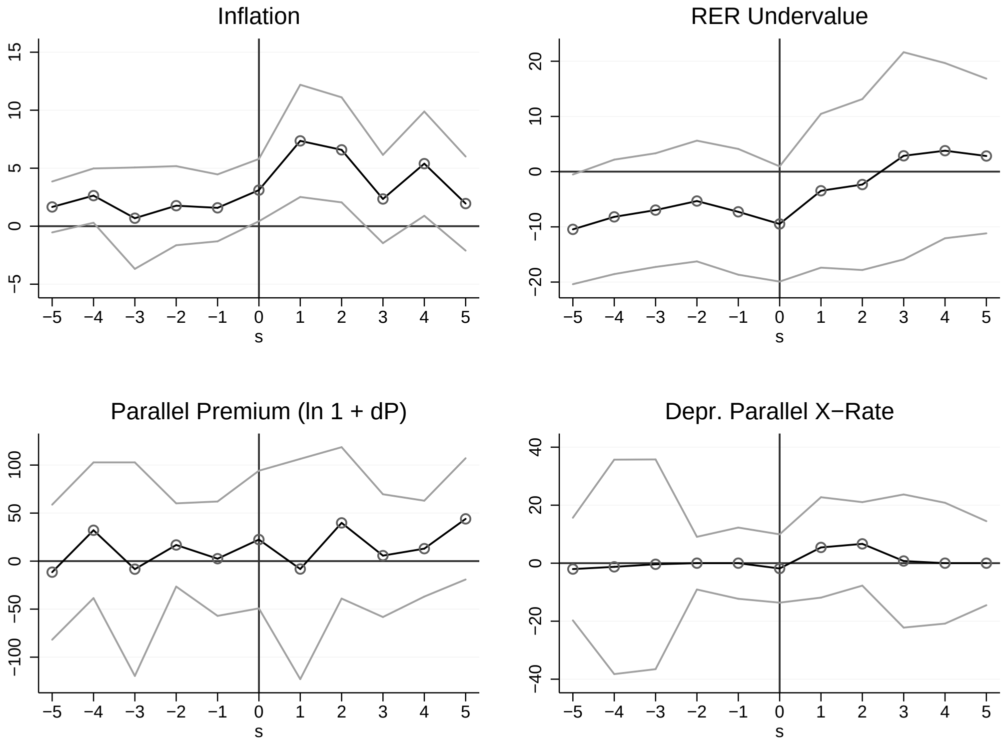
growth collapses substantially. In the first year of the slump growth is 13.7% less relative to tranquil times and remains depressed in the five years after, varying between -2 and -3% from year two to five. Since the window is symmetric around the break date, these results do not depend on the duration of the decline or the duration of the entire slump.

In the run up to the slump actual output is a few percent over potential output in five years before the downbreak (ranging from 7.6% to -0.7%), moderates to slightly below normal levels just prior to the break date and spikes again dramatically to about 10% over capacity just in the last year of positive growth. Once the slump occurs, output remains below potential in the five years after the break and is still -3.7% below potential at year five. The actual output gap is likely to be understated as the computation method used here adjusts potential output downwards relatively quickly in advance of large slumps.[18] A different estimate of the output gap which would be more in line with our definition of a slump could be computed using the implied growth rates from a linear time-trend before the downbreak. Similar to GDP per capita growth, this alternate gap would be strongly negative on the first year of the slump and then recover depending on the depth of the slumps and how long they last on average.

Turning to consumer prices and exchange rates, Figure 4 shows that inflation is slightly elevated in the five years before the downbreak but these estimates have wide confidence intervals. However, inflation strongly and significantly increases during the slump, peaking at 6-7% above the median rate in the first two years of decline. While this pattern is expected, it can have many causes, such as macroeconomic mismanagement, intentional external devaluation, or higher inflation being used as a means of avoiding other and potentially more coordinated policy responses. The real exchange rate, as measured by an undervaluation index proposed by Rodrik (2008), is overvalued relative to tranquil times in the five years before the slump and when the slump hits, but then depreciates slowly towards normal levels just after the downbreak. A large strand of the empirical growth literate argues that overvaluation hurts growth prospects and may signal

[18]The output gap is computed using a Hodrick-Prescott filter with smoothing parameter 100. The choice of smoothing parameter follows Gourinchas and Obstfeld (2012) and is motivated by filtering out cycles with a higher periodicity than the U.S. business cycle. Changing it to 6.25 (the standard value for annual time-series) only moderately dampens the estimated overrun from $s = -5$ to $s = -1$.

Figure 4 – Prices and Exchange Rates

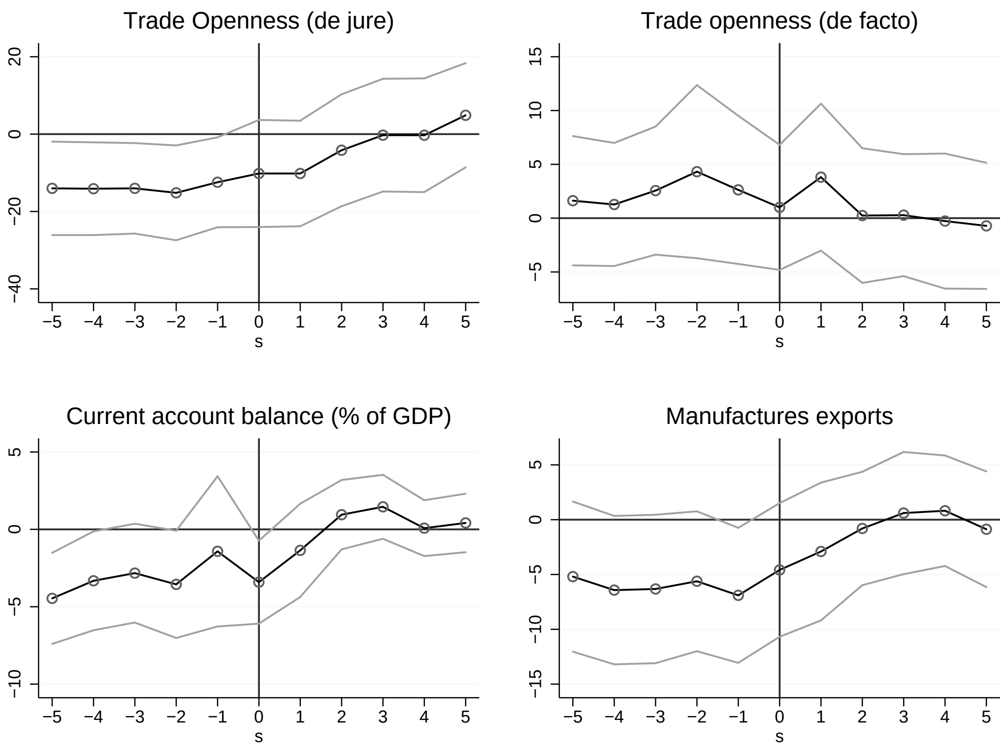


the advent of several types of crises.[19] Figure 4 suggests that the benefits of exchange rate undervaluation or detriments of overvaluation are visible in the data to some extent, but the uncertainty associated with these estimates is (too) high. Further, depreciations or appreciations in the parallel market rate and black market premia show no significant trends, but draw on a much smaller dataset from Reinhart and Rogoff (2004). For all but the real exchange rate undervaluation index, the graphs in Figure 4 are derived from median fixed-effects regressions with bootstrapped double-clustered standard errors[20] to reduce the influence of large outliers caused by episodes of hyperinflation.

[19] Although Rodrik (2008) provides evidence of the benefits of undervaluation, he views the overvaluation and growth literature sceptically. However, in the "early warning signals" literature overvaluation of the real exchange rate systematically emerges as a robust predictor of financial, currency and banking crises (Eichengreen et al., 1995; Bussière and Fratzscher, 2006; Gourinchas and Obstfeld, 2012; Frankel and Saravelos, 2012).

[20] For the quantile regressions, we apply the results from Cameron et al. (2011) and Thompson (2011) to bootstrapping. We estimate the following VCE matrix: $\hat{V}[\hat{\boldsymbol{\beta}}]^* = \hat{V}^i[\hat{\boldsymbol{\beta}}]^* + \hat{V}^t[\hat{\boldsymbol{\beta}}]^* - \hat{V}^{i\cap t}[\hat{\boldsymbol{\beta}}]^*$, where i denotes block sampling from countries, t denotes block sampling from years, $i \cap t$ denotes sampling from country-years, and the superscript '*' refers to bootstrap quantities. The method simply replaces the variance-covariance components with their bootstrap equivalents, which is an asymptotically valid approach (Cameron et al., 2011, see in particular, p. 243). Spectral decomposition is used to fix non-positive semi-definiteness by replacing negative eigenvalues with zero.

Figure 5 – Trade & Exports I



What about trade and export performance? In the upper panel of Figure 5, we use two measures of trade openness to capture whether the well-accepted principle that trade openness is good for growth also holds in the reverse sense that less openness coincides with the occurrence of slumps.[21] The the upward trend in the β_s-coefficients for the *de jure* binary measure of trade openness by Sachs and Warner (1995)[22] suggests that most countries in our sample restricted trade more heavily in the run up to a slump, relative to tranquil times, and then slowly eased these restrictions. This effect is very significant and large, as the estimates indicate that a country is 12.4-14.1% less likely to be open at any given year in the five years before the downbreak. Interestingly, a comparable effect is not evident when examining *de facto* trade openness.[23] De facto trade flows are somewhat elevated before the break occurs, but revert to normal levels thereafter and the differences in means are measured with substantial imprecision. Two other variables

[21] There is a lot of evidence is in favor of this proposition (e.g. see Winters, 2004), however Rodriguez and Rodrik (1999) criticize the usefulness of the measures used here and in the earlier literature.

[22] We use the updated version of their data as presented in Wacziarg and Welch (2008).

[23] Measured as $(X + M)/GDP$ adjusted for structural characteristics of trade, see the footnotes to the table of summary statistics and data sources in Appendix C for details.

behave similarly. The current account balance is somewhat lower relative to normal times before the slump hits but then naturally improves as the relative price of imports rises and export prices decline. This trend is mirrored by the share of manufacturing exports as percent of total exports, which is depressed before the slump but increases successively in the five years during/ after. Most of these trends are thus very close to what we would expect slumps to look like.

Figure 6 – Trade & Exports II

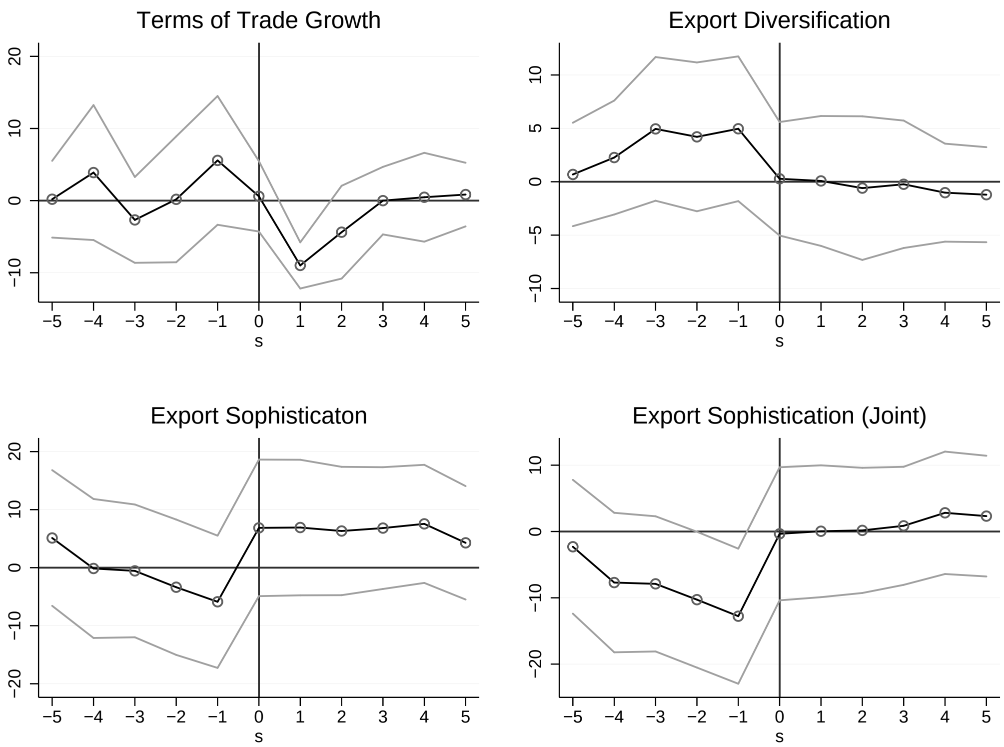


Figure 6 shows how other potentially key measures of trade performance evolve, focusing more on structural characteristics of trade and traded products. A well-established empirical result is that terms of trade shocks spur output volatility in general and may even cause growth collapses in particular (Rodrik, 1999). We measure terms of trade shocks as the annual growth rate of the net barter terms of trade. For the slumps in our sample, terms of trade shocks do not on average precede the downbreak, but the terms of trade worsen markedly relative to normal times in the first two years of the downturn (-9% and -4.4%). This effect is most likely due to a depreciating currency. Next, we examine the structure of a country's export portfolio. Narrow export baskets could leave those countries more vulnerable to demand/supply shifts in just a few industries, while those with more diversified export baskets may be more insulated against such shocks. On the contrary, we find that the conditional expectation of export diversification, measured as one minus the Herfindahl concentration index, is higher before a slump begins and declines to normal level as it progresses but these differences are not significantly different from zero.[24] Further, Hausmann, Hwang and Rodrik (2007) have suggested that higher export sophistication (higher quality and productivity of the export basket) benefits growth directly. The lower part of Figure 6 examines this relationship. The first graph shows the conditional expectation of the original Hausmann et al. (2007) measure, which suggests that export baskets are of less quality in the two years before the break but improve relative to tranquil times from the eve of the slump onwards. We also construct a longer series which gives a more intuitive result, namely export sophistication is significantly lower in the run up to the slump but then adjusts to normal levels from the eve of the slump onwards.[25] For the latter series, this effect is significant at the 5%-level in the two years before the break date.

Figures 7 and 8 show trends in various indicators for financial macroeconomic balance sheet, financial openness, and financial development. A widespread conception is that financial development and financial globalization benefit growth by reducing (consumption) volatility through lowering interest rates, broadening access to credit, and better allocating resources across the domestic economy (and global economy[26]). While this notion draws on some support from several empirical studies (e.g. King and Levine, 1993; Beck, Levine and Loayza, 2000), the question of causality is still unresolved and often found to run both ways or work through indirect channels, such as technology

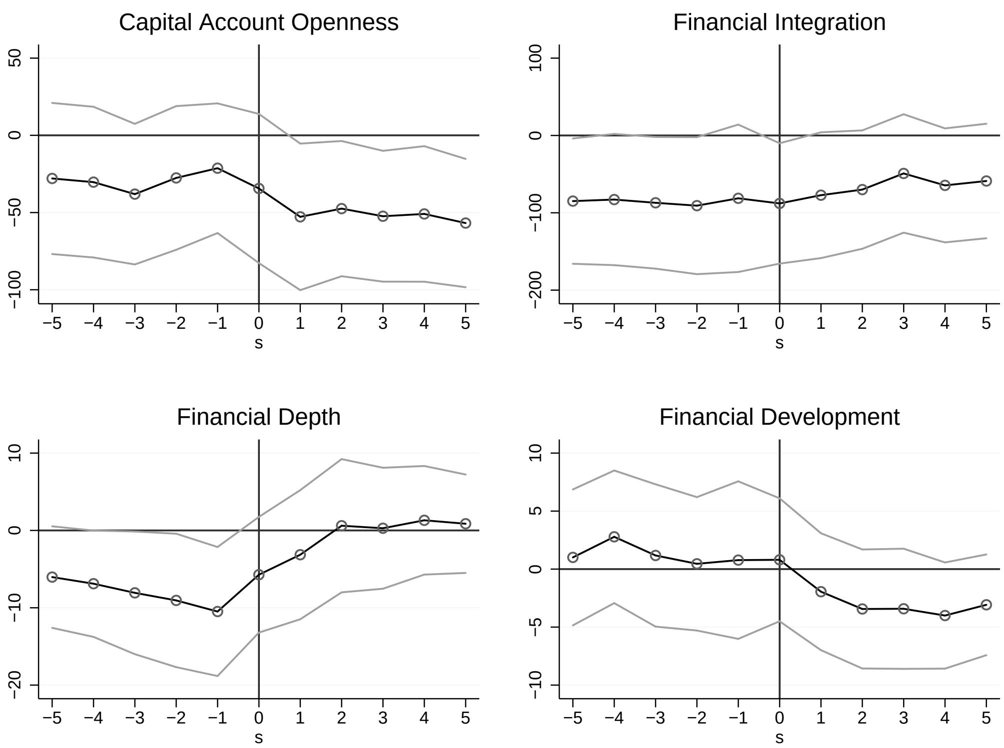
[24] To investigate whether this is caused by a time trend in export diversification, we re-estimate the underlying model with two-way fixed effects. This pushes the conditional means consistently below zero over the entire event window, while the standard errors remain large.

[25] The original measures is extended with data from the World Bank using the same method (http://go.worldbank.org/N69D6V9FE0). The difference in means was adjusted via simple OLS regression. Interestingly, when estimating a two-way fixed effects model, the results for both series suggest that sophistication is higher than at normal times over the entire event window.

[26] This is a standard implication of neoclassical growth models, but in reality there is a lack of large capital flows from rich to poor countries, which is commonly known as Lucas' paradox (Lucas, 1990).

spillovers or institutions. Equivalently to trade openness, we contrast the findings from *de jure* financial openness using an indicator of capital account restrictions (Chinn and Ito, 2006) and a measure of *de facto* financial integration, as a country's capital account may be open but real flows are few and vice versa (for a discussion, see Kose, Prasad, Rogoff and Wei, 2009).

Figure 7 – Finances I



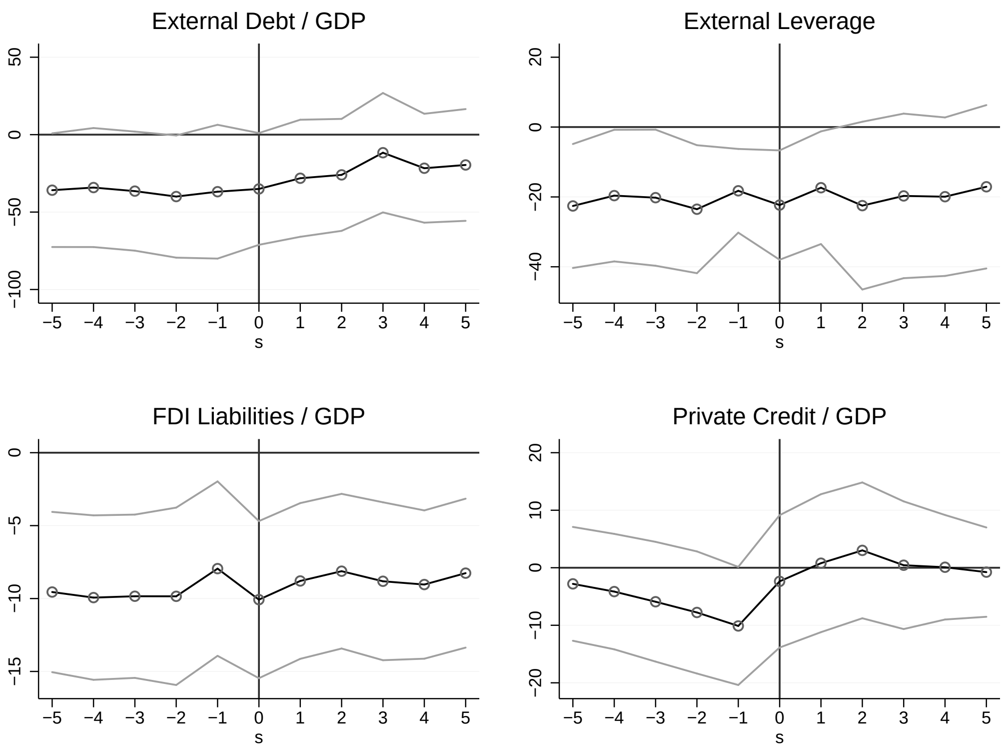
Three out of four indicators in Figure 7 exhibit a similar pattern before the slump occurs but diverge thereafter. There are more restrictions on the capital account relative to tranquil times before the slump occurs and restrictions increase further after the downbreak (between 47-56% less open throughout the five years after/ during a slump).[27] Similarly, financial integration measured as the sum of assets and liabilities over GDP is depressed before and after.[28] Liquid liabilities over GDP – a key indicator of financial depth – are significantly lower, by about -6% to -10%, in the years before the crisis but then adjust upwards to normal levels. The upward drift in several of the financial

[27]With time dummies, this curve shifts up to a more open capital account before the start of a slump and less openness thereafter (yet the standard errors are large and include zero for all estimated points).

[28]This curve also shifts towards zero if a two-way model is specified.

indicators may be owed to the denominator (GDP) shrinking faster than, in this case, the liabilities present in the financial system. We find only weak evidence that a lack of financial integration and financial depth help to explain why some countries are more vulnerable to slumps. When examining the role of more specific financial institutions using a common indicator of financial development (Deposit Money over Central Bank Assets), financial development is elevated before the break date and then declining during the slump. This may be in part due to an expansion of the Central Bank's balance sheet, possibly coinciding with a contraction of deposit money. However, almost none of these differences are significant at conventional levels.

Figure 8 – Finances II



In the case of external balances the results are very clear (Figure 8). Most of these slumps do not appear to be debt driven. External debt is very low relative to normal times before the slump occurs, then increases by about 10% but still remains lower than otherwise but this variation is measured with great uncertainty.[29] Further, Gourinchas

[29]This may be due to the time trends involved in building up external debt. Using a two-way fixed effects model shifts the curve up around zero at all event times in the 11-year window.

and Obstfeld (2012) devise a leverage ratio for countries in an empirical analogy to how leverage of firms is defined – a broader concept than just external debt. Similarly to debt levels, this measure indicates that the countries in our sample are significantly less reliant on external financing in the 11 year window than at normal times.[30] Not only debt is low, the stock of FDI liabilities is also about 8-10% lower throughout the 11 year period relative to normal times, suggesting that periods before and during slumps are associated with comparatively little FDI inflows (which are usually considered particularly desirable and stable investment flows). Taken together, this implies that most of the countries in our sample are not well integrated into global finance in the run up to a slump, or according to some indicators throughout the entire event window. Contrary to this pattern, private credit to GDP is depressed just before the slump, much like financial depth (Figure 7). Hence, contractions in credit may indicate upcoming slumps.

Figure 9 – Institutions & Politics

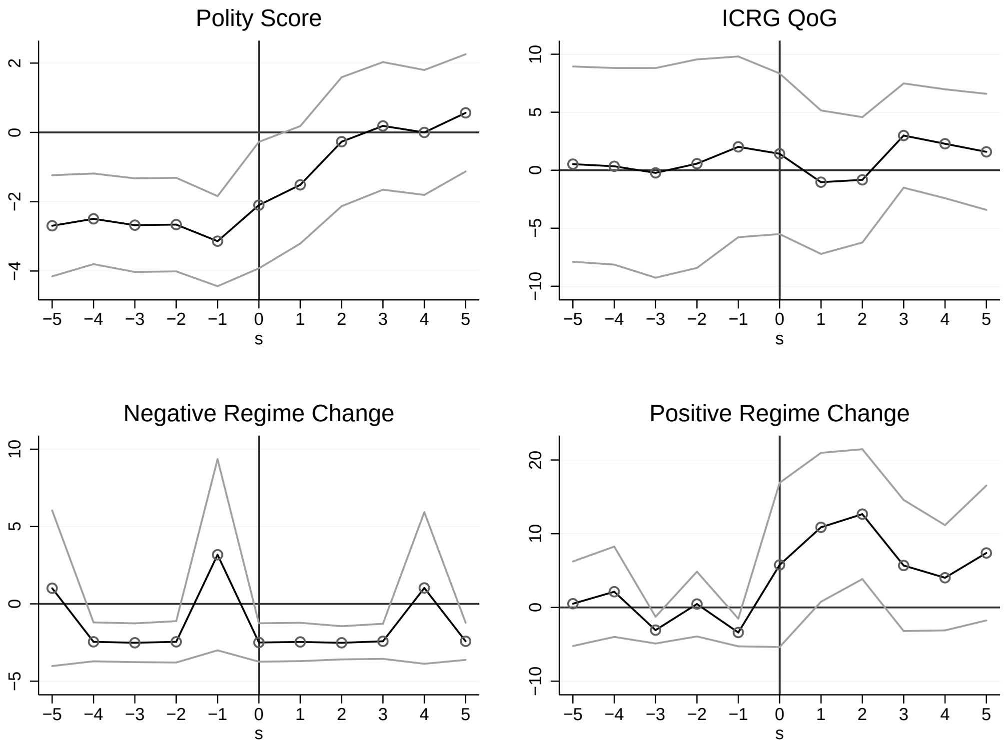


Figure 9 shows a panel of graphs describing how certain institutional and political

[30]In model with additional time-effects, the ratio also hovers around zero with no significant differences at any time s. Similar results are obtained in a two-way model for FDI liabilities / GDP. All of these additional graphs and model estimates are available from the authors upon request.

dynamics evolve around the time of the break date. In many ways, these results are the most remarkable of this section. The previous literature suggests that better institutions reduce macroeconomic volatility and lead to better polices in general (Acemoglu et al., 2003). In addition, the literature finds a positive relationship running from democratic institutions and political stability to reduced output volatility in particular (Mobarak, 2005; Klomp and de Haan, 2009). The Polity score is much lower before a slump occurs, but increases towards tranquil levels thereafter. In the five years before a slump, the conditional expectation is between 2.5 and 3.1 points lower than in tranquil times and until the break date these differences are significant at the 5%-level. Furthermore, all the subcomponents of the combined polity score, including *constraints on the executive*, exhibit very similar trends depending on their scale (not shown).[31] This strongly suggests that prior deficiencies in institutions increase vulnerability to negative growth shocks and may ultimately also affect slump duration. However, the Quality of Government indicator does not show the same pattern, yet it also offers less time-coverage and is conceptually more limited. We interpret these different trends as an indication that broader institutional characteristics in excess of 'bureaucratic quality' are picked up by the Polity IV data. The time profile of the probabilities of negative and positive regime changes, measured as a minimum three-point change in the Polity score, underline this point. There is little evidence that negative regime changes precede downbreaks or occur with heightened probability thereafter, but there is a definitive upward trend in the probability of positive regime changes from the eve of the slump onwards. In the first and second year of a slump, the probability of a positive regime change is 10-12% higher than in tranquil times and while these point estimates are individually highly significant, the *successive* probability over several years would be even higher.[32] Slumps are thus both preceded underdeveloped institutions and then present policy windows as sharply negative growth opens up space for reform. A more encompassing interpretation is that given prior institutional deficiencies, slumps bring about a form of *creative political*

[31] Additional models and graphs demonstrating this point are available from the authors upon request.

[32] Since test statistics of such successive effects are mutually dependent and it is unreasonable to assume independence over "distance time" s, they cannot be calculated in a straightforward manner.

destruction by altering power relations and increasing the pressure on governments to pursue institutional change (North et al., 2009).[33]

Figure 10 – Social & Political Conflict

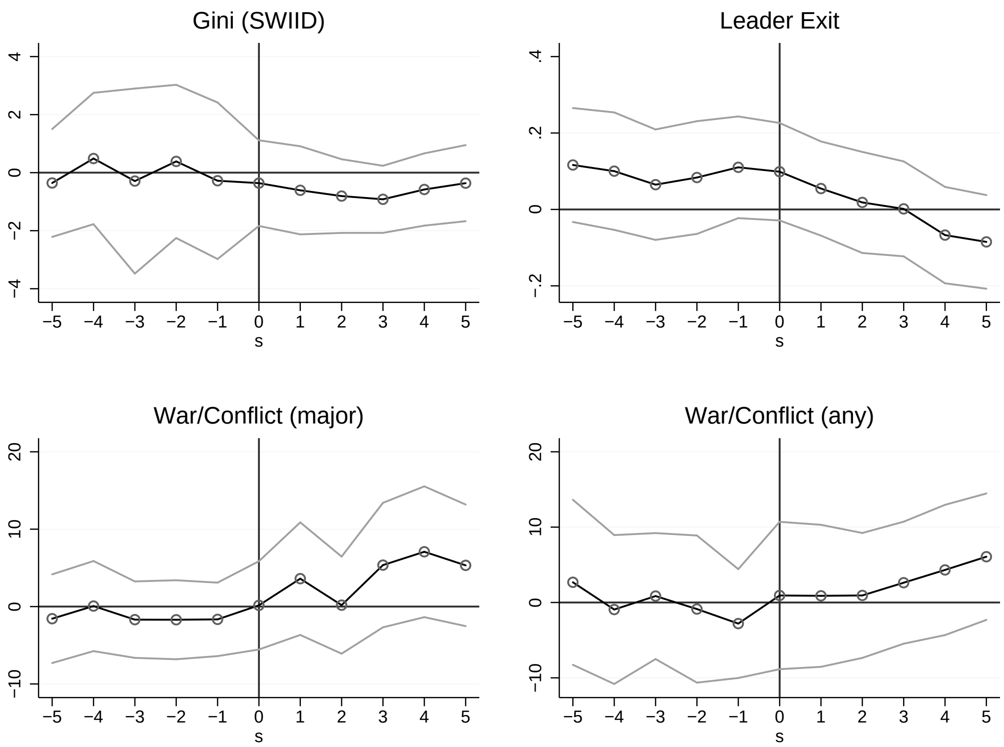


Figure 10 shows a set of measures which are sometimes interpreted as the degree of open and latent social conflict challenging the conflict management capacity of a country's institutions (e.g. Rodrik, 1999). The picture these indicators present is mixed. Inequality, as measured by the Gini, is not elevated before or after and even shows a very small negative trend during/after slumps.[34] This is not too surprising. On the one hand, income inequality is a deep-seated social phenomenon implying that we typically see few

[33]This is a common theme in the literature on the political economy of institutions, where the distribution of wealth among groups levying political power – now and in the future – is often at the center of an equilibrium solution or "political settlement". Weak institutions can both be the cause of declining overall wealth, by providing incentives to political and economy actors to seek rents, and declining wealth may then bring about political realignments as the bargaining position of actors changes (among many others, see Acemoglu et al., 2004; Acemoglu and Robinson, 2006; North et al., 2009). Greif and Laitin (2004) go even further by arguing that even seemingly stable institutions can be self-undermining in the long-run if they do not continuously broaden the set of situations in which they are supported.

[34]If we use the UTIP inequality data the trend is different. Household inequality is marginally lower relative to tranquil time in the run up to a slump and then increases back to normal level, but none of the coefficients are significantly different from zero. We prefer the data from Solt (2009), as it represent the most detailed effort to standardize and benchmark data across a variety of sources to-date.

swift changes. On the other hand, even if we expect overall inequality to change in response to crises it is not clear which part of the income distribution is most affected. Crises do not necessarily just hit the poor but may also have large negative effects on capital incomes and many other income sources, so that "churning" under the surface can make the overall impact on the Gini ambiguous. However, this finding does not preclude that differing levels of initial inequality could be associated with the duration of declines, as we will investigate in the subsequent section.

Next, we examine three additional indicators of outright conflict. Coups d'états, assassinations, but also deaths of leaders in office and other forms of abrupt government changes that could tip a (institutionally weak) country into crisis are measured by the variable irregular 'leader exit' (Goemans et al., 2009). The Figure 10 illustrates that the probability of an irregular exit is elevated relative to tranquil times prior to the first year of the slump, but the difference is not statistically significant from zero and the effect smoothly declines after the downbreak. Political turmoil is thus only weakly linked to subsequent crises. Outright wars or major conflicts[35] between state and non-state actors are another extreme form of social conflict that could in many ways destroy the economic base of a country and cause slumps to occur, as well as prolong their duration. We find little systematic evidence in favor of the former. The probability of war increases from the eve of the slump onwards, yet there is no indication that – on average – wars tip countries into slumps. The effect is even weaker when we use a lower threshold designed to capture more low intensity conflicts[36] such as ongoing civil strife and other forms of sectarian violence. There is a slight upward trend from two years after the downbreak onwards, but the coefficients are both quantitatively small and statistically insignificant.

In sum, this section outlined the characteristics of slumps and identified several factors associated with the decline phase.[37] Many of these indicators and economic aggregates evolve in the expected manner but often include a mix of endogenous policy responses.

[35] Defined as any armed conflict coded as "war" in the PRIO Armed Conflict Database; that is, any ongoing conflict with 1,000 battle related deaths in any given year.

[36] War/Conflict (any) is coded as unity if there are at least 25 battle related deaths in any given year.

[37] We did not attempt to "explain" the onset of crises in this part, for such an analysis see the "early warning signals" literature cited in the text, but also Bluhm, Crombrugghe and Szirmai (2012) on the difference between onset and continuation of stagnation spells.

For example, higher inflation, a depreciating real exchange rate and a re-balancing of the current account are both testament of the strong price pressures faced by these economies but also of the necessary adjustments that ultimately help to stabilize the fall. However, other covariates behave in interesting ways around the break date. The difference between de facto trade flows and de jure openness is striking and suggests that trade restrictions play an important role. Similarly, export structures, especially the quality of exports, is unusually low before the break and may thus hold explanatory power over slump duration. Additionally, several indicators of financial development and financial integration either switch means around the time the slump hits or remain permanently below the levels of tranquil times throughout. While this exercise could certainly be extended further with more policy variables, the most interesting and unexpected pattern finding is a switch from significantly lower quality institutions in the run up to a slump back to mean scores occurring in the first two years after the downbreak. This indicates that weaker institutions precede the beginning of a slump, while the slump itself offers a window of opportunity for institutional improvements.

5 The duration of declines

5.1 Method

To answer our main research question, we use an approach that is often interchangeably referred to as duration analysis, event history analysis or survival analysis. Duration methods were first developed in medical statistics, particularly for drug trials measuring the survival chances of patients under different treatments, but have since been generalized and applied in numerous disciplines. Contrary to linear regression or regression with binary dependent variables, duration analysis models the probability of an event occurring at a certain time *conditional* on the event not having occurred before. For example, in the case of mortality due to a terminal disease, the typical event is the time of death of a patient who is considered at-risk from the first signs of the disease onwards. Survival analysis then asks "how long do we expect patients with this terminal disease to survive"? To answer this question, we need to model the probability of dying in the first year of the disease, then the probability of dying in the second year conditional on surviving the first, and so on. A separate framework is required, because durations have several properties making them unsuitable for classical linear analysis. First, durations are always positive. Second, the distribution of durations is often non-normal, non-symmetric and multi-modal. Third, an observed duration may be right-censored, i.e., all we know is that the event has not occurred yet.

There are two major approaches to analyzing such data: semi-parametric Cox proportional hazards regression and parametric models (e.g. exponential, Weibull, log-normal). Cox regression forms so-called risk sets of the sample population at the observed event times and then computes the probability of the event occurring in each particular risk set given the observed data. Most importantly, Cox regression leaves the baseline chances of the event occurring unspecified and only assumes that this baseline hazard is scaled up and down proportionally. Under certain assumptions, this approach can be shown to be equivalent to several conditional logistic regressions using the at-risk population each time the event occurs (Box-Steffensmeier and Jones, 2004, p. 58). Parametric regressions, on the contrary, specify a full model of the duration process itself, but different parametrization originate from different vantage points.

In all parametric models, the underlying duration process is formalized either via a hazard model or an accelerated failure time (AFT) model. Hazard models usually begin with a log-linear specification of the intensity process, which stands for the instantaneous probability of an event occurring at time $\tilde{t}$ conditional on that it has not occurred before.[38] Given this hazard rate, it is possible to derive the survival function which captures the cumulative probability of the event not having occurred until the observed time. If the observation exits the initial state at the observed duration (i.e., the patient dies in the example above), then the model estimates the probability that the event occurred at that particular time given the probability of remaining in the initial state until that time (e.g. probability of dying conditional on survival). If there is no observed exit, only the

[38] We denote analysis time as $\tilde{t}$, where $\tilde{t} \equiv t - t_0$ so that we can refer to the calendar times t and t_0 when necessary. Analogously, the last observed duration is $\tilde{t}_D = t - t_0$, where $t = \hat{t}_{min}$ and $t_0 = \widehat{tb}_1$.

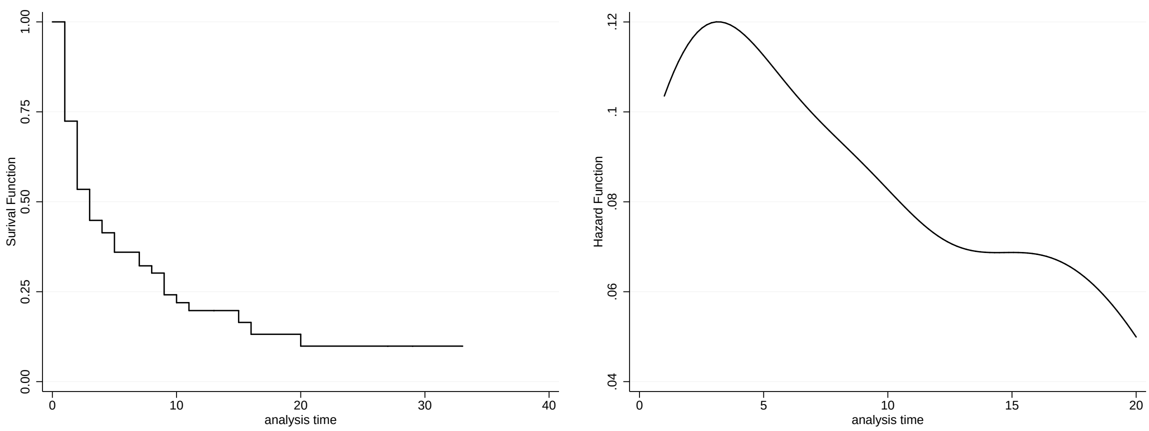
survival probability is counted and the observation is censored. Accelerated failure time models specify the same underlying relationships, but directly begin with a log-linear model of the duration to derive the survival function and thus bear more similarities to linear regression. Log-normal models do not imply proportionality of the hazard rate, but some models can be specified in both metrics. A detailed and technical description of how log-normal models are parametrized and estimated via Maximum Likelihood can be found in Section D of the Appendix.

The term "accelerated failure" derives from the interpretation of the effects estimated by AFT models. If the coefficient on the covariate of interest is positive, then the expected duration until the event is prolonged by higher values of the covariate. In our case, this is equivalent to delayed exit from the decline phase (later start of the recovery). If the coefficient associated with the variable of interest is negative, then the expected duration is shortened and time passes more quickly (time is "accelerated"). In our case, the exit from the decline phase will occur earlier due to the effect of the covariate in question. The expect decline is shorter and the recovery begins earlier. If the coefficient is zero, or statistically not different from zero, then the null hypothesis that the variable does not affect the expected duration cannot be rejected.

Figure 11 – Unconditional Survival and Hazard Functions

(a) Kaplan-Meier Survival Function

(b) Gaussian Hazard Function



Note(s): The hazard function has been smoothed using a Gaussian kernel with bandwidth parameter 3 and no boundary adjustment. The Kaplan-Meier estimate shows the implied actual exit rate at each unit of analysis time.

All parametric duration models make certain assumptions about the shape of the hazard function. In other words, they assume a particular shape of the time-profile of the conditional probability that exit from the initial state occurs at analysis time $\tilde{t}$. The simplest parametric model, the exponential model, assumes that the hazard is constant over the entire duration process. Models with a Weibull or Gompertz distribution allow for flat, monotonically increasing or monotonically decreasing hazard rates. Log-normal and log-logistic models offer a non-monotonic function that is first increasing and then decreasing. The generalized gamma distribution is very flexible and encompasses the exponential, Weibull and log-normal distributions but is more demanding to estimate.

There is no strong theoretical prior for the duration of declines to follow any one particular hazard shape. We may expect some countries to exit rather quickly and others to take longer, but we do not know ex ante if remaining in the decline phase for very long leads to a deterioration of fundamentals and thus a decreasing hazard, or if the probability of exit is actually increasing because countries are bound to enter a strong recovery eventually. As Figure 11 shows, the smoothed unconditional hazard is downward sloping after the first couple of years. However, the shape of the conditional hazard (with covariates) may be very different. We take a flexible approach by first relying on a log-normal parametrization of the hazard shape and then testing the robustness of our preferred specification under different assumptions.

For all country-spells, we specify versions of the following log-normal AFT model of the duration until exit of the decline phase:

$$\tilde{t} \equiv \ln(t - t_0) = \mathbf{x}'_{t_0}\boldsymbol{\beta} + \delta USI_t + \sigma\epsilon_t \tag{5}$$

where $\tilde{t}$ is the duration at time t (so that, $\tilde{t} = \tilde{t}_D$ at $t = \hat{t}_{min}$), $\boldsymbol{\beta}$ is a vector of coefficients for a $\mathbf{x}_{t_0}$ vector of covariates at t_0, USI_t is the real US interest rate at time t, and ϵ_t is distributed $\mathcal{N}(0,1)$, with a scale parameter σ. Further, for each country-spell define a censoring indicator that is one if the duration is censored and otherwise.

An important complication in survival data with time-varying covariates is the problem of possible feedback effects running from the duration to the observed covariate. If any such feedback exists, the estimated coefficients will be biased and the usual test statistics invalid (Lancaster, 1990; Kalbfleisch and Prentice, 2002). In some instances, joint modeling of the covariate and the duration outcomes can achieve valid inference in the presence of feedback effects, but with multiple endogenous measures this quickly becomes a daunting task involving many limiting assumptions on the data generating process. In order to avoid such problems altogether, we simply take the pre-slump value of the explanatory valuable at $t_0 = \widehat{tb}_1$. This rules out any possible effects running from the duration to, say, higher inflation, stronger institutions, or any other of the possibly endogenous covariates. Only the real US interest rate is assumed to be exogenous in all model specifications, so there is no need to fix the US interest rate at its pre-slump value.

Another technical issue is that countries can have several recurrent slumps. In practice, this is a minor concern, since only 8 of the 58 spells in our data are not the first spell for a given country. To account for the dependence of the parameter estimates across spells of the same country, we allow their variances to be correlated (clustered) across countries using the standard Huber/White/Sandwich estimator of the variance-covariance matrix, but we assume the sequence of spells does not matter. In the robustness section, we examine how well the results hold up when relaxing this assumption.

Dealing with a maximum of only 48 exits in 58 decline spells, representing 348 years of decline over the entire period of 1950 to 2008 in total, requires a careful approach to model selection. While the maximum sample size is statistically large enough to identify reasonably robust results, we match these episodes with data over the almost six decades spanned by them. Simply including many covariates with different patterns of missing data then easily leads to sample sizes that are too small by conventional standards. Even at more moderate sample sizes, care needs to be taken to guard against overfitting.[39] To arrive at a parsimonious model specification, we first fit variable-by-variable regressions including only a set of minimal controls and then build summary models selecting those variables that passed the bar of explaining a significant amount of the (close to unconditional) variation. This approach borrows from Berg et al. (2012)[40]

[39] Overfitting can occur when there are too many variables relative to observations. The model comes close to being fully determined but near irrelevant as it describes mostly noise/ minor details and not general patterns. A often used rule of thumb is to have at least five failures (exits) per included variable.

[40] Apart from studying a very different type of growth episode, our method differs in several more

who faced similar limitations when studying the duration of growth accelerations. While we do explicitly aim to test the influence of institutions on the duration of declines, many alternative theories exist and the empirical growth literature is abundant with covariates that may explain declines phases. Hence, the following section also serves to select those covariates and associated theories receiving most support from the data.

5.2 Variable-by-Variable Regressions

The null model in Table 2 shows the reference log-likelihood for the full sample including no covariates other than a constant. For models of similar sample size only, the log-likelihood (denoted $\log \mathcal{L}$) can be used as a reference to assess potential improvements by the addition of other variables. The second model only includes initial GDP per capita and a constant, while the third model accounts for US interest rate movements on the duration but omits initial GDP. The signs are as expected in these two very basic models, since in the AFT metric a positive coefficient implies that (log) time takes longer to pass, and *vice versa*. Higher initial GDP is associated with shorter declines, while increases in the US interest rate lead to a contraction of global dollar supply and thus predict longer declines. To account for these effects, all subsequent models also condition on initial GDP and the real interest rate on 3-month US treasuries as a set of minimal controls.

Table 2 – Base Models

	Coefficient	SE	p-value	Exits	Spells	Years in Spell	$\log \mathcal{L}$
Constant Only	1.346	0.180	0.00	48	58	348	-87.86
Initial ln GDP per capita	-0.124	0.144	0.39	48	58	348	-87.44
Real US Interest Rate	0.096	0.047	0.04	48	58	348	-86.55

Each model refers to one row in Tables 3 to Table 7. We only report the coefficient, standard error and p-value of the variable of interest and omit these estimates for initial GDP per capita, the real US interest rate and a constant. Table 3 shows simple duration regressions with prices and exchange rates as covariates. Inflation has a very small, marginally negative and insignificant coefficient, suggesting that initially elevated

respects. For example, Berg et al. (2012) also include changes of endogenous time-varying covariates in their models, which we exclude on grounds of not violating strict exogeneity. Also, we do not rely on a Weibull shape as our primary hazard parametrization.

inflation has very little to do with the duration of downturns.[41] This contrasts directly with many panel studies in the growth regressions literature often finding that higher inflation reduces growth, which may in part be due the conditional nature of our sample.[42] Inflation is correlated with short-run output dynamics but the duration of big slumps is not driven by initial inflationary pressures. Similarly, initial undervaluation of the exchange rate does not significantly affect the duration but has a negative coefficient, in line with studies suggesting that undervaluation benefits growth (Rodrik, 2007) or the traditional view that overvaluation is harmful. Depreciations in the parallel rate point in a direction consistent with the coefficient on the exchange rate index but also have no statistically discernible effect, while upward shifts in the premium paid in the parallel market are significantly correlated with longer declines. However, the latter effect is measured with (too) few degrees of freedom to be considered further.

Table 3 – Prices

	Coefficient	SE	p-value	Exits	Spells	Years in Spell	$\log\mathcal{L}$
Inflation ($\ln(1+\pi)$)	-0.002	0.004	0.68	38	45	234	-64.62
Real X-Rate Underval	-0.144	0.333	0.67	48	58	348	-86.13
Parallel X-Rate	0.179	0.166	0.28	28	31	122	-40.72
Parallel Premium	0.269	0.136	0.05	20	22	64	-22.96

Notes: All regressions also condition on (log) initial GDP per capita and real US interest rates.

The preliminary results on trade performance are particularly interesting. The earlier section showed how strongly trade performance and export structures react in the build up to and response to a slump. In terms of these factors driving the duration of a declines, Table 4 reveals a nuanced picture. De jure openness is strongly and significantly related with shorter declines, while de facto openness moderately but significantly increases the time until the trough is reached. In other words, open economies, as measured by the Wacziarg and Welch (2008) dummy, experience substantially shorter declines but economies that are strongly integrated in terms of actual trade flows and thus more heavily reliant on world trade experience moderately longer declines. Omitted variables may be at work, but this ex ante difference is remarkable. Further, export structure

[41] Even when conditioning only on initial inflation, we only find a coefficient of -.004298 with an associated p-value for a two-sided test of 0.508.

[42] The sample only includes large, unexpected and pronounced slumps; that is, we discard many short-run business cycles and possibly along with that the usual effects of inflation on these cycles.

matters in interesting ways. A higher initial share of manufacturing exports in total exports shortens the length of declines, export diversification has the opposite sign than expected, and both measures of sophistication of a country's export basked are strongly associated with shorter declines. Taken together, we find that manufacturing exports and exports of high value products help reduce the duration of a decline, yet it is not diversified exports *per se* but more sophisticated exports that have a protective effect. The current account balance and terms of trade[43] are not significant predictors but their signs are as expected. An initially more balanced current account and better terms of trade are associated with shorter declines.

Table 4 – Trade and Exports

	Coefficient	SE	p-value	Exits	Spells	Years in Spell	$\log \mathcal{L}$
Trade Openness (de jure)	-1.019	0.304	0.00	43	52	316	-74.89
Trade Openness (de facto)	0.017	0.005	0.00	48	58	348	-79.33
Current Account Balance	-0.006	2.672	1.00	27	34	222	-47.79
Terms of Trade Growth	-0.704	1.722	0.68	24	27	164	-36.63
Manufactures (% Exports)	-0.016	0.006	0.01	27	34	203	-45.79
Export Diversification	0.015	0.008	0.07	24	31	236	-42.29
Export Sophistication (1)	-2.131	0.574	0.00	28	34	241	-45.86
Export Sophistication (2)	-1.771	0.440	0.00	32	40	290	-55.59

Notes: All regressions also condition on (log) initial GDP per capita and real US interest rates.

Most of the financial variables do not significantly affect the duration of decline spells. The coefficients for six out of eight covariates in Table 5 are insignificant at conventional levels, as indicated by large standard errors and p-values. However, two variables stand out. First, financial depth has a substantial negative effect on the expected duration, suggesting that more developed financial systems help to buffer negative growth shocks. Second, the coefficient of private credit provided by deposit money banks and other financial institutions underscores this effect. It is also significant at the 5%-level and points in the same direction as the coefficient on financial depth.

The insignificant covariates also tell a story. Several factors commonly linked to growth are *not* systematically correlated with shorter or longer declines phases in our sample. Most notably and corroborating the picture of the earlier section, the coefficients

[43]The terms of trade could also be measured as a time-varying covariate since they are at least in part external to any one particular country. The results remain qualitatively and quantitatively similar if we include the time dimension.

of debt to GDP and external leverage have extremely large confidence intervals and are virtually indistinguishable from zero. The duration of declines is unlikely to be related with large debt levels or over-leveraged international investment positions. Similarly, higher FDI stocks relative to the size of the economy have a negative effect on the expected duration, in line with the theoretical expectation, but this effect is not statistically significant. In addition, the positive signs but large standard errors on capital account openness and financial integration (sum of assets and liabilities to GDP) suggest that financial globalization plays no role in determining slump duration. This also holds for the degree of financial development within the private sector vis-à-vis the public sector measured as the sum of deposit money assets over central bank assets.

Table 5 – Finances

	Coefficient	SE	p-value	Exits	Spells	Years in Spell	log $\mathcal{L}$
Capital Account Openness	-0.016	0.125	0.90	32	41	275	-59.63
Financial Integration	0.000	0.003	0.88	35	43	271	-61.67
Financial Depth	-0.015	0.005	0.00	26	33	195	-44.81
Financial Development	0.006	0.009	0.55	31	39	266	-57.87
Debt / GDP	0.000	0.007	0.98	35	43	271	-61.69
External Leverage	0.310	1.376	0.82	35	43	271	-61.64
FDI Liabilities / GDP	-0.004	0.018	0.83	35	43	271	-61.67
Private Credit / GDP	-0.013	0.004	0.00	28	35	198	-47.09

Notes: All regressions also condition on (log) initial GDP per capita and real US interest rates.

Table 6 investigates a whole array of variables related to institutions, politics and social conflict. Most notably, nearly all of the institutional variables have the expected sign, meaningful effect sizes and small standard errors. Expanding on the results of the previous section, all of the Polity IV variables measuring various aspects of the political institutions of a country are strongly associated with the subsequent duration of declines. The only variable which does not have such an effect is regime durability (the elapsed time since the last regime change).[44] In addition, the model with *executive constraints* has a smaller log-likelihood than either the model with the aggregate Polity IV score or models with any other of its sub-components. Coincidentally, constraints on the executive is also our preferred proxy of institutional quality for two reasons. On the one hand, it is

[44] As in Section 4, regime change is defined as a three-point change in the Polity score, but now refers to both negative and positive transitions. In models including measures for positive or negative regime changes at the break date, neither coefficients are anywhere near significant (results available on request).

widely used in the empirical literature as a measure of property rights institutions and has already been linked to macroeconomic volatility (e.g. Acemoglu et al., 2003; Acemoglu and Johnson, 2005). On the other hand, it is more conceptually rooted in the economic theory of institutions than any of the broader measures capturing wider aspects of the political regime (e.g. democracy or autocracy). The direction of the relationship between institutions and duration of declines is also confirmed by an alternative indicator. The ICRG's Quality of Government indicator has the same sign, but offers a considerably smaller number of observations for the relevant period. Taken together, these models suggest that institutions, especially property rights institutions, have a potentially very significant impact on the duration of declines that requires further attention.

Table 6 – Institutions, Politics and Conflict

	Coefficient	SE	p-value	Exits	Spells	Years in Spell	log $\mathcal{L}$
Polity IV Score	-0.064	0.018	0.00	47	57	346	-80.27
Democracy Score	-0.118	0.032	0.00	47	57	346	-80.43
Autocracy Score	0.127	0.038	0.00	47	57	346	-80.57
Executive Recruitment	-0.163	0.057	0.00	47	57	346	-81.90
Executive Constraints	-0.218	0.065	0.00	47	57	346	-79.70
Political Competition	-0.122	0.038	0.00	47	57	346	-80.89
Regime Durability	-0.002	0.005	0.72	47	57	346	-84.96
Quality of Gov (ICRG)	-0.038	0.009	0.00	14	18	98	-18.00
Fractionalization (ELF1)	0.018	0.007	0.01	48	58	348	-83.82
Fractionalization (ELF15)	0.018	0.004	0.00	48	58	348	-78.07
Gini (SWIID)	0.045	0.023	0.05	22	27	137	-34.73
Gini (UTIP)	0.028	0.034	0.40	25	29	199	-42.67
Irregular Leader Exit	0.424	0.360	0.24	47	57	346	-84.18
War / Conflict (major)	0.179	0.875	0.84	48	58	348	-86.19
War / Conflict (any)	0.460	0.556	0.41	47	57	346	-84.55

Notes: All regressions also condition on (log) initial GDP per capita and real US interest rates.

What about measures of outright or latent social and political conflict that are often investigated in conjunction with institutional characteristics? Table 6 also reveals some interesting findings in this regard. Neither the irregular ousting/ disposition of political leaders nor the occurrence of major or minor wars and conflicts have a significant effect on the length of slumps but their coefficients point in the expected direction. However, higher ethno-linguistic fractionalization and income inequality are strongly associated with longer declines. Building on the seminal contributions by Easterly and Levine (1997) and Alesina, Baqir and Easterly (1999), Desmet, Ortuno-Ortín and Wacziarg

(2012) recently developed a very detailed set of measures for ethnic cleavages. They compute these measures at different levels of the language tree and thus capture the effects of linguistic diversification into increasingly narrower language groups. We use two measures at the extremes of the spectrum, which we could not include Section 4 on the anatomy of slumps due to their time-invariant nature. Fractionalization (ELF1) is the most aggregate level, capturing only crude distinctions such as Indo-European versus non-Indo European languages, and Fractionalization (ELF15) represents the most disaggregate level, differentiating among the language groups known today. In their paper, Desmet et al. (2012) show that aggregate fractionalization does not explain much of the variation in growth rates across countries, but the disaggregate level strongly predicts growth differentials. Interestingly, we find very little difference between the ELF1 measure and the ELF15 measure in these basic models but the latter has a somewhat smaller standard error.

With respect to inequality, our preferred measure of income inequality from Solt (2009) predicts longer spells at the 5% significance level. The coefficient of the inequality measure from the University of Texas Inequality Project (UTIP) is similar in size but statistically insignificant. This suggests that higher inequality may not only shorten the length of growth accelerations, as in Berg et al. (2012), but is also associated with longer declines. However, the latter effect could be driven by omitted variables. In addition, the poor quality and comparability of cross-country income distribution data implies that these effects need to be interpreted with caution.[45]

Countless other determinants have been suggested in the growth regressions literature. Without attempting an exhaustive analysis, we show how a few "usual suspects" are associated with the duration of declines in Table 7. For example, a key feature of poverty trap theories is that certain health and other conditions interact with the level of GDP per capita and produce multiple equilibria or so-called low-level traps. To capture the

[45] Not all authors take this into account. For example, Berg et al. (2012) emphasize that higher inequality reduces the duration of growth accelerations as a key result of their analysis, but do so without providing a critical discussion of the underlying inequality data. Their models are based on UNU-WIDER's Worldwide Income Inequality Database (WIID), whose data is not comparable across countries and often not even within countries without a battery of further adjustments. See Solt (2009), who provides an excellent discussion of these issues and offers possible remedies.

Table 7 – Other covariates

	Coefficient	SE	p-value	Exits	Spells	Years in Spell	log $\mathcal{L}$
Infant Mortality	0.014	0.006	0.02	48	58	348	-83.18
Life Expectancy	-0.060	0.030	0.05	48	58	348	-83.03
Schooling (Primary)	-0.356	0.096	0.00	46	56	327	-76.39
Schooling (Secondary)	-0.448	0.165	0.01	46	56	327	-79.76
Schooling (All)	-0.254	0.063	0.00	46	56	327	-76.17
Telephones per capita	-0.021	0.014	0.13	30	38	257	-52.57

Notes: All regressions also condition on (log) initial GDP per capita and real US interest rates.

role of such co-dependencies on the duration of declines, we investigate how mortality and life expectancy fare in our duration set-up. In this set of simple models, lower infant mortality and higher life expectancy are both associated with a shorter duration of decline spells. Further, both endogenous growth theory and modernization theory attach a large weight to human capital accumulation as a driver of growth. As Table 7 shows, countries where the population has more years of schooling (primary, secondary or total) experience shorter slumps – an effect that is also sizable compared to other coefficients. We also examine if infrastructure affects the duration (using telephone lines per capita as a proxy) and find it reduces the length of declines albeit insignificantly at the 10% level.

While the relationships uncovered in this section still have to prove robust in an expanded multivariate framework, the preceding analysis reduces the list of possible covariates to a smaller set of promising predictors. On the trade side, both de jure and de facto trade openness, as well as export structures, emerged as potentially relevant, while of the financial variables, only financial depth and private credit are strongly associated with duration outcomes. Debt and financial globalization do not play an important role in our sample of slumps, while human capital and other basic characteristics may have an effect. The particularly strong correlation between the duration of declines and institutional characteristics requires further examination. In addition schooling, mortality and life expectancy may play a role. The next section addresses the issue of omitted variables and provides an interpretation of the estimated effects.

5.3 Summary Models

We now turn to a set of summary models, highlighting which determinants of the duration of decline spells are robust to controlling for other potentially important effects. As the number of variables that can be included in a single run remains limited due to missing data and overfitting concerns, we present three sets of summary models. First, we examine how the effects of institutions and fractionalization change when other variables are added. Second, we present a set of results using our preferred specification as a base but adding other variables in thematic groups. Third, we show an expanded set of summary regressions highlighting the non-linearities involved in the effects of institutions and fractionalization on the duration of declines – a feature that has received too little attention in the empirical literature so far.

Table 8 (on the next page) reports the first set of summary regressions. The table is organized as follows. In addition to initial GDP per capita and the real US interest rate, all regressions include executive constraints as our primary proxy for institutions and ethno-linguistic fractionalization at the disaggregate level (ELF15). Subsequently, we enter each of the variables that were significant in previous section separately into the model in order to examine how this changes the partial regression coefficients.

The broad patterns are very interesting. Above all, the effect of fractionalization is extremely robust in all but one model[46] and varies only within a narrow band. Further, the coefficient of executive constraints is significant at the 5% or 1%-level in most of the regressions and has a stable negative sign throughout. However, it becomes small when we control for export sophistication, private credit to GDP and financial depth. The standard error of executive constraints also widens substantially when controlling for inequality, but the size of the coefficient remains stable. This is more due to the sharp reduction in sample size – a sample selection effect – when inequality enters the regression, rather than the partial effect of inequality itself.[47] In the case of export sophistication

[46]When also controlling for export diversification, we lose nearly half the same and ELF15 becomes insignificant. However, the sign remains stable and the standard error increases only marginally.

[47]We investigated this issue by running a regression on exactly the same estimation sample as if inequality were included, but dropping inequality. In this exercise, the coefficient of executive constraints is -.163 with a standard error of 0.102 which is just insignificant at the 10%-level and only about 0.01

Table 8 – Summary Models I

Variables	(1) $\ln \tilde{t}$	(2) $\ln \tilde{t}$	(3) $\ln \tilde{t}$	(4) $\ln \tilde{t}$	(5) $\ln \tilde{t}$	(6) $\ln \tilde{t}$	(7) $\ln \tilde{t}$	(8) $\ln \tilde{t}$	(9) $\ln \tilde{t}$	(10) $\ln \tilde{t}$	(11) $\ln \tilde{t}$	(12) $\ln \tilde{t}$
Executive Constraints	-0.195***	-0.176**	-0.149**	-0.178**	-0.025	-0.012	-0.024	-0.134	-0.178***	-0.155***	-0.156***	-0.172***
	(0.068)	(0.070)	(0.059)	(0.085)	(0.068)	(0.072)	(0.078)	(0.114)	(0.058)	(0.060)	(0.060)	(0.064)
Fractionalization (ELF15)	0.014***	0.019***	0.012***	0.008	0.015***	0.019***	0.019***	0.014***	0.016***	0.016***	0.016***	0.015***
	(0.005)	(0.004)	(0.004)	(0.006)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)
Initial ln GDP per capita	0.230	0.402***	0.089	0.120	0.423***	0.282**	0.286**	0.520**	0.188	0.374**	0.399**	0.401***
	(0.153)	(0.118)	(0.116)	(0.155)	(0.159)	(0.132)	(0.131)	(0.244)	(0.114)	(0.170)	(0.184)	(0.124)
Real US Interest Rate	0.080	0.066	0.078*	0.130	0.095*	0.106**	0.086	-0.019	0.088*	0.099**	0.097**	0.070
	(0.057)	(0.046)	(0.046)	(0.082)	(0.049)	(0.051)	(0.054)	(0.077)	(0.048)	(0.046)	(0.046)	(0.045)
Inflation ($\ln(1+\pi)$)	-0.003											
	(0.004)											
Trade Openness (de jure)		-0.807**										
		(0.329)										
Trade Openness (de facto)			0.009**									
			(0.004)									
Export Diversification				0.005								
				(0.010)								
Export Sophistication (2)					-1.500***							
					(0.439)							
Private Credit / GDP						-0.017***						
						(0.005)						
Financial Depth							-0.018**					
							(0.007)					
Gini (SWIID)								0.022				
								(0.022)				
Fractionalization (ELF1)									0.002			
									(0.009)			
Infant Mortality										0.007		
										(0.005)		
Life Expectancy											-0.029	
											(0.023)	
Schooling (All)												-0.088
												(0.069)
Constant	-0.632	-2.080**	0.532	0.441	10.035***	-1.324	-1.176	-4.164**	-0.479	-2.538	-0.542	-1.786**
	(1.219)	(0.966)	(1.005)	(1.476)	(3.166)	(1.039)	(1.030)	(2.004)	(0.930)	(1.724)	(0.936)	(0.884)
Exits	37	42	47	24	32	27	25	22	47	47	47	45
Spells	44	51	57	31	40	34	32	27	57	57	57	55
Years of Decline	232	314	346	236	290	196	193	137	346	346	346	325
Log-L	-53.908	-61.318	-70.346	-39.655	-50.304	-38.267	-36.857	-31.316	-72.051	-71.252	-71.192	-66.142
Pseudo-R^2	0.174	0.222	0.188	0.111	0.177	0.230	0.209	0.158	0.168	0.177	0.178	0.202

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

and the two financial variables, two factors could be driving the changes in the coefficient on institutions. On the one hand, this could be a sample selection effect as in the case of the inequality. On the other hand, these measures are strongly correlated both with institutional quality and initial GDP per capita. We find that in all three cases, sample selection is not the main culprit[48] but that there are strong grounds to suspect that both the ability to produce more sophisticated exports and sustain a complex financial system with a higher share of liquid liabilities and private credit must first be preceded by well-developed institutions. In fact, we can characterize these complex features of modern economies as outcomes of institutional development (e.g., see Acemoglu et al., 2003, for a similar point and evidence thereof). We return to this issue when discussing how we arrive at our preferred specification.

Several of the effects found in the previous variable-by-variable regressions turn out to be not robust in a multivariate setting. The coefficient of export diversification is now insignificant, as is the coefficient of inequality. The former now even points in the wrong direction, while higher inequality still predicts longer spells. Both do not add much explanatory power to the model. Further, the results also help to determine which of the two measures of fractionalization should be preferred in our context. Complementing the results of Desmet et al. (2012), we find that when controlling for both the disaggregate and aggregate measures of fractionalization at the same time, the disaggregate measure (ELF15) captures more of the relevant variation in the duration of declines. All of the "other" covariates, that is, life expectancy, infant mortality and years of schooling, are insignificant in the expanded models and have hardly any effect on the partial coefficients of institutions or fractionalization.

Table 9 takes a different approach to addressing the issue of omitted variables. Instead of adding variables one-by-one, we now add groups of variables measuring similar yet different aspects of a certain theme, such as trade or finance. Most of the previous

less than in the model with inequality. Hence, we conclude that this is mainly a sample selection effect.

[48] Contrary to the case of inequality, keeping the estimation sample constant but dropping the "troublemaker" reveals that the smaller sample results in some added imprecision but cannot fully account for the reduction in the coefficient's size for either of these three variables. All of these results are available upon request.

Table 9 – Summary Models II

VARIABLES	(1) $\ln \tilde{t}$	(2) $\ln \tilde{t}$	(3) $\ln \tilde{t}$	(4) $\ln \tilde{t}$	(5) $\ln \tilde{t}$	(6) $\ln \tilde{t}$
Executive Constraints	-0.138**	-0.156**	-0.007	-0.172***	-0.111*	-0.178***
	(0.062)	(0.075)	(0.075)	(0.063)	(0.064)	(0.058)
Fractionalization (ELF15)	0.014***	0.010*	0.019***	0.016***	0.014***	0.016***
	(0.005)	(0.006)	(0.005)	(0.004)	(0.005)	(0.004)
Initial ln GDP per capita	0.259*	0.610***	0.293**	0.559***	0.459***	0.197*
	(0.134)	(0.186)	(0.132)	(0.171)	(0.164)	(0.106)
Real US Interest Rate	0.053	0.133	0.098*	0.081*	0.074*	0.087*
	(0.044)	(0.085)	(0.054)	(0.044)	(0.042)	(0.048)
Trade Openness (de jure)	-0.818***					
	(0.314)					
Trade Openness (de facto)	0.010*					
	(0.005)					
Export Diversification		-0.017				
		(0.011)				
Export Sophistication (2)		-2.038***				
		(0.765)				
Private Credit / GDP			-0.011			
			(0.008)			
Financial Depth			-0.007			
			(0.013)			
Infant Mortality				0.003		
				(0.009)		
Life Expectancy				-0.020		
				(0.038)		
Schooling (All)				-0.026		
				(0.074)		
Constant	-0.677	14.123**	-1.346	-2.426	-3.346**	-0.553
	(1.216)	(5.672)	(1.035)	(2.633)	(1.401)	(0.868)
Region FE	NO	NO	NO	NO	YES	NO
Exits	42	22	25	45	47	47
Spells	51	28	32	55	57	57
Years of Decline	314	215	193	325	346	346
Log-L	-59.269	-33.830	-36.422	-65.321	-63.827	-72.090
Pseudo-R^2	0.248	0.172	0.218	0.211	0.263	0.168

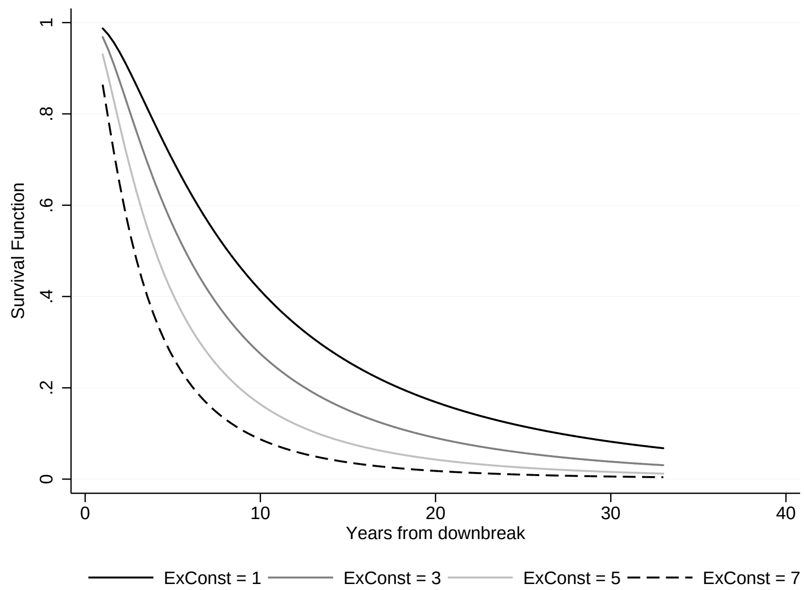
Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

results are confirmed, but Table 9 also introduces a few refinements. First and foremost, the effect of ethno-linguistic fractionalization remains very robust. Second, interesting patterns emerge for the three groups of macroeconomic variables. Model (1) shows that de jure trade openness still holds substantial explanatory power over the expected duration and de facto openness continues to have a marginally significant positive effect, yet both of these covariates have little effect on the coefficients and standard errors of institutions and fractionalization. Model (2) highlights that the coefficient of executive constraints is actually robust to the inclusion of export sophistication once the degree of export diversification is also controlled for.[49] For the financial variables, Model (3) reveals that while the model with private credit of GDP and financial depth still reduces

[49] The coefficient of ELF15 is somewhat less significant in this model, but this is due to the smaller sample size.

the coefficient on executive constraints by a lot, these variables are not very robust themselves. Including both of these correlated but different measures suffices to make their coefficient insignificant.[50] Model (4) confirms that neither of life expectancy, infant mortality of schooling have robust effects on the duration. In addition, in Model (5) we take a different approach to addressing the issue of omitted variables by including region dummies to account for regionally shared heterogeneity that is otherwise not captured by the observed covariates. Both the coefficients and standard errors of institutions and fractionalization remain within their common range, providing further support for the assertion that the effect of institutions is reasonably robust. Model (6) in the last column is our preferred and most parsimonious specification. This model captures most of the effects we are interested in and all of the measured effects, as well as their standard errors, remain well within the usual range. Taken together, these models show that the effect of fractionalization is very robust and the effect of institutions is only strongly affected by the financial depth and credit – two measures that, according to our interpretation, are observed institutional outcomes.

Figure 12 – Predicted survival functions



So far we have not investigated if the substantive effects implied by these models

[50]This also applies to many other models where additional controls are used on top either one of the financial measures.

have a meaningful size. Figure 12 examines this point by plotting the survival functions predicted by our preferred specification over four different values of executive constraints (while keeping all other variables at their sample mean). As is readily observed, the effect of institutions on the expected duration is very large but still plausible. In most cases, we would compute the predicted mean durations by integrating these curves from analysis time zero (the onset of the decline) to infinity. However, for the log-normal model there is a simpler way, since mean and median are equivalent and can be easily estimated by the exponentiated linear prediction.[51] Conditional on having entered a slump as defined by our method, a country with the lowest score on the executive constraints measure is expected to decline for about 9.1 years, while a country with the highest score is expected to decline only for about 3.1 years. The mean of executive constraints in the estimation sample is only about 2.4, implying an expected duration of 7.7 years.

As a last set of summary results, Table 10 reproduces Table 9 with one important change. Instead of imposing that institutions and fractionalization have a linearly additive effect on the expected duration of declines, we now allow for an interaction effect between the two. The rationale behind this is simple. Given a political economy in which (latent) social conflict challenges the ability of political actors to take coordinated action, better institutions – property rights or otherwise – may help to overcome this negative effect. However, countries with a very high degree of fractionalization may require particularly strong institutions just to compensate for the negative effect of the former. Likewise, countries with much greater degree of ethno-linguistic homogeneity may make do with less developed institutions to achieve a similar degree of social coordination. This hypothesis is a less restrictive variant of the idea that there is a multiplicative effect between social conflict and institutions in response to external shocks (Rodrik, 1999).[52] Table 10 shows

[51]This is a special property of the log-normal model. Since $\sigma\epsilon \sim \mathcal{N}(0, \sigma)$, which is symmetric at zero, we have $E[\ln \tilde{t}|\mathbf{x}] = Q_{0.5}[\ln \tilde{t}|\mathbf{x}]$ because $E[\ln \tilde{t}|\mathbf{x}] = \mathbf{x}'\boldsymbol{\beta} + E[\sigma\epsilon|\mathbf{x}]$ and $E[\sigma\epsilon|\mathbf{x}] = Q_{0.5}[\sigma\epsilon|\mathbf{x}] = 0$ by construction. The expected median and mean are both estimated by the linear prediction, as the predicted median of $\ln \tilde{t}$ is the median of the error distribution (i.e. zero) plus the linear prediction. Most other error distributions invalidate this equivalence.

[52]Rodrik (1999) shows that such a multiplicative effect exists when looking at growth rate differentials, but constrains the effect to be multiplicative only by not including the base categories of either institutions or measures of conflict. In this respect, this is very different to proposing the existence of a non-linear interaction as we suggest here.

Table 10 – Summary Models III

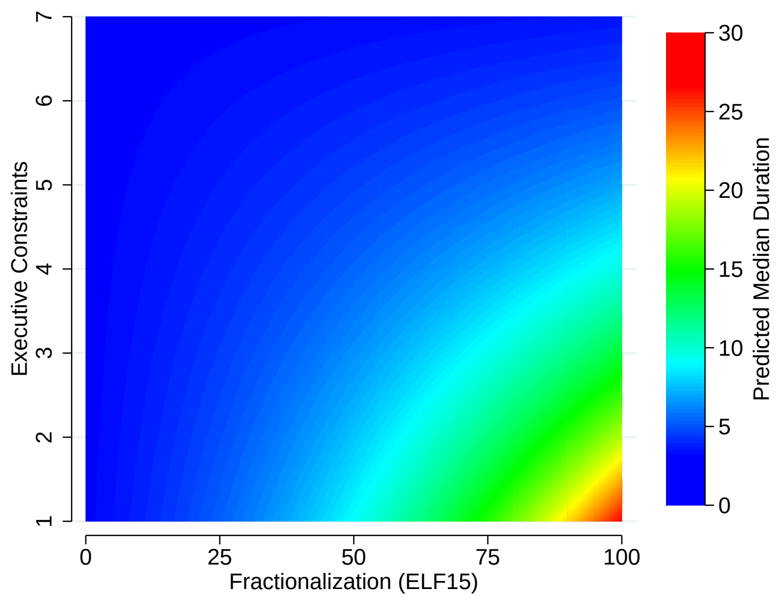
VARIABLES	(1) $\ln \tilde{t}$	(2) $\ln \tilde{t}$	(3) $\ln \tilde{t}$	(4) $\ln \tilde{t}$	(5) $\ln \tilde{t}$	(6) $\ln \tilde{t}$
Executive Constraints$^{\sim}$	-0.280***	-0.220*	-0.091	-0.280***	-0.217**	-0.288***
	(0.096)	(0.113)	(0.101)	(0.090)	(0.087)	(0.080)
Fractionalization (ELF15)$^{\sim}$	0.016***	0.012**	0.019***	0.017***	0.017***	0.018***
	(0.004)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Executive Constraints$^{\sim}$ × ELF15$^{\sim}$	-0.004***	-0.003	-0.003	-0.004**	-0.003**	-0.004***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Initial ln GDP per capita	0.265*	0.642***	0.238*	0.533***	0.439***	0.198*
	(0.138)	(0.183)	(0.135)	(0.175)	(0.163)	(0.106)
Real US Interest Rate	0.059	0.136	0.100*	0.084*	0.076*	0.098**
	(0.042)	(0.089)	(0.054)	(0.043)	(0.042)	(0.047)
Trade Openness (de jure)	-0.757**					
	(0.318)					
Trade Openness (de facto)	0.012**					
	(0.005)					
Export Diversification		-0.019				
		(0.012)				
Export Sophistication (2)		-2.173***				
		(0.792)				
Private Credit / GDP			-0.008			
			(0.009)			
Financial Depth			-0.008			
			(0.011)			
Infant Mortality				0.001		
				(0.007)		
Life Expectancy				-0.020		
				(0.030)		
Schooling (All)				-0.045		
				(0.070)		
Constant	-0.144	15.389**	0.213	-1.425	-2.552*	0.025
	(1.153)	(6.005)	(1.247)	(2.066)	(1.362)	(0.872)
Region FE	NO	NO	NO	NO	YES	NO
Exits	42	22	25	45	47	47
Spells	51	28	32	55	57	57
Years of Decline	314	215	193	325	346	346
Log-L	-55.312	-33.303	-35.514	-62.746	-61.681	-69.540
Pseudo-R^2	0.298	0.185	0.238	0.243	0.288	0.197

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

that there is substantial evidence of such an effect and that the interaction term always points in the hypothesized direction. In order to ease the interpretation of the coefficients, we subtract the sample average from the institutions and fractionalizaton variables before estimating each model (denoted with superscript '$\sim$'). This has the following effect. If either one of the two variables is at their mean value, then the interaction term is zero and the only relevant coefficient is the non-interacted variant. As a result, the coefficient of the executive constraints variable directly measures the effects at the average level of fractionalization, and vice versa. For values other than the mean, the coefficient on the interaction term need to be taken into account.

Table 10 suggests that in the same models where we find a robust effect of institutions, we also find a significant interaction effect between executive constraints and ethno-linguistic fractionalization. In most models, the effect of one variable at the mean of the other is at least as significant as in the models without an interaction effect and the interaction term is negative and significant at the 1%-level in all versions but models (2) and (3). Even in the latter two models, the sign, size and standard errors of the interaction term are similar to the other results. Since our preferred specification is nested in Model (6) of this table, testing the null hypothesis that the interaction term is zero is equivalent to a test that the expanded model fits the data better.[53] We refer to this model as the expanded version of our preferred specification, or simply the interaction model.

Figure 13 – Median exit times: interaction model (interpolated predictions)



Does the non-linear interaction model make sense in terms of its substantive predictions? Figure 13 visualizes how the predicted mean/median exit times vary over the entire range of the theoretical distribution of executive constraints and ethno-linguistic fractionalization. The figure clearly illustrates that it would be difficult to understand the effects of institutions without accounting for the effects of fractionalization. The increasing curvature of the contours is the graphical representation of the interaction effect. Since missing data points are interpolated, the graph needs to be interpreted with

[53]This is corroborated by other formal tests or measures of fit. A likelihood ratio test picks the interaction model over the simple model and the pseudo-R^2 improves to 0.197 over 0.168.

caution, particularly at the extremes of both scales.[54] As hypothesized, the greater the degree of fractionalization, the better institutions are required to achieve shorter declines. This effect is very strong, particularly at lower degrees of fractionalization/ greater homogeneity, where the model suggests that comparatively short declines will happen regardless of the score on executive constraints. The wide range of this distribution is actually covered in reality. The sampling distribution of executive constraints covers the entire theoretical range (scores 1 to 7) and ethno-linguistic fractionalization ranges from a near perfect homogeneity (0.07) to near perfect fractionalization (96). A positive interpretation of this result would be that stronger institutions have the potential to overcome even very high degrees of latent social conflict/ fractionalization.

5.4 Robustness

In this part, we briefly show that our main results are robust to the choice of hazard shape, presence of unobserved heterogeneity, exclusion of influential groups of observations, and different ways of accounting for recurrent spells. We run a battery of statistical tests for these common sources of model misspecification and describe in more detail how the choice of hazard function relates to the time process underlying the different models.

Table 11 tackles the issues of choice of functional form and model selection. To aid a direct comparison, we again report the preferred specification in column one and then show the estimates from fitting five models each using a different hazard function. Model (2) uses a log-logistic hazard instead of the log-normal shape, but the parameter estimates do not change much. This is not too surprising. The log-logistic distribution is very similar to the log-normal, in that it offers a non-monotonic shape that is either first increasing and then decreasing or monotonically decreasing throughout. This model has an additional parameter (γ) which tells us which of the two is the case. The estimated shape parameter ($\ln \gamma$) is negative and significant, implying that the hazard is first increasing then decreasing just as in the log-normal model. The log-likelihood is close to

[54]Figure 13 is asking a lot from comparatively few data points. First, we estimate the model using 57 spells, then we predict are large set of combinations between the two variables, and then we interpolate these predictions over the entire continuous range shown in the graph.

that of model (1) but not better, indicating a similar or, at best, minimally worse fit.

Table 11 – Robustness: functional form

Variables	(1) Log-normal $\ln \tilde{t}$	(2) Log-logistic $\ln \tilde{t}$	(3) Exponential PH $\ln \tilde{t}$	(4) Weibull PH $\ln \tilde{t}$	(5) Gompertz PH $\ln \tilde{t}$	(6) Cox PH $\ln \tilde{t}$
	Coefficients		*Hazard Ratios*			
Executive Constraints	-0.178***	-0.185***	1.229***	1.263***	1.222***	1.194***
	(0.058)	(0.067)	(0.074)	(0.089)	(0.071)	(0.070)
Fractionalization (ELF15)	0.016***	0.016***	0.978***	0.974***	0.979***	0.981***
	(0.004)	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)
Initial ln GDP per capita	0.197*	0.235**	0.787	0.765	0.786*	0.814
	(0.106)	(0.112)	(0.119)	(0.146)	(0.113)	(0.119)
Real US Interest Rate	0.087*	0.084*	0.947	0.928	0.949	0.952
	(0.048)	(0.051)	(0.058)	(0.061)	(0.057)	(0.054)
$\ln \gamma$ (Log-Logistic)		-0.580***				
		(0.105)				
$\ln p$ (Weibull)				0.198**		
				(0.087)		
γ (Gompertz)					-0.015	
					(0.030)	
Constant	-0.553	-0.856	1.830	1.723	1.928	
	(0.868)	(0.901)	(2.432)	(2.884)	(2.448)	
VCE	cluster	cluster	cluster	cluster	cluster	cluster
Exits	47	47	47	47	47	47
Spells	57	57	57	57	57	57
Years of Decline	346	346	346	346	346	346
Log-L	-72.090	-73.286	-75.295	-73.940	-75.192	-149.794
AIC	156.180	158.571	160.590	159.879	162.384	307.588
Pseudo-R^2	0.168	0.164	0.208	0.210	0.160	0.071

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

Models (3) to (6) in Table 11 have a different interpretation than all of the AFT models shown previously. We no longer report coefficients but instead *hazard ratios*, since these models are proportional hazards (PH) models by nature and only the Weibull and exponential distribution also fit into an equivalent AFT formulation. Their interpretation is as follows. A hazard ratio of unity means the probability of exiting the decline phase is unaffected by the covariate. A hazard ratio greater than one implies an increase in the hazard rate, so a higher instantaneous probability of exiting the decline, and a hazard ratio below one implies a lower instantaneous probability of exiting the spell. The first model in column four is the exponential or constant hazard model. Here too, the results are quantitatively and qualitatively very similar (given the altered interpretation), but the log-likelihood decreases somewhat and we have no reason to suspect a constant hazard. Model (4) uses a Weibull parametrization which allows for monotonically increasing or decreasing hazard rates. This model also has a shape parameter (p) which allows testing for a constant hazard and, if constancy is rejected, indicates whether the rate increases or decreases. A Wald test of the null hypothesis that $\ln p = 0$ rejects, so the baseline hazard is not constant. However, the Weibull model suggests an increasing hazard contrary to all other parametrizations. The Gompertz model in column five also offers a monotonically increasing or decreasing hazard. As before, the results remain similar and the shape is estimated as monotonically decreasing (as indicated by $\gamma < 0$).

How do we assess which hazard shape fits the data best and makes the most sense? The question of fit is easily answered by the highest log-likelihood, which in our case clearly prefers the log-normal model but is not much lower for the log-logistic and Weibull models. However, a comparison of likelihoods does not account for the additional parameters introduced by the varying functional forms. The Akaike information criterion (AIC) can account for such differences and is a common tool for selecting among non-nested models. The AIC is lowest for the log-normal model, confirming our choice. While these two indicators of model fit and penalized model fit help choosing among models, they do not tell us what the underlying baseline hazard truly looks like. For this we turn to model (6) in the last column, where we specify a Cox proportional hazard model[55] which leaves the baseline hazard unspecified. Then, we extract the shape of the hazard function from the model. The predicted baseline hazard of the Cox model agrees that the probability of exiting a spell first increases and then decreases. Figure 14 in Appendix E illustrates this by plotting all of the estimated baseline hazard functions. Such a non-monotonic shape lends itself to the following interpretation. In the first few years of a decline, the economy is suffering from a harsh but possibly temporary shock and is thus more likely recover quickly. However, the longer the decline lasts, the more the economic fundamentals deteriorate and it becomes increasingly difficult for countries to enter the recovery phase.

Next, we turn to the issues of unobserved effects, influential observations and recurrent spells. The previous section has already shown that the effects in our preferred specification are robust to the inclusion of regional fixed-effects. Model (1) in Table 12

[55]This model is also known as the Anderson-Gill (AG) model of recurrent events (Andersen and Gill, 1982). However, the AG model still assumes that repeated events are constitutionally independent, an assumption which we relax further below.

Table 12 – Robustness: heterogeneity, multiple failures and dropping regions

Variables	(1) Full $\ln \tilde{t}$	(2) No SSA $\ln \tilde{t}$	(3) No MNA $\ln \tilde{t}$	(4) No LAC $\ln \tilde{t}$	(5) No multiple $\ln \tilde{t}$	(6) PWP multiple $\ln \tilde{t}$
Executive Constraints	-0.161***	-0.159***	-0.199***	-0.189**	-0.199***	1.263***
	(0.061)	(0.055)	(0.071)	(0.074)	(0.064)	(0.096)
Fractionalization (ELF15)	0.012**	0.005	0.018***	0.016***	0.015***	0.981***
	(0.006)	(0.004)	(0.005)	(0.005)	(0.004)	(0.006)
Initial ln GDP per capita	0.213*	0.358***	0.263*	0.179	0.196*	0.759
	(0.111)	(0.101)	(0.139)	(0.115)	(0.111)	(0.137)
Real US Interest Rate	0.091**	0.103**	0.066	0.090	0.086	0.940
	(0.039)	(0.042)	(0.048)	(0.065)	(0.060)	(0.067)
Constant	-0.767	-1.843**	-1.055	-0.393	-0.405	
	(0.943)	(0.936)	(1.063)	(0.870)	(0.900)	
VCE	–	cluster	cluster	cluster	cluster	cluster
Frailties	shared	–	–	–	–	–
Strata	–	–	–	–	–	spell #
Exits	47	40	43	34	40	47
Spells	57	44	50	43	50	57
Years of Decline	346	178	294	271	312	346
Log-L	-71.867	-50.584	-64.255	-54.643	-63.715	-123.435
Pseudo-R^2	0.111	0.151	0.163	0.172	0.162	0.095

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

goes a step further and includes country-level effects in the model. Each country now has an unobserved individual effect or so-called gamma distributed frailty. These frailties are the duration analysis equivalent of random effects in linear models. The term "frailty" derives from the notion that any one (clinical) subject may be more "frail" than the average, that is, more disposed to experiencing a certain event than others. Just as in linear models, these random effects are assumed to be uncorrelated with any of the other included covariates (which is unlikely to hold). The results show that the main effects are very robust to controlling for this type of restricted country-specific heterogeneity. Further, there is only weak evidence in favor of unobserved heterogeneity altogether.[56]

Models (2) to (4) examine if any of our main results are driven by specific regions with particularly long slumps. We address this issue by re-estimating our preferred model multiple times, each time removing one of the three regions with the longest spells. Model (2) drops all declines in Sub-Saharan Africa (SSA) and reveals an interesting additional finding. While the coefficient of fractionalization (ELF15) is very robust in the previous models, its size and significance is clearly driven by African observations. Without those, the coefficient keeps the same sign but shrinks substantially and becomes

[56]A Likelihood Ratio test for the presence of shared frailties ($\mathbb{H}_0 : \theta = 0$) rejects the null ($p = 0.252$).

insignificant at conventional levels. Since Sub-Saharan Africa has the greatest ethno-linguistic heterogeneity of all regions, this result is not too surprising and does not necessarily invalidate the presence of such an effect in general.[57] The interaction model proposed earlier may thus be more relevant to understanding the effects of institutions and fractionalization in Africa than elsewhere.[58] On the other hand, Models (3) and (4) show that the parameter estimates are not sensitive to excluding either the entire Middle East and North Africa (MNA) or all of Latin America and the Caribbean (LAC).

Until now, we have assumed that multiple spells of the same type are indistinguishable among another. The last two columns of Table 12 investigate if this relatively strong form of conditional independence is a reasonable assumption. Model (5) shows that our findings are robust to excluding all spells other than the first, which rules out any dependency across recurrent spells. The coefficient of executive constraints becomes even larger and the effect of fractionalization is virtually unchanged. Model (6) takes a different approach and specifies a conditional risk set model or stratified Cox model due to Prentice, Williams and Peterson (1981, henceforth PWP). The PWP model accounts for ordering of the events but assumes that a subject cannot experience another event until the previous event has occurred. In our case, this is a natural assumption, as – by definition – a country cannot exit a second decline phase before having left the first and so on. The PWP model is easily estimated by stratifying on the spell number and allowing the errors to be clustered across countries. Again, the results remain qualitatively and quantitatively similar, although the reported hazard ratios cannot be directly compared to the coefficients of the log-normal model.

In sum, we find that a log-normal hazard shape is not only a flexible assumption but also fits the data very well. It is preferred by model selection criteria and shaped similarly to the baseline hazard estimated via a semi-parametric Cox model. Further, the results are robust to allowing for a restricted form of unobserved heterogeneity, dropping of influential regions, and accounting for dependency among recurrent events. An important

[57] In out sample, the average ELF15 score in Sub-Saharan Africa is 87 out of 100, compared to 62 in the Middle East and North Africa and 34 in Latin America and the Caribbean.

[58] Tables for the robustness of the interaction model are available from the authors upon request.

additional insight is that the effect of fractionalization is driven by Sub-Saharan Africa, where fractionalization is highest and declines last the longest on average.

6 Conclusion

This paper makes three contributions to a burgeoning literature on growth episodes and structural breaks in growth performances. First, we show that a restricted structural change approach, as in Papell and Prodan (2011b), works well as an inferential method for identifying slumps, big recessions or growth collapses in a large sample of countries. We find a substantial number of slumps in developing and developed countries alike, suggesting that severe growth volatility is a ubiquitous phenomenon in the post-war period. Second, the slumps we identify have interesting characteristics, in excess of the expected macroeconomic symptoms, some of which have received little attention so far. Most prominently, we find systematic evidence of institutional underdevelopment before slumps hit and positive institutional change during and in the immediate aftermath of slumps. Hence, institutions may not only cause growth, but volatility can also contribute to institutional change. Severe economic crises bring about what we call creative political destruction and raise the pressure for institutional reform in a very broad sense. Third, relying on an inferential approach to identifying slumps provides much clearer results than an earlier study of the duration of growth collapses (Hausmann et al., 2008). We find robust evidence that the duration until the exit of the decline phase depends on institutions and particularly strongly on ethno-linguistic fractionalization. Further, we show that this effect is likely to be non-linear, in a more unrestricted manner than suggested previously.

As a whole, our results lend broad support to political economy theories stressing the respective roles of institutions and social conflict. Effective social coordination and responses to slumps are hampered by a high degree of (latent) social tension as captured by ethno-linguistic fractionalization but particularly strong institutions can put the coordination and legal mechanisms in place that are able contain or resolve these conflicts within the institutional framework. On the other hand, these results give rise to the interesting proposition that in less fragmented societies, as defined by ethnic cleavages, institutions are much less important as a determinant of the length of declines. These results do not suggest that sound macroeconomic policies as such do not matter, but they provide some indication that these policies may be secondary to more fundamental factors. In addition, while the previous literature has stressed the role of positive growth spurts, we show that slumps matter a lot and that the decline phase can last very long in some cases. In fact, given that growth has been found comparatively easy to ignite but difficult to sustain, a comparison of the relative effects of slumps versus accelerations on long-run GDP levels would be an interesting extension of our findings.

Many avenues are still left unexplored by this paper in particular and by the related literature in general. For example, we did not analyze the determinants of the depth of slumps, which is a natural extension to a study of their duration. More work can be done on nesting different models of restricted structural change and statistically testing which pattern fits the data better. Last but not least, much of the growth episodes literature still remains void of convincing causal analysis. Future research should focus more on exploring the causal factors that are behind the occurrence, duration and magnitude of different growth episodes – a very challenging but equally exciting area of research.

References

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

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

Acemoglu, D., J. A. Robinson, and T. Verdier. 2004. "Alfred Marshall Lecture: Kleptocracy and divide-and-rule: a model of personal rule." *Journal of the European Economic Association*, *2* (2-3): 162–192.

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, and J. A. Robinson. 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics*, *117* (4): 1231–1294.

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.

Alesina, A., R. Baqir, and W. Easterly. 1999. "Public goods and ethnic divisions." *The Quarterly Journal of Economics*, *114* (4): 1243–1284.

Andersen, P. K. and R. D. Gill. 1982. "Cox's Regression Model for Counting Processes: A Large Sample Study." *The Annals of Statistics*, *10* (4): 1100–1120.

Andrews, D. W. K. 1993. "Tests for parameter instability and structural change with unknown change point." *Econometrica: Journal of the Econometric Society*, pp. 821–856.

Andrews, D. W. K. and W. Ploberger. 1994. "Optimal tests when a nuisance parameter is present only under the alternative." *Econometrica: Journal of the Econometric Society*, pp. 1383–1414.

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. "A New Data Set of Educational Attainment in the World, 1950-2010." Working Paper 15902, National Bureau of Economic Research 2010.

Beck, T., A. Demirgüç-Kunt, and R. Levine. 2010. "Financial Institutions and Markets across Countries and over Time: The Updated Financial Development and Structure Database." *The World Bank Economic Review*, *24* (1): 77–92.

Beck, T., R. Levine, and N. Loayza. 2000. "Finance and the sources of growth." *Journal of Financial Economics*, *58* (12): 261 – 300.

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.

Bluhm, R., D. de Crombrugghe, and A. Szirmai. "Explaining the dynamics of stagnation: An empirical examination of the North, Wallis and Weingast approach." UNU-MERIT Working Paper Series 040, United Nations University, Maastricht Economic and social Research and training centre on Innovation and Technology 2012.

Box-Steffensmeier, J. M. and B. S. Jones. 2004. *Event history modeling: A guide for social scientists*, Cambridge University Press.

Bussière, M. and M. Fratzscher. 2006. "Towards a new early warning system of financial crises." *Journal of International Money and Finance*, *25* (6): 953–973.

Calvo, G. A., A. Izquierdo, and E. Talvi. "Phoenix Miracles in Emerging Markets: Recovering without Credit from Systemic Financial Crises." Working Paper 12101, National Bureau of Economic Research 2006.

Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2011. "Robust Inference With Multiway Clustering." *Journal of Business & Economic Statistics*, *29* (2): 238–249.

Cerra, V. and S. C. Saxena. 2008. "Growth Dynamics: The Myth of Economic Recovery." *American Economic Review*, *98* (1): 439–57.

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

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

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

Easterly, W. and R. Levine. 1997. "Africa's growth tragedy: policies and ethnic divisions." *The Quarterly Journal of Economics*, *112* (4): 1203–1250.

Easterly, W., M. Kremer, L. Pritchett, and L.H. Summers. 1993. "Good Policy or Good Luck?" *Journal of Monetary Economics*, *32* (3): 459–483.

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

Frankel, J. and G. Saravelos. 2012. "Can leading indicators assess country vulnerability? Evidence from the 2008–09 global financial crisis." *Journal of International Economics*, *87* (2): 216 – 231.

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

Goemans, H. E., K. S. Gleditsch, and G. Chiozza. 2009. "Introducing Archigos: A Dataset of Political Leaders." *Journal of Peace Research*, *46* (2): 269–283.

Gourinchas, P.-O. and M. Obstfeld. 2012. "Stories of the Twentieth Century for the Twenty-First." *American Economic Journal: Macroeconomics*, *4* (1): 226–265.

Greif, A. and D. D. Laitin. 2004. "A Theory of Endogenous Institutional Change." *American Political Science Review*, *98* (04): 633–652.

Gutierrez, R. G. 2002. "Parametric frailty and shared frailty survival models." *Stata Journal*, *2* (1): 22–44.

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

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

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

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

Jones, B. F. and B. A. Olken. 2008. "The Anatomy of Start-Stop Growth." *The Review of Economics and Statistics*, *90* (3): pp. 582–587.

Jonung, L. and T. Hagberg. "How costly was the crisis of the 1990s? A comparative analysis of the deepest crises in Finland and Sweden over the last 130 years." Economic Papers 224, European Commission 2005.

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

Kaminsky, G. L. and C. M. Reinhart. 1999. "The twin crises: the causes of banking and balance-of-payments problems." *American Economic Review*, *89* (3): 473–500.

King, R. G. and R. Levine. 1993. "Finance and growth: Schumpeter might be right." *The Quarterly Journal of Economics*, *108* (3): 717–737.

Klomp, J. and J. de Haan. 2009. "Political institutions and economic volatility." *European Journal of Political Economy*, *25* (3): 311–326.

Kose, M. A., E. Prasad, K. Rogoff, and S. J. Wei. 2009. "Financial globalization: A reappraisal." *IMF Staff Papers*, *56* (1): 8–62.

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

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

Lucas, R. E. 1990. "Why doesn't capital flow from rich to poor countries?" *The American Economic Review*, *80* (2): 92–96.

MacKinnon, J.G. 2009. "Bootstrap Hypothesis Testing." in D.A. Belsley and E.J. Kontoghiorghes, eds., *Handbook of Computational Econometrics*, John Wiley & Sons, Ltd, chapter 6, pp. 183–213.

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

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

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

Papell, D. H. and R. Prodan. "Long-Run Time Series Tests of Constant Steady-State Growth." Manuscript, Univeristy of Houston 2011.

Papell, D. H. and R. Prodan. "The Statistical Behavior of GDP after Financial Crises and Severe Recessions." in "Proceedings of the 56th Economic Conference, Federal Reserve Bank of Boston" 2011.

Perron, P. and Z. Qu. 2006. "Estimating restricted structural change models." *Journal of Econometrics*, *134* (2): 373 – 399.

Petersen, M. A. 2009. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." *Review of Financial Studies*, *22* (1): 435–480.

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

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

Prodan, R. 2008. "Potential Pitfalls in Determining Multiple Structural Changes With an Application to Purchasing Power Parity." *Journal of Business & Economic Statistics*, *26* (1): 50–65.

Reddy, S. and C. Minoiu. 2009. "Real Income Stagnation of Countries 1960–2001." *Journal of Development Studies*, *45* (1): 1–23.

Reinhart, C. M. and K. S. Rogoff. 2004. "The Modern History of Exchange Rate Arrangements: A Reinterpretation." *The Quarterly Journal of Economics*, *119* (1): 1–48.

Rodriguez, F. and D. Rodrik. "Trade Policy and Economic Growth: A Skeptic's Guide to Cross-National Evidence." Working Paper 7081, National Bureau of Economic Research 1999.

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

Rodrik, D. 2007. *One economics, many recipes: globalization, institutions, and economic growth*, Princeton, NJ: Princeton University Press.

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

Sachs, J. D. and A. Warner. 1995. "Economic reform and the process of global integration." *Brookings papers on economic activity*, *1995* (1): 1–118.

Silverberg, G. and B. Verspagen. 2003. "Long Memory and Economic Growth in the World Economy Since the 19th Century." in Govindan Rangarajan and Mingzhou Ding, eds., *Processes with Long-Range Correlations*, Vol. 621 of *Lecture Notes in Physics*, Springer Berlin Heidelberg, pp. 270–285.

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

Thompson, S. B. 2011. "Simple formulas for standard errors that cluster by both firm and time." *Journal of Financial Economics*, *99* (1): 1–10.

Wacziarg, R. and K. H. Welch. 2008. "Trade Liberalization and Growth: New Evidence." *The World Bank Economic Review*, *22* (2): 187–231.

Winters, L. A. 2004. "Trade Liberalisation and Economic Performance: An Overview." *The Economic Journal*, *114* (493): F4–F21.

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

Zivot, E. and D. W. K Andrews. 1992. "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis." *Journal of Business & Economic Statistics*, *10* (3): 251–270.

A Appendix: Estimation of Structural Breaks

A.1 Sequential procedure for testing and dating breaks

The procedure described here is a modification of Bai's (1997) sequential likelihood ratio tests for structural change – see also the extensions in Bai and Perron (1998) and in Bai (1999). We make an important simplifying assumption, namely, that all output series are regime-wise trend-stationary. Verifying this assumption is beyond the scope of this paper, as testing for unit roots in the presence of structural breaks (with sufficient power and size) is still contested territory and our output series have only a moderate time dimension ($T < 60$).[59] Our implementation of the sequential procedure involves six steps:

1. Determine the optimal $AR(p)$ trend model using the Schwarz 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:

$$y_t = \alpha+\beta t+\gamma_1\mathbf{1}(t > tb_1)+\gamma_{12}(t-tb_1)\mathbf{1}(t > tb_1)+\gamma_2(t-tb_2)\mathbf{1}(t > tb_2)+\sum_{i=1}^{p}\delta_i y_{t-i}+\epsilon_t$$

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

3. Define trimming parameter τ, where typically $\tau \in [0.05, 0.25]$. The resulting estimation sample (Λ_τ) runs over $[\tau T, (1-\tau)T]$.[60] The breaks are in the ranges $\widehat{tb}_1 \in [\tau T, (1-\tau)T - h]$ and $\widehat{tb}_2 \in [\tau T + h, (1-\tau)T]$. We set $\tau = 0.05$.

4. Compute the sup-W test statistic of the null of no break versus two breaks ($\mathbb{H}_0$:

[59] However, using longer time-horizons Papell and Prodan (2011a) recently documented that the unit root hypothesis can be rejected for many countries once endogenous structural breaks are accounted for.

[60] For simplicity of exposition, we suppress an additional index running over the sub-samples (defined in Step 6). In other words, here T refers to the number of observations of the currently active sample.

$\gamma_1 = \gamma_{12} = \gamma_2 = 0$) over $t \in \Lambda_\tau$, under the *restrictions* that $\beta > 0$ and $\gamma_1 < 0$:

$$\sup_{t\in\Lambda_\tau} W(\widehat{tb}_1^1, \ldots, \widehat{tb}_1^c; \widehat{tb}_2^1, \ldots, \widehat{tb}_2^d; q) = \sup_{t\in\Lambda_\tau} \left(\frac{T-k_m}{q}\right) \frac{SSR_T^r - SSR_T^u}{SSR_T^u}$$

where $q = 3$, k_m is the number of parameters and SSR_T^r denotes the sum of squared residuals under $\mathbb{H}_0$ and SSR_T^u are the sum of squared residuals under $\mathbb{H}_A$.

5. The critical value and empirical p-value of $\sup_{t\in\Lambda_\tau} W(\widehat{tb}_1, \widehat{tb}_2; q)$ bootstrapped, as in finite samples comparable asymptotic tests often have poor size and power.[61]

6. If the sup-W statistic is significant at the desired level, the sample is split into two new sub-samples from the beginning to the first break and from the third break to the end, then the procedure restarts at (4) using the estimated AR-order (p) from before. If either the bootstrapped $\sup W^*$ tests rejects, or the sample gets too small ($T \le 20$), then the procedure stops and all break pairs have been found.

A.2 Bootstrapping the sup-Wald statistic

There have been several suggestions on how to best bootstrap structural change tests in particular or other popular time-series tests in general. For example, Hansen (2000) suggests employing a fixed-design bootstrap allowing for non-stationarity, lagged dependent variables and conditional heteroskedasticity. MacKinnon (2009), on the contrary, shows that the recursive bootstrap of Diebold and Chen (1996) gives results superior to most other bootstrap types (fixed-parameter, Sieve, pairs, block, double block) and the asymptotic test in a simple application of an AR(1) model with an endogenous break. Similarly, Papell and Prodan (2011b) also favor a recursive bootstrap but do not compare it to other methods. Hence, we use a recursive bootstrap similar to Diebold and Chen (1996) as comparing these methods systematically is also well beyond the scope of this paper.[62] In line with usual notation, we denote all bootstrap quantities with the

[61] See, for example, Prodan (2008) who documents such poor finite sample properties for the Bai-Perron multiple structural breaks procedure and recommends the bootstrap.

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

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 Schwartz criterion as before and obtain the residuals:

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

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

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

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

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

4. Rerun the break search algorithm on the bootstrap series $\{y^*_t\}$, including determination of the optimal AR(p) model, and compute bootstrapped $\sup W^*$ test statistic $\sup^i_{t\in\Lambda_\tau} W^*(\widehat{tb}^*_1, \widehat{tb}^*_2; q)$, where i indexes the current bootstrap iteration.

5. Repeat from Step (2) until $i = B$, 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 the estimated bootstrap test statistics are greater than the originally estimated test statistic:

$$\hat{p}^* = \frac{1}{B}\sum_{i=1}^{B} \mathbf{1}\left(\sup^{i}_{\widehat{tb}^*_1, \widehat{tb}^*_2 \in \Lambda_\tau} W^*(\widehat{tb}^*_1, \widehat{tb}^*_2; q) > \sup_{\widehat{tb}_1, \widehat{tb}_2 \in \Lambda_\tau} W(\widehat{tb}_1, \widehat{tb}_2; q) \right)$$

and the critical value corresponds to the $i^{\text{th}} = (1 - \alpha^s)B$ element of the sorted vector of bootstrap statistics $\Gamma = [\sup^{min}_{t\in\Lambda_\tau} W^*(\cdot), \ldots, \sup^{max}_{t\in\Lambda_\tau} W^*(\cdot)]$, where α^s is the desired significance level (10% throughout the text, if not otherwise noted).

B Appendix: List of Episodes

Table 13 – Global Parameters

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

Table 14 – Estimated and Filtered Breaks with Troughs: 58 Episodes*

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

Continued on next page

Table 14 – *Continued from previous page*

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

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

C Appendix: Data Sources and Summary Statistics

Table 15 – Summary Statistics: break date to trough

Variable	Source	Mean	Std. Dev.	$T \times N$
	Institutions, Politics & Conflict			
Polity Score	Polity IV	-1.905	6.991	346
Democracy	Polity IV	2.733	3.607	330
Autocracy	Polity IV	4.688	3.743	330
Executive Recruitment	Polity IV	4.915	2.27	330
Executive Constraints	Polity IV	3.176	2.282	330
Political Competition	Polity IV	4.109	3.384	330
Regime Durability	Polity IV	18.138	22.704	347
Negative Regime Change	Polity IV	0.014	0.119	347
Positive Regime Change	Polity IV	0.095	0.294	347
Quality of Government (ICRG)	ICRG/QoG	44.633	18.92	186
Fractionalization (ELF1)	Desmet et al. (2012)	0.184	0.187	348
Fractionalization (ELF15)	Desmet et al. (2012)	0.637	0.307	348
Inequality (SWIID)	Solt (2009)	45.826	11.653	192
Inequality (UTIP)	UTIP/EHII	42.787	7.343	150
Irregular Leader Exit	Goemans et al. (2009)	0.387	0.488	344
War/Conflict (major)	Gleditsch et al. (2002)	0.124	0.33	348
War/Conflict (any)	Gleditsch et al. (2002)	0.231	0.422	347
	Macro I: Prices & Exchange Rates			
Inflation $[\ln(1+\pi)]$	WDI/IFS	22.888	43.975	292
Real X-Rate Value	Rodrik (2008) & PWT 7.0	0.067	0.536	348
Parallel X-Rate $[\ln(1+\pi)]$	Reinhart and Rogoff (2004)	1.281	2.04	153
Black Market Premium $[\ln(1+\pi)]$	Reinhart and Rogoff (2004)	2.719	2.852	95
	Macro II: Trade & Exports			
Current Account (% GDP)	WDI	-3.984	6.695	254
Δ Terms of Trade	WDI & IFS	-0.041	0.177	224
Trade Openness[a]	PWT 7.0 / CEPII	2.321	30.478	348
Manufactures (% Exports)	WDI	22.665	19.982	198
Trade Liberalization	Wacziarg and Welch (2008)	0.232	0.423	306
Export Sophistication (1)	Hausmann et al. (2007)	8.43	0.418	234
Export Sophistication (2)	World Bank / Hausmann et al. (2007)[b]	8.446	0.39	291
Export Diversification (Herfindahl)	WITS/COMTRADE	34.091	24.581	264
	Macro III: Finance			
Capital Account Openness	Chinn and Ito (2006)	-0.491	1.282	304
Financial Integration	Lane and Milesi-Ferretti (2007)	115.305	88.175	309
Financial Depth	Beck et al. (2010)	32.355	18.679	245
Financial Development	Beck et al. (2010)	68.402	22.178	271
Bank Deposits (% GDP)	Beck et al. (2010)	0.295	0.251	248
Private Credit (% GDP)	Beck et al. (2010)	26.249	23.529	248
Domestic Savings (% GDP)	WDI	18	14.561	308
FDI Liabilities (% GDP)	Lane and Milesi-Ferretti (2007)	15.11	15.664	309
External Debt (% GDP)	Lane and Milesi-Ferretti (2007)	65.216	59.176	309
External Leverage[c]	Beck et al. (2010)	1.653	3.271	307
	Other Determinants			
Initial GDP (log)	PWT 7.0	8.198	1.208	348
Real U.S. 3-Month T-Bill Rate	FRED	1.897	2.439	348
Infant Mortality[d]	Word Population Prospects	73.364	40.229	348
Life Expectancy[d]	Word Population Prospects	58.629	10.554	348
Telephones (per 100 ppl)	WDI	5.24	9.778	312
Education (primary)	Barro and Lee (2010)	3.14	1.715	327
Education (secondary)	Barro and Lee (2010)	1.123	0.834	327
Education (all)	Barro and Lee (2010)	4.437	2.471	327

[a] Trade openness is measures as the residual of a regression of a country's size, population, landlockedness, internal distance, GDP and GDP^2 on $(X+M)/GDP$ using the entire PWT 7.0 data.

[b] We extend the HHR series with data from the World Bank's Export Diversification toolkit. The joint series are the predicted values of a (pooled) regression of the World Bank data onto the HHR data.

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

[d] Converted into annual data by interpolation. If the average is for the years 1950-55, we assume it is reached in the 1952 and linearly interpolate to the middle of the next group (1957), and so on.

D Appendix: Duration Method

Several problems arise when modeling durations that OLS has difficulty dealing with. First, duration data is strictly non-negative. Second, durations are conditional, an event can only occur in $\tilde{t}+1$ if it has not occurred in $\tilde{t}$.[63] Third, observations are often censored, implying that the event has not occurred yet but may later on.

Maximum likelihood estimation (MLE) techniques of discrete events usually model the density $f(\tilde{t}) = Pr(\tilde{T} = \tilde{t})$, i.e. the probability of the event occurring at time $\tilde{t}$. However, that probability is only a small component of what defines a duration. A more relevant probability is cumulative probability of an event occurring at different moments in time. For duration analysis in general, it is much more helpful to model the reverse, that is the cumulative probability that the event has not yet occurred. Formally, the survivor function is the reverse of the c.d.f. of $\tilde{T}$, it is defined as $S(\tilde{t}) = 1 - F(\tilde{t}) = 1 - \int_0^{\tilde{t}} f(s)ds \equiv Pr(\tilde{T} > \tilde{t})$ and returns the probability of surviving beyond some time $\tilde{t}$. For most practical purposes we are less interested in aggregate statistics on how long it takes until failure has occurred, but we would like to know the probability of an event occurring in time $\tilde{t}$ *conditional* on it not having occurred before. For this purpose, duration models rely on the hazard function $\lambda(\tilde{t})$, which returns the instantaneous probability of an event occurring given that is has not occurred yet (Lancaster, 1990, p. 7):

$$\lambda(\tilde{t}) = \lim_{d\tilde{t} -> 0} \frac{\Pr(\tilde{t} \leq \tilde{T} < \tilde{t} + d\tilde{t} \mid \tilde{T} \geq \tilde{t})}{d\tilde{t}} = \frac{f(\tilde{t})}{S(\tilde{t})}$$

Defining parameters for either the density, cumulative density, survivor function or hazard function suffices to derive the others and fully describe the duration experience. Hazard models generally begin with the hazard function $\lambda(\tilde{t})$, or cumulative hazard $\Lambda(\tilde{t}) = \int_0^{\tilde{t}} \lambda(s)ds$, and then derive the survival function as $S(\tilde{t}) = \exp(-\Lambda(\tilde{t}))$, which implies the density function $f(\tilde{t}) = \lambda(\tilde{t}) \exp(-\Lambda(\tilde{t}))$.

[63] This refers to exits from a discrete state. In a small departure from usual notation, we use $\tilde{t} = t - t_0$ to denote analysis time for all country-spells, as t is used throughout the paper to denote calender time.

Log-normal Accelerated Failure Time (AFT) models

The two major types of modelling approaches in duration analysis have a different motivation and associated notation. Proportional hazard models specify a model of the hazard function that is scaled up and down depending on the value of the covariates and their parameters, while Accelerated Failure Time models directly specify a model of (log) duration. Sometimes this is just a change in notation, but several models (including the log-normal model) are only fully specified in the AFT metric.

AFT models are defined and interpreted much alike classical linear regression. The log-normal AFT model specifies the process of analysis time as: $\ln(\tilde{t}) = \beta_0 + \mathbf{x}'\boldsymbol{\beta} + \sigma\epsilon$, where $\boldsymbol{\beta}$ has elements $\beta_1, \ldots, \beta_k$ and log-normal refers to the distribution of the log of the error term, so that ϵ itself is distributed $\mathcal{N}(0,1)$ and σ is a scale parameter. The name 'accelerated failure' or 'accelerated life' derives from the interpretation of the model. If $\mathbf{x}'\boldsymbol{\beta} = 0$, then time passes at its normal speed. If $\mathbf{x}'\boldsymbol{\beta} < 0$, time passes more quickly (the event occurs earlier) and if $\mathbf{x}'\boldsymbol{\beta} > 0$, time passes more slowly (the event occurs later).

Specifying the analysis time to be distributed log-normal implies the following relationships in terms of the relevant functions describing the duration. At the baseline, that is $\boldsymbol{\beta} = \mathbf{0}$, the expected survival time is $\mathrm{E}[\ln t|\boldsymbol{\beta} = \mathbf{0}] = \beta_0$. Hence the baseline survival curve and the baseline hazard are:

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

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal p.d.f. and normal c.d.f, respectively.

Introducing (time-invariant) covariates is equivalent to scaling the baseline survival functions. The conditional survival curve is defined as $S(\tilde{t}|\mathbf{x}) = S_0(\tilde{t})\left(\exp(-\mathbf{x}'\boldsymbol{\beta})\tilde{t}\right)$. This implies $S(\tilde{t}|\mathbf{x}) = 1 - \Phi\left((\ln \tilde{t} - (\beta_0 + \mathbf{x}'\boldsymbol{\beta}))\sigma^{-1}\right)$; that is, the intercept can be absorbed into $\boldsymbol{\beta}$. The density and cumulative probability functions are defined implicitly.[64]

Time-varying covariates introduce two complications. First, the hazard rate at each

[64] It follows that an expression for the (log) hazard function conditional on the covariates is $\ln \lambda(\tilde{t}|\mathbf{x}) = \ln \lambda_0\left(\tilde{t}\exp(-\mathbf{x}'\boldsymbol{\beta})\right) - \mathbf{x}'\boldsymbol{\beta}$. An important point is that these hazards are not proportional which is why the log-normal model is only fully specified in the AFT metric.

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

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

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

We can now derive the likelihoods for log-normal duration data required to estimate the parameter vector.[65] Suppose we know the event occurs at $\tilde{t}_i$, the likelihood contribution of an observation i at time $j = \tilde{t}_i$ then is $\mathcal{L}_i = f(j) = S(j)\lambda(j)$. The likelihood contribution of an observation that has not failed at time j, so that $j < \tilde{t}_i$, then is just the probability of survival until j: $\mathcal{L}_i = f(j) = S(j)$. Hence, right-censoring is essentially nothing else than an observation at analysis time j that is still in the sample but has not yet failed and thus extends easily to (exogenous) time-varying covariates.

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

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

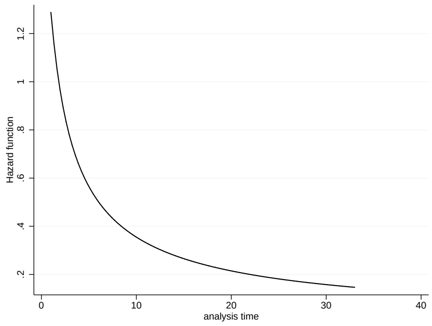
where $\alpha_j(\cdot) = \exp[-\int_{\alpha_j - 1}^{\alpha_j} \lambda(s, \cdot)ds]$ measures survival over the given interval and c_i indicates if observation i is censored. The inner sum (first term) is the probability of survival until $\tilde{t}_i - 1$ and the second term is the conditional probability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

[65] This does not apply to frailty models where the likelihoods are more involved, see Gutierrez (2002).

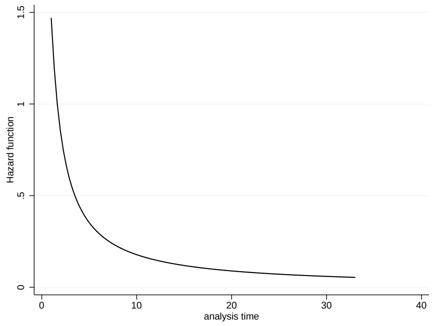
E Appendix: Baseline Hazards

Figure 14 – Robustness: predicted baseline hazards by functional form

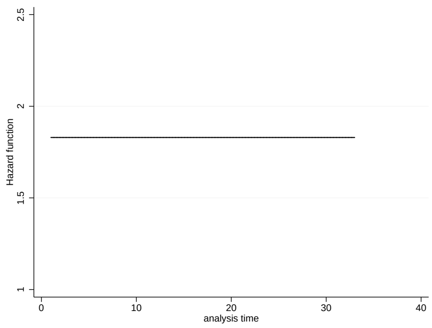
(a) Log-normal



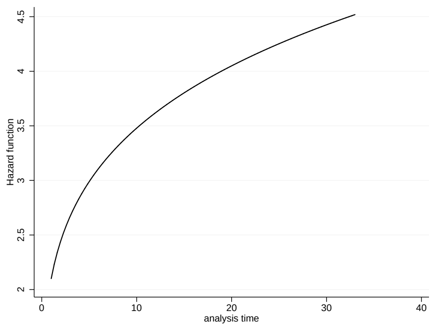
(b) Log-logistic



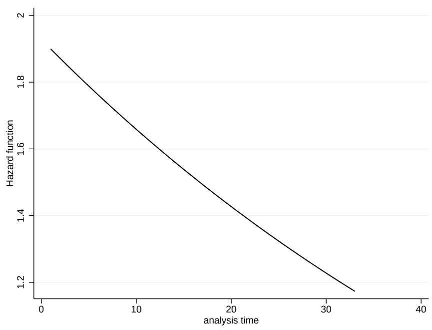
(c) Exponential



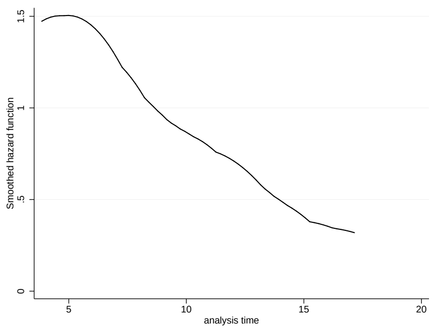
(d) Weibull



(e) Gompertz



(f) Semi-parametric Cox



Notes(s): The log-normal and log-logistic models do not have a proportional hazards interpretation. Although the hazard shape at $\boldsymbol{\beta} = \mathbf{0}$ is monotonically decreasing in both cases, at many other values the hazard function will be increasing and then decreasing, just as noted in the text.