KING'S BUSINESS SCHOOL

Qatar Centre for Global Banking & Finance

The scarring effects of deep contractions*

David Aikman, Mathias Drehmann, Mikael Juselius, Xiaochuan Xing

Working paper No. 2022/1 | October 2022

The scarring effects of deep contractions*

David Aikman[†], Mathias Drehmann[‡], Mikael Juselius[§], Xiaochuan Xing[¶]

October 2022

Abstract

We find that deep contractions have highly persistent scarring effects, depressing the level of GDP at least a decade hence. Drawing on a panel of 24 advanced and emerging economies from 1970 to the present, we show that these effects are nonlinear and asymmetric: there is no such persistence following less severe contractions or large expansions. While scarring after financial crises is well known, it also occurred after the deep contractions of the 1970s and 1980s that followed energy price shocks and restrictive monetary policy to combat high inflation. These results are very robust and have important implications for policy making and macro modelling.

Key words: Hysteresis; nonlinearity; financial crises; monetary policy; oil shocks.

JEL code: E32; E37; G01

*The views expressed are those of the authors and do not necessarily reflect those of the BIS or the Bank of Finland. The authors would like to thank Claudio Borio, Stijn Claessens, Esa Jokivuolle, Egon Zakrajsek and colleagues at the Bank for International Settlements and the Bank of Finland for helpful comments and suggestions.

[†]King's College London.

[‡]Bank for International Settlements.

[§]Bank of Finland.

[¶]Yale University.

1 Introduction

Since the beginning of the new millennium, the global economy has experienced several major contractions that were characterised by non-linearities and scarring effects. The Great Financial Crisis (GFC) is the best know example, when GDP remained below the pre-shock trend path in many countries for a long time.[1] More recently, the deep recessions following the outbreak of the Covid-19 pandemic and the economic fallout from the Russian invasion of Ukraine have raised the fears of long-term scarring.[2] At the same time, we still know little about the determinants of scarring more generally and the circumstances under which it becomes more probable, even though these are critical questions for both policymakers and researchers. For example, whether or not contractions cause long-term scarring has radically different implications for macroeconomic policy and the cost-benefit analysis of mitigating policy actions.[3] And from an economic research perspective, hysteresis effects challenge the still dominant paradigm of linear Dynamic Stochastic General Equilibrium models, which continue to be the workhorse models used by many academics and practitioners.

In this paper, we develop a new statistical test for scarring based on the properties of long-horizon growth rates. The approach is very simple and intuitive. We first define contractions as time periods where the (standardised) annual real GDP growth rate is below the median, and order such events in terms of their severity. We then calculate multi-year real GDP growth rates (up to 10 years) and compare those from the origin of contractions – the quarter immediately preceding the drop in GDP – with those calculated from all other points in the sample. If contractions have only transitory effects, the level of GDP will converge back to trend and we will observe no significant difference in long-term growth rates from the origin of contractions vis-à-vis the rest of the sample. In contrast, if contractions cause scarring, long-term growth rates will remain depressed

[1] See, for instance, Ball (2014).
[2] For a very recent contribution see e.g. Financial Stability Board (2022).
[3] See Cerra et al. (forthcoming) for discussion.

relative to other periods, regardless of the horizon.[4]

Using a panel of 24 advanced and emerging economies from 1970 to the present, we find significant non-linearities in that there is a tipping point in recovery dynamics: contractions whose severity exceeds a certain threshold have highly persistent costs whose effects can be observed in the level of GDP a decade hence. This phenomenon is not present for less severe contractions, the effects of which dissipate with the forecast horizon. Empirically, this tipping point occurs at around the 20th percentile, and we find that the more severe the initial contraction, the larger the scarring effects in the long run. At the same time, we find no evidence of similar persistence following the largest expansions. These findings are economically significant: real GDP growth in the ten years after very severe contractions is almost one standard deviation weaker than otherwise, which is equivalent to a 4.25% drop in the level of GDP for an average advanced economy.

Our second main finding is that the potential for scarring reflects the size of the contraction rather than the reasons it occurred. Several of the largest contractions in our dataset are associated with financial crises, phenomena which are known to have highly persistent effects.[5] But many others were triggered by different factors, including sharp increases in energy prices or restrictive monetary policy actions taken in response to high inflation. While it is hard to disentangle the drivers of any particular recession episode and many factors tend to coincide (e.g., restrictive monetary policy actions and oil prices shocks in the late 70s), we use judgement to allocate recessions to specific types depending on what we view as the dominant narrative of the event in question. Our surprising finding is that when we condition on these recession types separately, we find that all have long-term effects on the level of GDP, with contractions associated

[4]There is a read-across between our method and the old and inconclusive debate on whether GDP contains a unit root. As we discuss in the next section, our approach does not rely on the trend versus difference stationary distinction (see e.g., Christiano and Eichenbaum, 1989 for a summary of this literature). While it does share similarities with other non-parametric approaches, such as Reinhart and Rogoff (2014) and Blanchard et al. (2015), it does not involve judgment about e.g., turning points and definitions of recovery phases.

[5]See e.g., Cerra and Saxena (2008), Reinhart and Rogoff (2009), Claessens et al. (2012), Jorda et al. (2013), and Jordà et al. (2015).

with supply shocks caused by energy market disruptions having the largest long-term effects.

The scarring caused by large contractions is also apparent in the unconditional distribution of multi-period GDP growth. While growth rates are unimodal with a significant left skew at short horizons,[6] we find that distributions of multi-period growth of 5 years and above are multi- and predominantly bimodal. Growth tends to be polarised into two cases: a 'normal growth' state and a 'depressed growth' state. That this phenomenon is visible in the unconditional growth distribution signals a regularity in the data-generating process, with large adverse shocks having similar magnitude both over time and across countries (in terms of standard deviations).

Our findings are robust. They are not driven by the Great Financial Crisis – we obtain similar results using a sample that ends in 2000. They are also not driven by long-term growth trends or cross-country growth differentials. In our baseline specification, we remove a long-run trend using a Hodrick-Prescott filter and normalise each country's multi-period growth rates by subtracting the mean and dividing by the standard deviation (both calculated using the full sample). We obtain similar findings using alternative approaches to detrending including the Hamilton projection filter (Hamilton, 2018) and subtracting 20-year rolling averages; we find similar results too using the raw unfiltered data. Our results are also not sensitive to the method used to define large contractions.

The results also do not reflect another potential concern that large economic contractions tend to follow unsustainable booms, in which case our finding of protracted weak growth could simply reflect a reversion to the (lower) trend path. To address this, we use an approach advocated by Blanchard et al. (2015) where we repeat our baseline analysis, but shift back the starting point used to calculate long-term growth rates by 2 and 3-years from the origin of the contraction. For example, if GDP begins declining in 1990Q1, we calculate 10-year growth rates from 1988Q1 and 1987Q1. Intuitively, this allows us to

[6] For recent evidence, see e.g., Jensen et al. (2020).

'look through' the effects of unsustainable booms that may precede contractions. We find that this has no impact on our findings.

In the last part of the paper, we explore what these results imply for models of the business cycle. Our findings are manifestly inconsistent with textbook Real Business Cycle and New Keynesian models. Even in their pre-linearized form, these models do not generate significant differences between large and small shocks;[7] the dynamics of these models are also typically saddlepath stable, implying the economy will return to steady state following a shock, whatever its magnitude. Our results seem more consistent with strands of the literature that emphasise the potential for contractions to cause long-term scarring via their impact on labour markets, capital accumulation or technical progress (see Cerra et al., forthcoming for a detailed review). Endogenous growth models feature prominently in this literature. But the nonlinearity and asymmetry in our results points to interacting mechanisms, such as non-Gaussian shocks and/or occasionally-binding constraints associated with the effective lower bound on monetary policy or financial constraints. Indeed, our finding that smaller contractions and economic expansions are not associated with long-lasting effects on GDP is inconsistent with the predictions of endogenous growth models, where all shocks will have long-term effects.

Our findings have important implications for policymakers. A key message is that large shocks have the potential to depress the growth trajectory of the economy far into the future. This places a premium on policy actions that reduce the potential for deep contractions in the first place – be it active monetary and fiscal policy reactions to cushion the impact of shocks or resilience-building macroprudential measures that reduce the potential for shocks to be amplified via the financial sector. As such, it underscores Olivier Blanchard's assessment after the GFC that it should be a major objective of policymakers to stay away from "dark corners" (Blanchard, 2014).[8] But our

[7] See Ascari et al. (2015) for an analysis of the properties of pre-linearised Real Business Cycle and New Keynesian models.

[8] Blanchard (2014) defines dark corners are those states of the world when the economy malfunctions badly and non-linearities kick in.

findings also carry a cautionary note: if a large recession does occur, the economy will likely not converge to its pre-recession trend. As such, standard measures of output gaps that guide monetary policy may not be accurate for some time. Moreover, such a deep recession is likely to augur a prolonged period of depressed GDP (relative to its pre-crisis path) with consequences for public and private debt sustainablity, amongst other things.

The rest of this paper is organised as followed. In section 2, we present our statistical method for estimating scarring effects and compare it to other approaches advanced in the literature. Section 3 describes our data and approach to identifying severe contractions. Section 4 presents our main findings, while section 5 examines their robustness to a range of alternative specifications. Section 6 discusses the implications of our results for macroeconomic modelling, and Section 7 concludes.

2 Detecting scarring effects

The most common approach to studying if shocks or economic contractions have long-lasting effects on output exploits statistical differences between trend stationary and difference stationary series.[9] However, neither of these two specifications provides a plausible description of output dynamics.

At one extreme, trend stationarity is hard to reconcile with downward step changes that can be seen from time to time in real output series. At the other extreme, unit-root dynamics impose a degree of symmetry that is also at odds with the data. Under this specification, both big and small contractions have permanent effects leading to more erratic behaviour in GDP than is observed in reality. Moreover, expansions also have permanent effects on output, yet it is very rare to see upward jumps in the level of real GDP. Forcing the data to choose between these two extremes can substantially distort

[9] E.g., Nelson and Plosser (1982), Campbell and Mankiw (1987), and Cochrane (1988).

inference. For instance, other types of non-stationarity, such as structural breaks and or regime-switching dynamics, tend to bias results toward finding unit-roots if not properly accounted for.[10]

Recent contributions to the literature have moved away from explicit functional specifications and have instead approached the problem in other ways. One strand attempts to identify and date peaks and troughs in real GDP and use this information to study characteristics of recovery periods.[11] This approach requires some arbitrary choices that can have large effects on the results. For example, peaks (troughs) are often identified by periods of positive (negative) growth, followed by consecutive periods of negative (positive) growth. But this only picks up extreme business cycle events as growth tends to remain positive during milder downturns. Moreover, an analysis of peaks and troughs is cyclical by construction and so is not well suited to studying the possibility of scarring.[12] An alternative approach attempts to identify shifts in trend output. For instance, Blanchard et al. (2015) compare linear trends estimated before and after large contractions and Ball (2014) compares OECD forecasts of potential GDP before and after the Great Financial Crises in 2007-2008. But both studies assume that output is on a sustainable trajectory over the estimation interval.

In this section, we propose an alternative simple test of scarring, which avoids some of the pitfalls of the aforementioned literature. The basic idea of our approach is illustrated in Figure 1. In the left-hand panel, a big permanent shock to (log) output, y_t, of size $-d$ occurs at time $t_0 + 1$. At any $t \leq t_0 - h$ or $t > t_0$, the h-period ahead growth rate, defined as $x_{t+h,t} \equiv h^{-1}(y_{t+h} - y_t)$, is some constant μ. However, the h-period ahead growth rate calculated at t_0 (blue lines) or the $h-1$ periods preceding it (orange line) is $\mu - d/h$.

[10]Indeed, the literature has failed to reach consensus regarding whether a unit-root exists in real GDP, see e.g., Darné and Charles (2012) and Cushman (2016).

[11]See Cerra and Saxena (2005), Claessens et al. (2012) , Jorda et al. (2013), and Reinhart and Rogoff (2014).

[12]For example, a trough may be followed by one or two quarters of high growth before the economy reverts back to trend growth. The business cycle dating method will only identify the turning-points even if such a short recovery may not make up for initial GDP losses

Clearly, as h increases the h-period ahead growth rate at t_0 converges toward μ at the rate $1/h$, but it turns out that it remains statistically different from μ regardless of h. In contrast, if the shock is transitory (middle panel), the h-period ahead growth rate will return to μ by some horizon $\bar{h}$ (in the middle panel we have set $\bar{h} = 40$).[13]

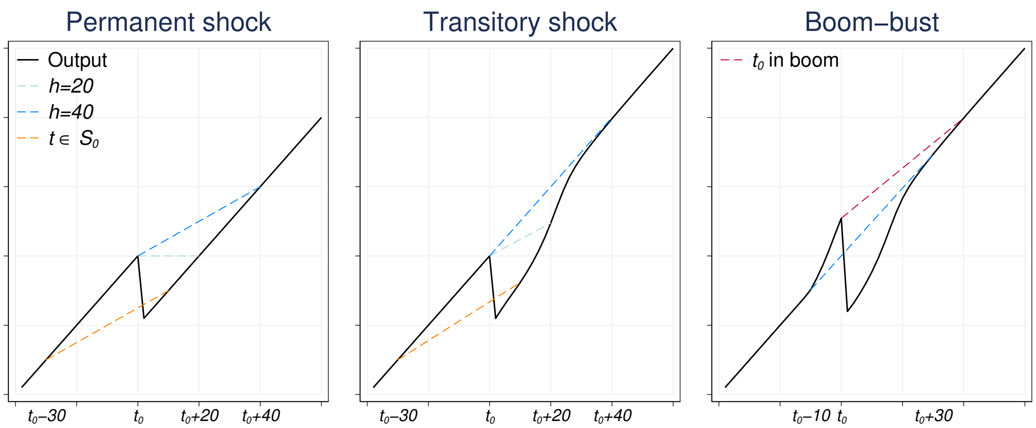

Figure 1: Illustrating the intuition behind our statistical test. The panels illustrate the impact of alternative shocks at t_0 on multi-period growth rates. From the perspective of t_0, a permanent shock (left hand panel) of size $-d$ at $t_0 + 1$ reduces the h-period ahead growth rate by $-d/h$. In contrast, if the shock is transitory, the average growth rate returns to μ for $h \geq \bar{h}$ for some $\bar{h}$ (middle panel). The right hand panel considers the case where there is a boom before the contraction.

We present our method formally in Appendix A, where we also consider the general case where there are many contractions, and discuss other issues such as booms before recessions, permanent effects of expansions, and growth slowdowns.

Implementing our approach is straightforward and amounts to:

1. Defining a contraction severity and identifying the starting points of all such contractions in the sample;

2. Calculating the h-period ahead growth rate from these points; and

3. Testing whether the mean of these growth rates is significantly below the mean of the h-period ahead growth rate at other points of the sample;

[13] It is sufficient that the h-period ahead growth rate at t_0 returns to some close neighborhood of μ by horizon $\bar{h}$, as would for instance be the case with a large transitory AR(1) shock. For ease of exposition, we maintain the simpler assumption of absolute convergence throughout this section.

4. Repeating this process for other severity levels.

We operationalise these steps as follows. We define contractions of different severity by their percentile in the historical distribution of (standardised) annual GDP growth (see Section 3). We start from the most severe contractions, which are those in the 0-5th percentile interval, and then consider consecutive 5-percentile intervals up to the median. This allow us to study contractions of different severity in an agnostic way.[14] We consider a variety of alternative approaches as robustness checks in section 5. We set the horizon h equal to 40, i.e., we look at growth 10-years ahead. While this choice is somewhat arbitrary, it is important for our method that h is sufficiently large and exceeds a reasonable estimate of the horizon at which transitory shocks will have dissipated. Note, however, that when we identify the t_0 of contractions in step 1, we are forced to discard some data to ensure that we can observe 10-year growth rates thereafter.[15] So a larger h entails discarding more data. We consider longer and shorter horizons as robustness tests. Finally, we test the null hypothesis that the difference between the mean of the 10-year ahead growth rates starting at t_0 and the mean of the remaining observations is zero. Since the 10-year growth rates involve overlapping samples by construction and are likely to be both serially and cross-sectionally correlated, we employ the stationary bootstrap resampling scheme originally proposed by Politis and Romano (1994) to construct a confidence region for the difference in the means.[16]

Several aspects of our approach are worth highlighting. First, for $h > 1$ there will be $h - 1$ time periods of subdued growth leading up to t_0 as is clear from the orange line in the left panel of Figure 1. And some of these growth rates (e.g., the one at $t_0 - h + 1$)

[14]This can be contrasted with approaches that rest on narrative identification of specific severe events, e.g., Cerra and Saxena (2008), prior to the analysis.

[15]Given our data end in 2019Q4, the last possible t_0 of a contraction is therefore 2009Q4.

[16]Following this approach, we generate pseudo samples of the same size as the original data, $N \times T$, by first sampling N countries with replacement from the original data using a uniform distribution. For each new country index, we then randomly select T time observations with replacement from the original data by picking blocks of geometrically distributed lengths from uniformly distributed starting points. In case a block runs over the end of the original sample, we continue from the beginning. We pick the geometric survival probability to achieve the optimal average block length in Patton et al. (2009), and use 10000 replications.

will be significantly lower that the rest by construction even if output eventually recovers (orange line in the middle panel). Hence, the $h-1$ observations leading up to t_0 – call this set S_0 – may not be equally informative for whether losses are permanent or not. This highlights the importance of picking the right t_0 points and selecting sufficiently large h.

Second, the discussion so far has assumed that the t_0 observations occur during normal periods, but it may be more realistic to assume that large contractions are preceded by booms. If so, this would bias our results toward finding significant permanent effects even if there are none (right hand panel, red line). We deal with this issue in the spirit of Blanchard et al. (2015) by using the points $\tilde{t}_0 = t_0 - k$ with $k = 8, 12$ in place of our identified t_0 as a robustness exercise. This implicitly assumes that booms last for at most two or three years.

Third, under the null of no permanent losses, the potential growth rate is accurately estimated. The reason is that the lower-than-average growth associated with the S_0 sets are offset by higher-than-average growth as output recovers. However, under the alternative hypothesis, this is no longer the case. For instance, suppose that the potential growth rate is μ but there are occasional permanent reductions in the output level. In this case, the estimated average growth rate is below the true one due to the influence of the S_0 sets, leading to underestimation of the output reductions. Moreover, if the reductions are of approximately equal size, the growth distributions becomes bimodal as h increases (and multi-modal if reductions are not of approximately equal size). One way to overcome this issue, is to throw away the S_0 sets before conducting the tests (this is similar in spirit to the assumption about trends in e.g., Blanchard et al., 2015). But while this works under the alternative, it biases results toward finding permanent losses under the null. For this reason, we only report the results from discarding the S_0 sets as a robustness exercise.

Fourth, our test cannot distinguish between the case where both contractions and expansions have permanent effects and the case where only one of these has such effects. We discuss a modification of our test that allows us to make this distinction in Appendix A. While we implement this as a robustness exercise, it is costly in terms of observations.

3 Data and identifying severe contractions

For our main analysis, we use quarterly time series from 1970Q1 to 2019Q4 – we consider alternative sample periods, including the pre-2000s and the Covid era in section 5. Our sample contains 24 countries for which we have long-run, seasonally-adjusted GDP series, 19 of which are advanced economies and 5 are emerging market economies (EMEs). This gives us a total 4,908 observations.[17]

We look at h-quarter ahead real GDP growth rates, defined as:

$$x_{i,t+h,t} = h^{-1}(y_{i,t+h} - y_{i,t}) \tag{1}$$

with $y_{i,t} = \ln(Y_{i,t})$ where $Y_{i,t}$ is the level of real GDP in country i at time t.[18]

To identify contractions and their magnitude, we take an agnostic approach based on percentiles of the distribution of pooled (standardised) annual real GDP growth rates. We focus in the main on 5 percentile interval ranges up to the median. More formally, let $F(x_{i,t+4,t})$ be the cumulative distribution of $x_{i,t+4,t}$. We classify contractions of severity

[17]The following countries are included: Australia, Austria, Belgium, Brazil (EME), Canada, Denmark, Finland, France, Germany, Italy, Japan, Korea (EME), Mexico (EME), the Netherlands, Norway, New Zealand, Portugal, Sweden, Singapore (EME), South Africa (EME), Spain, Switzerland, the United Kingdom, and the United States. Real seasonally adjusted GDP series are available from 1970Q1 for all countries, except Singapore (1975Q1) and Brazil (1980Q1). Data are taken from national sources.

[18]See Table B.1 (Appendix B) for summary statistics at the various horizons. We also found that the alternative growth rate transformation $X_{i,t+h,t} = h^{-1}(Y_{i,t+h} - Y_{i,t})/Y_{i,t}$ does not produce different results on non-detrended data. This transformation is not applicable for detrended series as such series revolve around zero by construction.

$\bar{p} = 5, 10, 15, ..., 50$ by the following condition

$$x_{i,t+4,t} \in \left(F^{-1}\left(\frac{\bar{p}-5}{100}\right), F^{-1}\left(\frac{\bar{p}}{100}\right) \right]$$

where $\bar{p}$ is the upper percentile of the 5 percentile interval that the contraction belongs to. As an alternative classification, we also consider

$$x_{i,t+4,t} \in \left(F_i^{-1}\left(\frac{\bar{p}-5}{100}\right), F_i^{-1}\left(\frac{\bar{p}}{100}\right) \right]$$

where $F_i(\cdot)$ is the county-specific distribution of $x_{i,t+4,t}$.

We make two adjustments to our GDP series prior to calculating percentiles. First, to address the possibility that the underlying trend growth rate has slowed in some countries in recent decades (e.g., Antolin-Diaz et al., 2017), we remove country-specific long-run trends from log levels of GDP. Our baseline specification is to use a Hodrick Prescott filter with a smoothing factor lambda of 400,000, equivalent to assuming a 20-year long cycle.[19] Such a smoothing factor has also been used in the literature on medium-term credit cycles. As robustness, we also consider: (a) no detrending; (b) a Hamilton projection filter based on country-specific local projections with a lag-length of four and projection horizon of either 10 or 20 years (see Hamilton, 2018); and (c) removing 10 or 20-year rolling averages from $x_{i,t+40,t}$. We focus on these relatively long projection horizons or windows for rolling averages because we assess 10-year growth rates. Second, to avoid average growth differentials between countries driving our results, we normalise all $x_{i,t+h,t}$ by removing the country-specific mean and dividing by the country-specific standard deviation, both calculated using the full sample of data.[20] While this normalisation ensures cross-country comparability, it is not driving our results. As we show in Section 5, results also hold at the country level.

[19]See Ravn and Uhlig (2002) on the relationship between cycle length and lambda.

[20]When we look at sub-samples, means and standard deviations are recalculated for this specific sample.

3.1 Classifying severe contractions

To gain a better intuition about the circumstances surrounding severe contractions, we classify these into one of the following four categories: (1) banking crisis; (2) restrictive monetary policy to combat high inflation; (3) oil shocks; and (4) other. In total, we have 198 extreme contractions (i.e., at or below the 5th percentile): 100 of these are classified as banking crisis-driven contractions; 51 are classified as monetary policy related, the latest of which occurred in the very early 1990s; 19 are the result of the oil shocks of the 1970s; and 9 have other causes. The full allocation is presented in Table B.2 and summarised in Figure B.1 in Appendix B.

We use a combination of simple rules of thumb and judgment to perform this classification. We first ask whether the severe event coincides with either a banking crisis (as defined by the superset of the Laeven and Valencia (2018), Reinhart and Rogoff (2009) and ESRB datasets (see Lo Duca et al. (2016)), a substantial monetary policy tightening (as defined by the increase in nominal interest rates during the period)[21], or an oil price shock (as defined by the 1973-74 and 1978-79 oil price shocks). In the case of multiple potential causes, we examine the published literature and allocate depending on our reading of the dominant driver of the event.

A few comments on this approach. First, we do not require the allocated cause to precede the event in question. The thought experiment is rather whether the severity of the recession cannot be understood without reference to the allocated category. For example, while the proximate cause of the subprime crisis that triggered the US recession of 2008 was arguably the modest tightening in the Fed Funds rate between 2004-2006, the scale of the subsequent recession unquestionably reflected amplification created by the financial crisis.

Second, we allocate some contractions as financial crisis-driven even if they do not

[21] Zarnowitz (1999) and more recently Blinder (2022) also argue that the severe recessions we identify in the US in the late 1970s and early 1980s were related to monetary policy tightening.

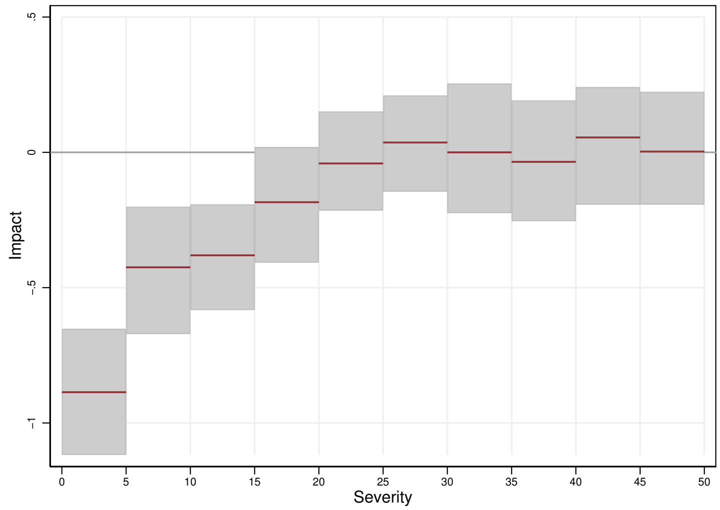

Figure 2: Difference in mean 10-year growth rates following contractions of different severity versus that calculated using all other points in the sample. The solid lines show points estimates of this difference for contractions in the percentile buckets indicated on the x-axis. The shaded areas are 95% confidence intervals. The y-axis is in standard deviations of 10-year real GDP growth.

appear in the Laeven-Valencia, Reinhart-Rogoff or ESRB datasets. Examples here include Italy and Japan's contractions in 2008. Given the exceptional tightening in global financial conditions triggered by the collapse of Lehman Brothers, and the crisis-driven downturn occurring in these countries' main trading partners, the banking crisis was clearly the dominant cause of this downturn even though there was not a domestic banking crisis at this point.

4 Results

Figure 2 summarises the main results of the paper. The solid lines show the estimated differences in mean 10-year growth rates following severe contractions (i.e., annual declines in real GDP in the percentile bucket indicated on the x-axis) compared to mean 10-year growth rates calculated from all other points in the sample. The shaded area shows 95% confidence intervals. To interpret the results, recall that the data are normalised by country-specific means and standard deviations, so the y-axis is expressed in standard deviations of the 10-year growth rate.

Two results stand out. First, the estimates in the left-hand region of the figure are significantly different to zero indicating that 10-year growth rates following severe contractions are detectably weaker than the mean growth rate over the rest of the sample. This effect is largest for the most severe contractions: the difference in 10-year growth rates is 0.9 standard deviations following the 5% largest annual falls in GDP, but a more modest 0.2 standard deviations for a contractions between the 15th and 20th percentiles. These reductions in the 10-year growth rates translate approximately into permanent losses in the level of real GDP of 4.75% and 1.05%, respectively, for a typical economy in our sample.[22] This is below the average loss estimate of 8.4% reported by Ball (2014) for the Great Recession.

Second, following less severe contractions, the economy returns to trend and we observe no long-term impact. This is evident from the fact that mean differences are very close to, and statistically indistinguishable from, zero for contractions above the 20th percentile (with the impact of contractions between the 15th and 20th percentiles having borderline statistical significance). Overall, these results point to the presence of important non-linearities in the data-generating process.

Does the proximate cause of a recession matter for its persistence? To answer this question, we exploit the classification scheme presented in section 3.1, which partitions contractions into those associated with financial crises, monetary policy tightenings to combat high inflation, oil and commodity price shocks, and those driven by other factors. In Figure 3, we repeat the analysis of Figure 2 conditioning on these different recession types. We focus on the most severe contractions, those at or below the 5th percentile. Our novel finding is that, while severe contractions associated with financial crises indeed

[22] At growth horizon h, the difference between the average growth rate at t_0 and other points is given by $-d/h$ which implies $-d = h(-d/h)$. Since we look at annualized numbers, we should use $h = 10$ rather than $h = 40$. Moreover, using the average standard deviation of the 10 year growth rate in Table B.1, we can calculate the level effects as $d = 0.9 \cdot 0.53 \cdot 10 \approx 4.75\%$ and $d = 0.2 \cdot 0.53 \cdot 10 \approx 1.05\%$, respectively.

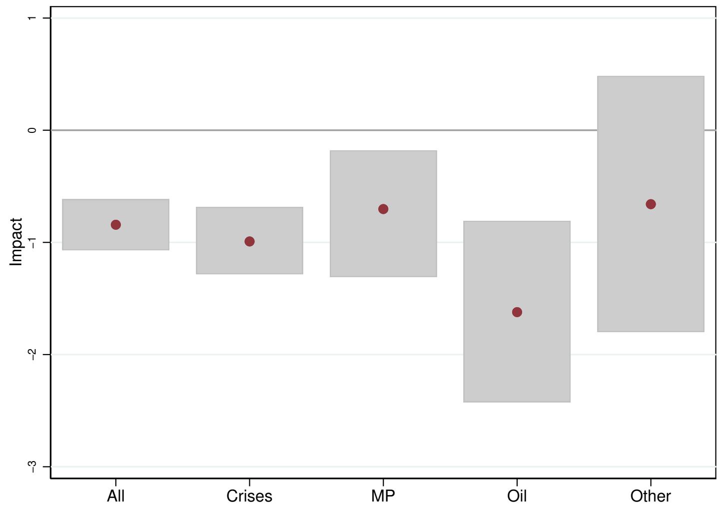

Figure 3: Difference in mean 10-year real growth rates following severe contractions (bottom 5th percentile) of different types versus that calculated using all other points in the sample. The shaded areas are 95% confidence intervals. Recession types are labelled on the x-axis. The y-axis is in standard deviations of 10-year real GDP growth.

have long-term effects, so do those associated with monetary policy tightenings and oil price shocks.[23] Indeed, while the point estimates of the growth shortfall following monetary policy tightenings is somewhat smaller than that following financial crises, oil price shocks generate materially larger growth shortfalls with 10-year growth rates 1.5 standard deviations weaker following such shocks. Interestingly, this resonates with the finding in Blanchard et al. (2015) that recessions that are associated with supply shocks are more likely to be followed by lower level of output compared to pre-recession trend.[24] Overall, these findings challenge the notion that it is only financial crises that generate scarring effects; the perhaps surprising message from Figure 3 is that all severe contractions have this characteristic.

An alternative way to present our findings is via the unconditional distribution of 10-year growth rates – this provides a useful sense-check as it does not depend on a

[23]The after-effects of severe contractions that we could not classify (i.e., those in the 'other' category) are imprecisely estimated due to the small number of observations.

[24]With respect to monetary policy tightenings, our findings are in line with Amador (2022) who studies the transmission of identified monetary policy shocks to total factor productivity and finds evidence of substantial hysterisis.

formal statistical test. As noted in Section 2, when contractions have long-term effects on the level of GDP, we should expect the distribution of multi-period growth rates to be multimodal rather than unimodal.

As Figure 4 shows, while the distribution in annual GDP growth is unimodal with a typical left skew, the distribution in 10-year growth rates is multimodal: one mode is located slightly to the right of zero; the other is located to the left; and there is perhaps even a third mode located further to the right, although this is less clear. Recall that our data have been standardised so zero corresponds to the mean growth rate across the full sample. The presence of a clear second mode to the left of zero indicates that weak 10-year growth outcomes are not uncommon and there is a degree of uniformity in their severity (as measured in country-specific standard deviations).

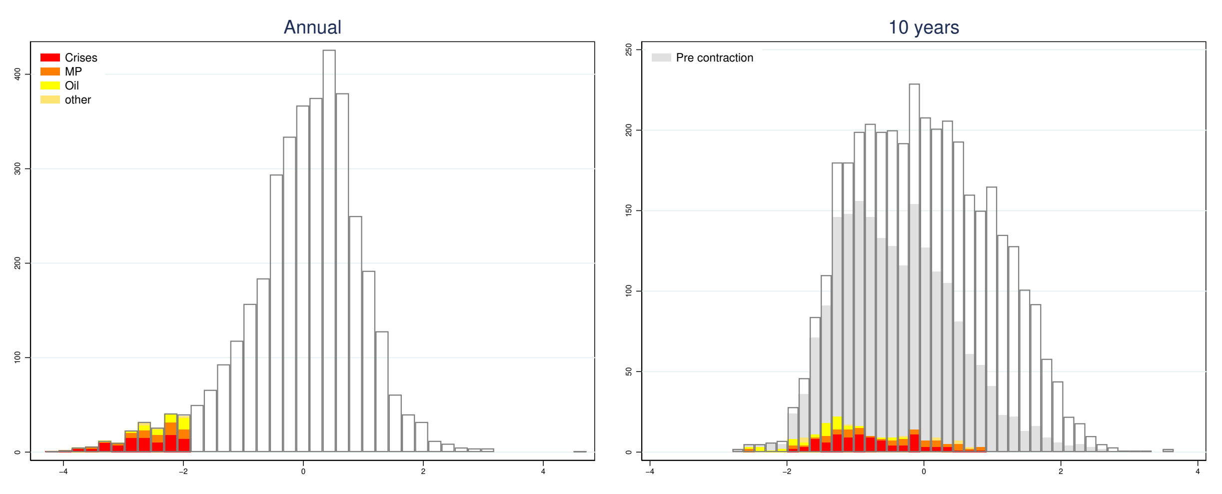

Figure 4: Distribution of annual and 10-year real growth rates (normalised) and different types of large contractions below or at the 5th percentile.

The figure also highlights growth realisations following severe contractions associated with financial crises, monetary policy tightenings, and oil price shocks. These growth outcomes are clustered around the left-hand mode, confirming our finding from Figure 3 that these contractions all have long-lasting impacts.

It is interesting to note that bimodality already emerges at shorter than 10 years (see

Figure B.2. It is already evident for 5 year growth rates and, with good will, at the 3-year horizon. And as would be expected from the 10-year results, at all horizons, the growth outcomes following severe contractions remain clustered around the left-hand mode.

We conclude this section by examining whether large *expansions* (i.e., high annual growth outturns) also have long-term, boosting effects. After all, theories of endogenous growth suggest that capital accumulation in expansions may have long-term effects on the level of GDP because of non-decreasing returns to scale effects. As we discuss in section A.4, some care is needed when exploring this issue using our method. To see why, suppose the true data generating process is asymmetric and that only severe contractions have long-term effects. There will be a gap in this situation between the 'normal' 10-year economic growth rate experienced outside of contractions and the mean calculated over the sample excluding strong expansions because the latter will reflect the drag caused by severe contractions. A naive application of our statistical test, therefore, risks delivering false positives. To address this issue, we run a 'two-sided' version of the test, where we adjust the sample mean by removing from the calculation the weakest (strongest) annual growth outcomes when testing whether strong expansions (severe contractions) have permanent effects.[25]

Figure 5 presents the results of these 2-sided tests alongside the '1-sided' estimates given by a naive application of the test. We focus on annual growth rates in the extreme tails, at the 5th and 95th percentiles. The estimated impact of severe contractions is marginally weaker in the 2-sided test as is to be expected, but remains economically and statistically significant. By contrast, the 2-sided estimate of the long-term impact of strong expansions is small and indistinguishable from zero. This is also strongly supported if we look at the unconditional 10-year growth distribution and plot 10-year growth outturns that follow large annual growth expansions. This is shown in Figure B.3

[25]An additional concern that we do not control for here is the possibility that the strongest growth outcomes in the sample sometimes reflect bounce-back recoveries from severe contractions. In this case, finding long-term effects of such 'expansions' is tantamount to finding that not the full effects of the initial severe contractions are felt permanently.

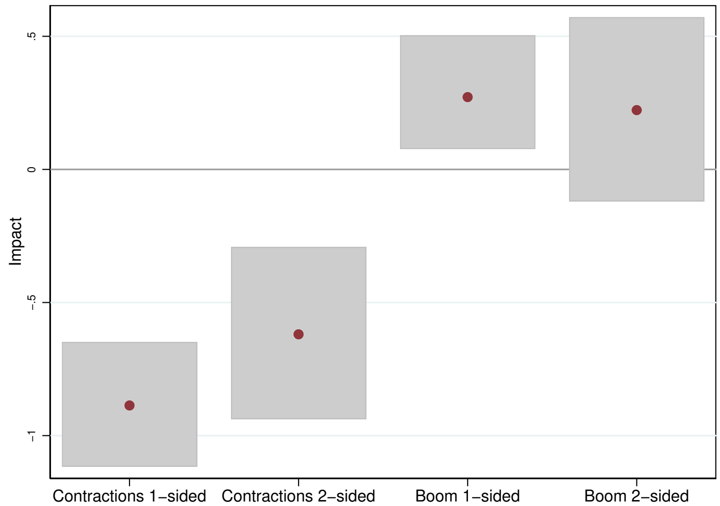

Figure 5: 1-sided and 2-sided estimates of the differences in mean 10-year real growth rates following severe contractions and strong expansions versus rest of the sample (The shaded areas are 95% confidence intervals). Severe contractions are annual growth outcomes at or below the 5th percentile; strong expansions are annual growth outcomes above the 95th percentile. '1-sided' refers to a comparison with the mean 10-year growth rate calculated using all other points in the sample For the '2-sided' test, we also remove strong expansions from the calculation of the mean when testing the long-term impact of severe contractions and vice versa. The y-axis is in standard deviations of 10-year real GDP growth

in Appendix B, highlighting that 10 year growth rates after large expansion are centred around the mean, albeit on average marginally positive in line with the intuition of the one-sided test.

The key takeaway is that the growth process is both non-linear and asymmetric; while large contractions appear to come with the risk of hysteresis and economic scarring, we do not observe long-term positive effects following periods of strong economic growth.

5 Robustness

In this section, we present a battery of robustness tests to explore the impact on our results of using different approaches to detrending the data, applying tests to different sub-samples, and using different methods to identify severe contractions.

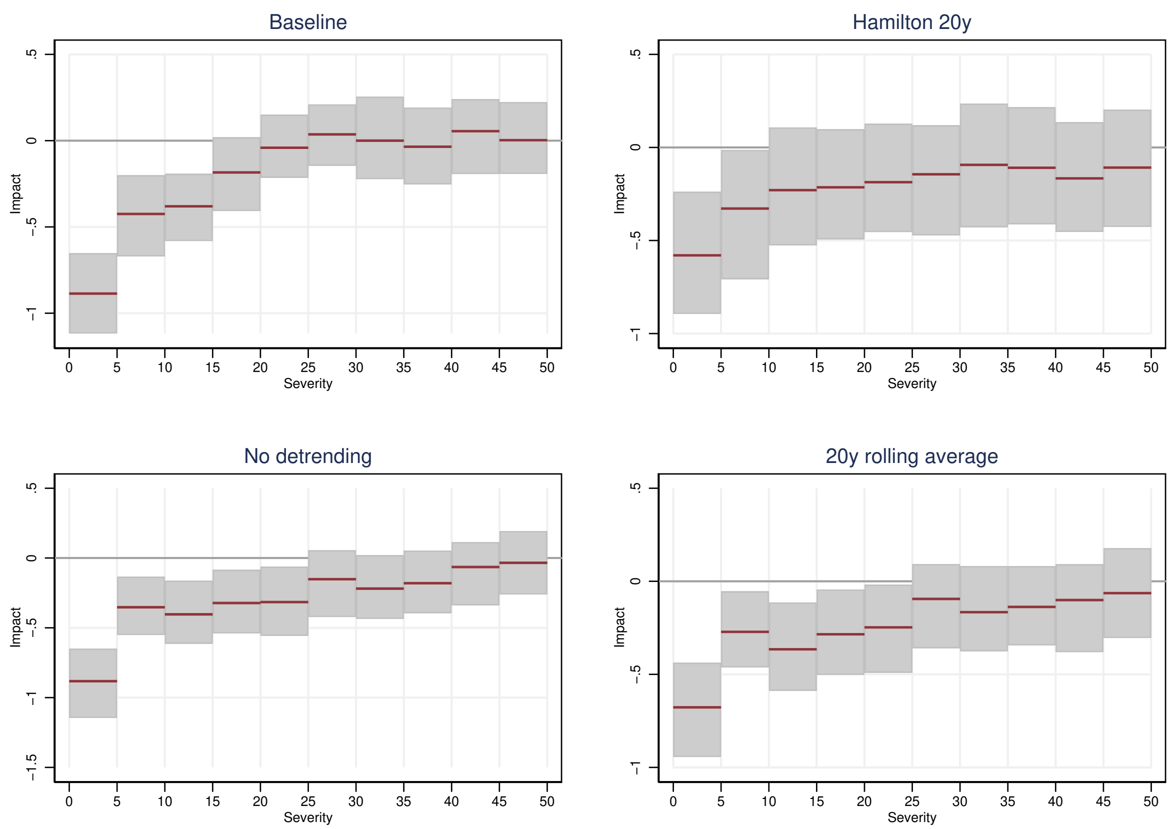

Figure 6: Mean 10-year real growth rate differences under alternative approaches to detrending the data (solid lines). The shaded areas are 95% confidence intervals. Hamilton 20y: detrending by applying the Hamilton projection filter based on country-specific local projections with a lag-length of four and a projection horizon of 20 years. 20y rolling average: removing 20-year rolling averages from the 10-year growth rate, $x_{i,t+10,t}$. The y-axis is in standard deviations of 10-year real GDP growth.

Detrending methods

Figure 6 presents the impact on our baseline results of applying a range of alternative detrending schemes. Recall that our baseline results were obtained using a slow-moving Hodrick-Prescott filter with smoothing factor 400,000. The panels show the impact of (a) not detrending the data at all; (b) applying the Hamilton projection filter based on country-specific local projections with a lag-length of four and a projection horizon of 20 years (see Hamilton, 2018);[26] and (c) removing 20-year rolling averages from the 10-year

[26]These forecast horizons are longer than the suggestions by Hamilton (2018) who proposes 5 years for

growth rate, $x_{i,t10,t}$.

Our baseline result that severe contractions have permanent effects goes through in all cases. Moreover, in all cases we continue to find that the estimated long-term impact diminishes as we consider the impact of less severe contractions. The main differences across these detrending schemes relates to the degree of imprecision in the estimates, which in some cases changes the specific threshold beyond which permanent effects disappear. In all alternative specifications, the estimated scarring effects of the most severe contractions continue to be statistically different to zero (the exception being the 10-year rolling average case). And there is no specification in which scarring effects are statistically detectable beyond the 25th percentile.

Sample splits

Our results are also robust to using different sample splits (Figure 7). We find no real difference between estimated effects for advanced economies or emerging markets; both display long-term scarring after severe contractions. This challenges the result reported in Aguiar and Gopinath (2007) that shocks to trend growth are the primary source of economic fluctuations in emerging market economies whereas advanced economies are better characterised as displaying transitory fluctuations around a stable trend. Similarly, we find no material impact of splitting our sample pre- and post 2000, although only very severe contractions have permanent effects in our post-2000 sample. Our baseline results are also unchanged if we restrict the focus to countries that have not experienced a financial crisis in our sample, corroborating our earlier finding that it is not only contractions associated with financial crises that have long-term effects.[27] Finally, our baseline results are also unaffected by including the Covid era within the sample.

the credit cycle. This is too short given we look at 10 year growth rates. For instance, in the stylised picture of a permanent shock in Figure 6, the Hamilton gap with a 5 year horizon would be zero 5 years after t_0 despite the permanent drop in output. As an additional robustness check, we also used a Hamilton gap with a 10 year horizon. Eve though this strongly biases us to find scarring, significant scarring effects are found for very severe contractions below the 5th percentile

[27]Countries without a financial crises are Australia, Canada, New Zealand, Singapore and South Africa.

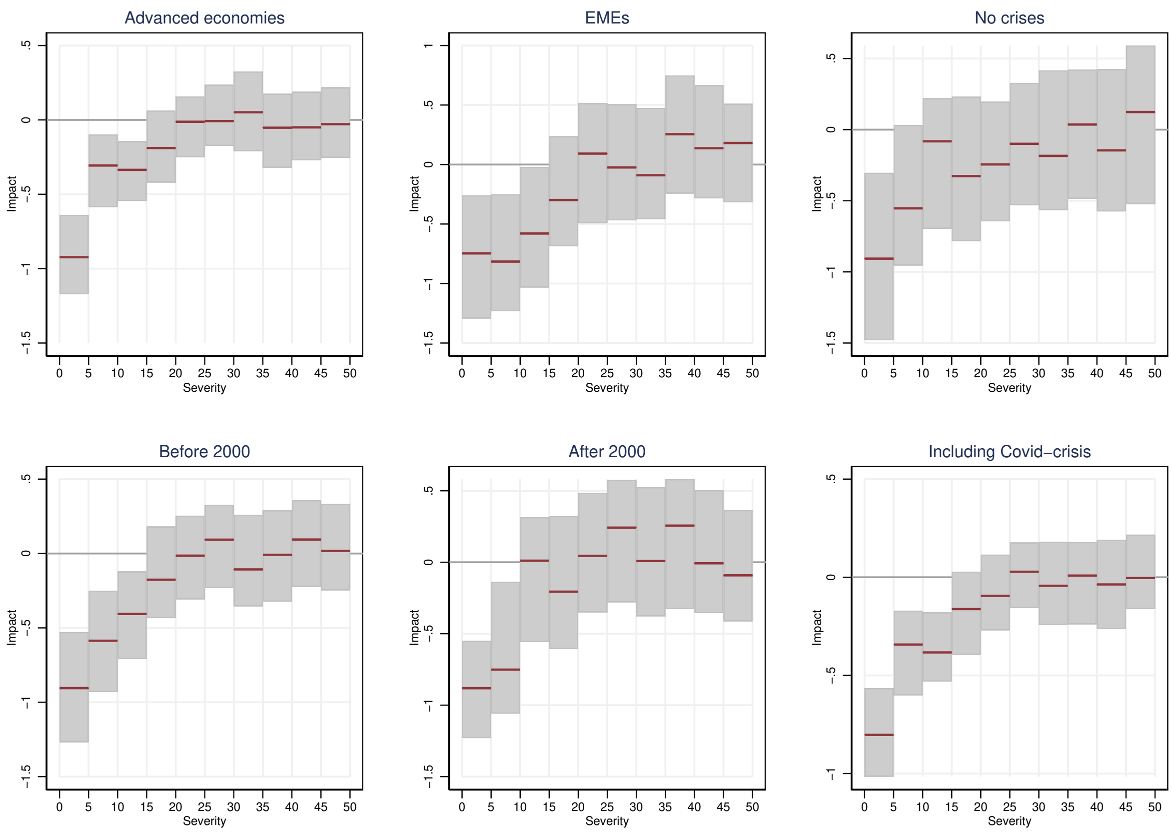

Figure 7: Mean 10-year real growth rate differences under alternative sub-samples of the data (solid lines). The shaded areas are 95% confidence intervals. No crises: only countries that did not experience a financial crises Including Covid-crisis: Sample extended up to 2021Q3. The y-axis is in standard deviations of 10-year real GDP growth

The horizon

The choice of the 10-year horizon also not drive the results (Figure 8). In this case we focus on the very severe contractions below the 5th percentile. By construction, at the one year horizon, the contractions are extremely different to normal times. Differences decrease as the horizon increases and flatten out after year 6. And they remain large and significant even after 15 years. Finally, our baseline results are also unaffected by including the Covid era within the sample.

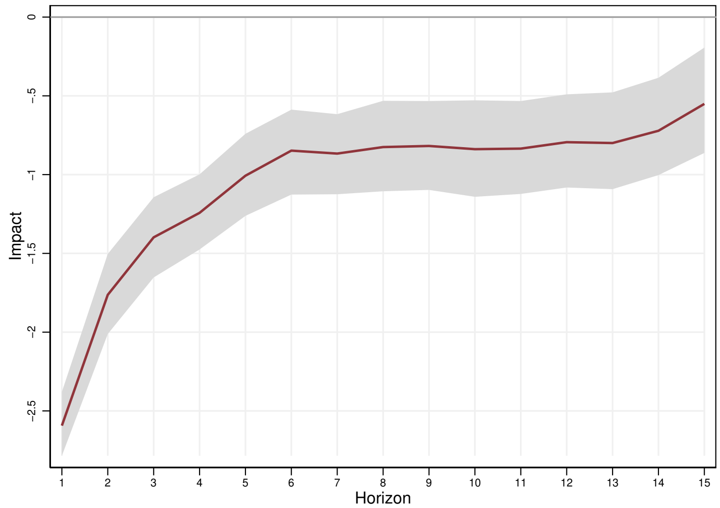

Figure 8: Mean 10-year real growth rate differences of very severe contractions (at or below the 5th percentile) for different horizons (in years). The shaded areas are 95% confidence intervals. The y-axis is in standard deviations of GDP growth at the specific horizon.

Methodology for defining severe contractions

We conclude the set of robustness exercises by exploring the impact of alternative definitions of severe contractions. Overall, our main finding that large contractions have scarring effects whereas smaller contractions do not continues to hold (Figure 9). In our baseline, we categorised contractions by their percentile of the (standardised) annual growth distribution in our pooled sample. We then compared 10-year growth rates from T_0 with 10-year growth rates for all periods in the sample. Here we examine the following alternatives: (a) defining contractions according to their percentiles of country-specific annual growth distributions rather than the pooled data; (b) using quarterly real GDP growth rather than annual GDP growth to classify contractions; (c) defining contractions based on the distribution of 'shocks' to growth rather than actual growth rates, where shocks are the residuals of country-specific AR(4) regressions; (d) comparing 10-year growth rates at T_0 with 10-year growth rates for all periods except those 10-year growth rates that contain T_0 (i.e., we drop S_0)[28]; and (e) controlling for booms up to

[28]Dropping S_0 reduces the sample size significantly as we drop 9 years and 3 quarters prior to each

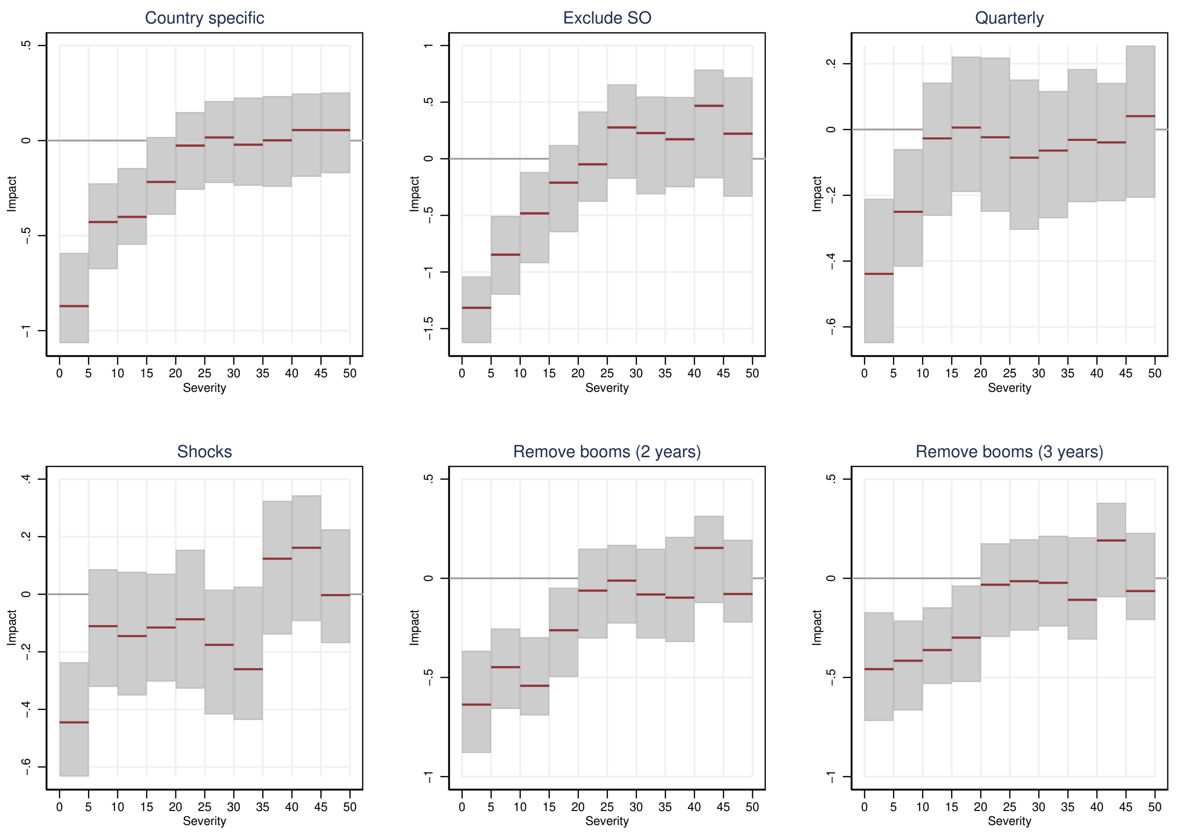

Figure 9: Mean 10-year real growth rate differences under alternative approaches to defining severe contractions (solid lines). The shaded areas are 95% confidence intervals. Country specific: defining contractions according to their percentiles of country-specific annual growth distributions; Quarterly: using quarterly real GDP growth to classify contractions; Shocks: defining contractions based on the distribution of 'shocks' identified as residuals of country-specific AR(4) regressions; Exclude S0: comparing 10-year growth rates at T_0 with 10-year growth rates for all periods except those 10-year growth rates that contain T_0; Remove booms (2/3 years): Using the points $\tilde{t}_0 = t_0 - k$ with $k = \tilde{8}, 12$ quarters in place of the identified t_0. The y-axis is in standard deviations of GDP growth at the 10-year horizon.

three years before contractions. As it is evident from Figure 9, none of these changes has a material impact on the key take-aways of the paper.

Given these findings, it is not surprising that our results also hold if we run the analysis at the country level (Figure B.4 in the Appendix). The most severe contractions have significant long-run effects for the vast majority of countries, albeit our estimates are

event. In turns out that for smaller sized contractions, this leads to so few observations that no tests can be run.

now less precise and these point estimates are statistically distinct from zero at the 5% level only for 11 of the 24 countries.[29]

6 Connection to macroeconomic models

This section discusses how our empirical results link to, and what they imply for, various strands of the theoretical literature on business cycles. [30]

It is clear at the outset that simple autoregressive models with Gaussian shocks are not consistent with our findings of non-linearity and asymmetry in the data generating process of GDP. Such models imply either trend stationarity or, if the roots of the model imply non-stationarity, that some shocks will have permanent effects regardless of their sign or size. This implies that the two benchmark models used in the macroeconomics literature – the Real Business Cycle (RBC) model and its close cousin the New Keynesian model – are also inconsistent with our findings.

The potential for temporary shocks to have highly persistent or even permanent effects on the economy arises naturally in endogenous growth models. But in contrast to our findings, these effects are typically symmetric in such models. In models of endogenous growth, temporary disturbances that change the amount of resources allocated to growth leave a permanent imprint on the level of output.[31] For example, a recession may make innovation less profitable resulting in reduced R&D spend for a period, permanently lowering the level of technology. This helps explain our finding that even recessions associated with monetary policy tightening – the archetypal temporary demand shock in New Keynesian models – can have permanent effects on the level of output. By the same logic, however, booms should also have this effect.[32] So while endogenous growth mod-

[29]These are: Austria, Australia, Belgium, Brazil, Switzerland, Japan, Korea, Mexico, Norway, New Zealand and South Africa.

[30]For an excellent recent comprehensive survey of models discussed in this section and the role they can play in generating hysteresis, see Cerra et al. (forthcoming).

[31]See King et al. (1988) and Stadler (1990), and the discussion in Fatas (2000).

[32]Indeed, the stories embodied in these models via which the growth process occurs – e.g., "changes in

els help us understand how some temporary shocks plausibly have permanent effects on the supply capacity of the economy, they do not explain why only large contractions have this effect.

Business cycle models with occasionally binding constraints and financial frictions generate asymmetric and nonlinear responses to shocks, but not permanent scarring. For example, in Brunnermeier and Sannikov (2014) credit-constrained entrepreneurs hold precautionary buffers to navigate moderately sized shocks so that the economy exhibits small fluctuations around the steady state in normal times. If shocks are sufficiently large, however, or if this buffer is eroded, entrepreneurs are forced to sell assets, depressing asset prices and kick-starting an adverse feedback loop because of tightening credit constraints. The stationary distribution of the model is bimodal: around the steady state, volatility is low and growth is strong. But once constraints bind, the economy quickly transitions to a depressed growth regime at which it remains for some time. [33]

A recent strand of the literature embeds occasionally binding financial constraints in models with endogenous growth. One example of this approach is Queralto (2020), who finds that financial frictions significantly depress medium-run productivity and output losses following a crisis.[34] Ikeda and Kurozumi (2019) develop a model in which adverse financial shocks can induce a slow recovery and examine the implications for optimal monetary policy. A recent paper by Bonciani et al. (2020) considered the implications for macroprudential policy, finding that optimal bank capital ratios are significantly higher

the utilization of factor inputs when demand changes can result in reorganization and the acquisition of new skills" (Stadler, 1990) – are arguably better suited to describing expansions than contractions.

[33]Other models feature similar dynamics. He and Krishnamurthy (2013) use a model with occasionally binding constraints to study the transition from normal states to rare systemic risk states. Adrian and Boyarchenko (2019) examine a similar economy, but relax the assumption that all household savings must be intermediated by the banking sector. Holden et al. (2020) present a model in which banks face occasionally binding borrowing constraints and costs to issuing fresh equity. Xing (2022) presents a related model in which leverage (the credit multiplier) is strongly procyclical, implying even tighter credit constraints during recessions that can further amplify the non-linear downward spiral.

[34]Borio et al. (2016) also provide evidence of how resource misallocations during credit booms persistently depresses productivity during the credit bust.

once we account for the slow recoveries from crises predicted by endogenous growth models.

Finally, a distinct nonlinear mechanism consistent with our empirical findings arises in models of so-called 'debt traps', where an economy can transition from a good equilibrium to a low growth state. A recent example of such a model is proposed by Mian et al. (2021), where savers have lower marginal propensities to consume than borrowers. As indebtedness increases, higher debt service costs reduce aggregate demand, and with it the equilibrium real interest rate. If the decline in the equilibrium rate is of sufficient magnitude and there are nominal rigidities in the economy, the effective lower bound may impair the capacity of monetary policy to stabilize the economy and the economy can enter a low-growth debt-trap in which borrowers' consumption is depressed by the weight of debt servicing costs and savers' consumption is insufficient to make up the difference.[35]

7 Conclusion

In this paper, we present new evidence demonstrating that the size of economic contractions rather than their cause is the better indicator of the potential for scarring/hysteresis effects.

Using a panel of 24 advanced and emerging economies from 1970 to the present, we find significant nonlinearities in the long-run effects of contractions, with contractions in the bottom 20% of the pooled annual growth distribution having effects that can be observed in the level of GDP a decade hence. We find no such evidence following less severe contractions or following large economic expansions. Our estimates are economically significant: real GDP growth in the 10-years after very severe contractions is almost

[35] A purely real model of such a debt trap is presented by Matsuyama (2007): when adverse shocks deplete entrepreneurs' net worth, they become heavily reliant on external borrowing and invest in low productivity projects with high pledgeable returns.

one standard deviation weaker than otherwise.

Perhaps surprisingly, our results do not simply reflect the impact of financial crises, whose after-effects are now well known to be highly persistent. Deep contractions associated with monetary policy tightenings to combat high inflation and with supply shocks caused by energy market disruptions also have material long-term effects, with the latter having the largest long-term costs overall. Our findings are robust to a battery of robustness tests, including alternative detrending approaches, sample splits and methods for defining severe contractions.

Our results have important implications for policymakers and business cycle modelling. A key message for policymakers is that large shocks tend to depress the growth trajectory of the economy far into the future, suggesting greater benefits from macroeconomic policies designed to mitigate the impact of large shocks and from monetary or fiscal policy responses designed to prevent small contractions becoming severe ones. For business cycle modelling, our results challenge the self-stabilising properties of workhorse Dynamic General Equilibrium Models, and instead support models that combine endogenous productivity growth with nonlinear constraints.

References

Adrian, T. and N. Boyarchenko (2019) "Liquidity policies and systemic risk," *American Economic Review*, 109 (4), 1263–89. 26

Aguiar, M. and G. Gopinath (2007) "Emerging Market Business Cycles: The Cycle Is the Trend," *Journal of Political Economy*, 115 (1), 69–102. 21

Amador, Sebastián (2022) "Hysteresis, endogenous growth, and monetary policy," University of California, Davis, Department of Economics, Working Papers 348. 16

Antolin-Diaz, J., T. Drechsel, and I. Petrella (2017) "Tracking the Slowdown in Long-Run GDP Growth," *Review of Economics and Statistics*, 99 (2), 343–356. 12, 37

Ascari, G., G. Fagiolo, and A. Roventini (2015) "Fat-tail distributions and business-cycle models," *Macroeconomic Dynamics*, 19 (2), 465–476. 5

Ball, Lawrence (2014) "Long-term damage from the Great Recession in OECD countries," NBER Working Paper, No 20185. 2, 7, 15

Blanchard, Olivier, Eugenio Cerutti, and Lawrence Summers (2015) "Inflation and Activity – Two Explorations and their Monetary Policy Implications," IMF Working Papers No 230. 3, 4, 7, 10, 16, 36

Blanchard, Olivier J (2014) "Where Danger Lurks: The recent financial crisis has taught us to pay attention to dark corners, where the economy can malfunction badly," *Finance & Development*, 51 (003). 5

Blinder, A. (2022) "Landings hard and soft: the Fed, 1965-2020," markus' academy, https://bcf.princeton.edu/events/alan-blinder-on-landings-hard-and-soft-the-fed-1965-2020/. 13

Bonciani, D., D. Gauthier, and D. Kanngiesser (2020) "Slow Recoveries, Endogenous Growth and Macro-prudential Policy," Available at SSRN. 26

Borio, Claudio, Enisse Kharroubi, Christian Upper, and Fabrizio Zampolli (2016) "Labour reallocation and productivity dynamics: financial causes, real consequences," BIS Working Papers No 534. 26

Brunnermeier, M. and Y. Sannikov (2014) "A macroeconomic model with a financial sector," *American Economic Review*, 104 (2), 379–421. 26

Campbell, John Y. and Gregory N. Mankiw (1987) "Are Output Fluctuations Transitory?" *Quarterly Journal of Economics*, 102 (4), 857–880. 6

Cerra, V., A. Fatas, and S. Saxena (forthcoming) "Hysteresis and Business Cycles," *Journal of Economic Literature*. 2, 5, 25

Cerra, V. and S. Saxena (2008) "Growth dynamics: The myth of economic recovery," *American Economic Review*, 98 (1), 439–457. 3, 9

Cerra, Valerie and Sweta C. Saxena (2005) "Growth Dynamics: The Myth of Economic Recovery," IMF Working Paper No 147. 7

Christiano, Lawrence J and Martin Eichenbaum (1989) "Temporal aggregation and the stock adjustment model of inventories," in *The Rational Expectations Equilibrium Inventory Model*, 70–108: Springer. 3

Claessens, Stijn, M. Ayhan Kose, and Marco E. Terrones (2012) "How do business and financial cycles interact?" *Journal of International Economics*, 87 (1), 178–190. 3, 7

Cochrane, John H. (1988) "How Big Is the Random Walk in GNP?" *Journal of Political Economy*, 96 (5), 893–920. 6

Cushman, David (2016) "A unit root in postwar U.S. real GDP still cannot be rejected, and yes, it matters," *Econ Journal Watch*, 13 (1), 5–45. 7

Darné, Olivier and Amélie Charles (2012) "A note on the uncertain trend in US real GNP: Evidence from robust unit root tests," *Economics Bulletin*, 32 (3), 2399–2406. 7

Fatas, Antonio (2000) "Do business cycles cast long shadows? Short-run persistence and economic growth," *Journal of Economic Growth*, 5 (2), 147–162. 25

Financial Stability Board (2022) "Exit strategies to support equitable recovery and address effects from COVID-19 scarring in the financial sector. Interim report," Interim report, 13 July 2022. 2

Hamilton, James D. (2018) "Why You Should Never Use the Hodrick-Prescott Filter," *Review of Economics and Statistics*, 100 (5), 831–843. 4, 12, 20

He, Z. and A. Krishnamurthy (2013) "Intermediary Asset Pricing," *American Economic Review*, 103 (2), 732–770. 26

Holden, T., P. Levine, and J. Swarbrick (2020) "Credit Crunches from Occasionally Binding Bank Borrowing Constraints," *Journal of Money, Credit and Banking*, 52 (2–3), 549–582. 26

Ikeda, D. and T. Kurozumi (2019) "Slow Post-financial Crisis Recovery and Monetary Polic," *American Economic Journal: Macroeconomics*, 11 (4), 82–112. 26

Jensen, H., I. Petrella, S. Hove Ravn, and E. Santoro (2020) "Leverage and Deepening Business-Cycle Skewness," *American Economic Journal: Macroeconomics*, 12, 245–281. 4

Jorda, O., M. Schularick, and A. Taylor (2013) "When credit bites back," *Journal of Money, Credit and Banking*, 45 (2), 3–28. 3, 7

Jordà, Òscar, Moritz Schularick, and Alan M Taylor (2015) "Leveraged bubbles," *Journal of Monetary Economics*, 76, S1–S20. 3

King, R., C.I. Plosser, and S.T. Rebelo (1988) "Production, Growth and Business Cycles: II. New Directions," *Journal of Monetary Economics*, 21 (2-3), 309–341. 25

Laeven, Luc and Fabian Valencia (2018) "Systemic Banking Crises Revisited," *IMF Working Paper* (WP/18/206). 13

Lo Duca, Marco, Anne Koban, Marisa Basten et al. (2016) "A New Database for Financial Crises in European Countries, Ecb/Esrb Eu Crises Database," *European Systemic Risk Board, Occasional Paper Series* (13). 13

Matsuyama, K. (2007) "Credit Traps and Credit Cycles," *American Economic Review*, 97 (1), 503–516. 27

Mian, Atif, Ludwig Straub, and Amir Sufi (2021) "Indebted Demand," *Quarterly Journal of Economics*, 136 (4), 2243–2307. 27

Nelson, Charles R. and Charles I. Plosser (1982) "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10 (2), 139–162. 6

Patton, Andrew, Dimitris N. Politis, and Halbert White (2009) "Correction to "Automatic Block-Length Selection for the Dependent Bootstrap" by D. Politis and H. White," *Econometric Reviews*, 28 (4), 372–375. 9

Politis, Dimitris N. and Joseph P. Romano (1994) "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89 (428), 1303–1313. 9

Queralto, A. (2020) "A model of slow recoveries from financial crises," *Journal of Monetary Economics*, 114 (C), 1–25. 26

Ravn, Morten O. and Harald Uhlig (2002) "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations," *Review of Economics and Statistics*, 84 (2), 371–376. 12

Reinhart, C. and K. Rogoff (2009) "The aftermath of financial crises," *American Economic Review*, 99 (2), 466–472. 3, 13

Reinhart, Carmen M. and Kenneth S. Rogoff (2014) "Recovery from Financial Crises: Evidence from 100 Episodes," *American Economic Review*, 104 (5), 50–55. 3, 7

Stadler, G.W. (1990) "Business Cycle Models with Endogenous Technology," *The American Economic Review*, 80 (4), 763–778. 25, 26

Xing, Xiaochuan (2022) "Collateral, leverage, and the business cycle," Available at SSRN 4119075. 26

Zarnowitz, Victor (1999) "Theory and history behind business cycles: are the 1990s the onset of a golden age?" *Journal of Economic Perspectives*, 13 (2), 69–90. 13

A Methodological aspects

In this appendix, we motivate our approach more formally and discuss some issues that arise.

A.1 Baseline model

To formalize this idea, let's assume for simplicity (and without loss of generality)[36] that y_{t+1} is a trend stationary process given e.g. by

$$y_{t+1} = \beta y_t + (1 - \beta L)\left(\alpha + \mu(t+1)\right) + \varepsilon_{t+1} \tag{A.1}$$

where $|\beta| < 1$, L is the lag operator and $\varepsilon_{t+1} \sim N(0, \sigma^2)$. The solution of (A.1) is

$$y_{t+1} = \alpha + \mu(t+1) + \sum_{i=0}^{\infty} \beta^i \varepsilon_{t+1-i} \tag{A.2}$$

implying that the h-period ahead growth rate follows

$$x_{t+h,t} = h^{-1}\left(\mu h + \sum_{i=0}^{\infty} \beta^i \left(\varepsilon_{t+h-i} - \varepsilon_{t-i}\right)\right) \tag{A.3}$$

with mean and variance

$$\mu_h \equiv E[x_{t+h,t}] = \mu \tag{A.4}$$

$$\sigma_h^2 \equiv E[x_{t+h,t} - \mu_h]^2 = h^{-2}\left(\frac{2\sigma^2(1-\beta^h)}{1-\beta^2}\right). \tag{A.5}$$

[36]For most stationary processes, the variance of the h-period ahead growth rate will have the form $h^{-2}\phi(h)$ where $\phi(h)$ converges to some finite number as h increases. This implies that the statements below will go through with minor modifications.

A.2 A one-off permanent shock

A one-off permanent reduction of size $-d$ in y_t can be modeled by adding $-(1-\beta L)d$ to the right hand side of (A.1) at time t_0+1. We assume that the permanent component is large with respect to the one-period ahead growth rate, $x_{t+1,t} = y_{t+1} - y_t$, i.e. we set $d \geq \lambda\sigma_1$, where $\sigma_1^2 = 2\sigma^2(1+\beta)^{-1}$ is the variance of $x_{t+1,t}$ from (A.5) and λ is some positive number (for instance 1.645 for a 5% one-sided significance level).

If this permanent reduction occurs, we have

$$y_{t+1} = \begin{cases} \alpha + \mu(t+1) + \sum_{i=0}^{\infty} \beta^i \varepsilon_{t+1-i} & \text{for } t \leq t_0 \\ \alpha - d + \mu(t+1) + \sum_{i=0}^{\infty} \beta^i \varepsilon_{t+1-i} & \text{for } t > t_0 \end{cases} \tag{A.6}$$

and therefore

$$x_{t+h,t} = \begin{cases} h^{-1}\left(\mu h + \sum_{i=0}^{\infty} \beta^i \left(\varepsilon_{t+h-i} - \varepsilon_{t-i}\right)\right) & \forall t \notin S_0 \\ h^{-1}\left(\mu h - d + \sum_{i=0}^{\infty} \beta^i \left(\varepsilon_{t+h-i} - \varepsilon_{t-i}\right)\right) & \forall t \in S_0 \end{cases}$$

where $S_0 = [t_0 - h + 1, t_0]$, which implies

$$E[x_{t+h,t}] = \mu_h = \begin{cases} \mu & \forall t \notin S_0 \\ \mu - d/h & \forall t \in S_0. \end{cases}$$

Specifically, the expected h-period ahead growth rate at t_0 is $\mu_{t_0,h} \equiv E[x_{t_0+h,t_0}] = \mu - d/h$. The variance is given by (A.5) in both cases.

Consider the sample estimator for μ_h outside the set S_0 given by $\bar{x}_{h,t\notin S_0} = (T - 2h)^{-1} \sum_{t=1,t\notin S_0}^{T-h} x_{t+h,t}$, where we have assumed $t_0 \geq h$ for simplicity so that the available number of observations for calculating $\bar{x}_{h,t\notin S_0}$ is $T-2h$. The central limit theorem implies that

$$\bar{x}_{h,t\notin S_0} \sim N(\mu, (T-2h)^{-1}\sigma_h^2). \tag{A.7}$$

for large T, where $N(.,.)$ denotes the normal distribution.

We want to show that $\bar{x}_{h,t\notin S_0}$ converges sufficiently fast in distribution around μ so that $\mu_{t_0,h}$ remains in the tail for all h, i.e. that $\mu - \mu_{t_0,h} \geq \lambda\sqrt{(T-2h)^{-1}\sigma_h^2} = \lambda \text{std}[\bar{x}_{h,t\notin S_0}]$. To do so, we first show that $\sigma_h^2 \leq h^{-1}\sigma_1^2$ holds for all $h \geq 1$. This follows from the fact that

$$\begin{aligned}
\Delta(h^2\sigma_h^2) &= h^2\sigma_h^2 - (h-1)^2\sigma_{h-1}^2 \\
&= \frac{2\sigma^2(1-\beta^h) - 2\sigma^2(1-\beta^{h-1})}{1-\beta^2} \\
&= \frac{2\sigma^2\beta^{h-1}(1-\beta)}{1-\beta^2} \\
&= \beta^{h-1}\sigma_1^2 \\
&< \sigma_1^2
\end{aligned}$$

which implies $h^2\sigma_h^2 = \sigma_1^2 + \sum_{j=2}^{h}\Delta(j^2\sigma_j^2) < h\sigma_1^2$ and therefore the inequality. Hence, as long as $T-2h \geq h$, we have $((T-2h))^{-1}\sigma_h^2 \leq ((T-2h))^{-1}(h^{-1}\sigma_1^2) \leq h^{-2}\sigma_1^2$, which implies that

$$\mu_h - \mu_{t_0,h} = dh^{-1} \geq \lambda\sigma_1 h^{-1} = \lambda\sqrt{h^{-2}\sigma_1^2} \geq \lambda\sqrt{(T-2h)^{-1}\sigma_h^2}. \tag{A.8}$$

In other words, x_{t_0+h,t_0} is expected to be an outlier vis-a-vis the distribution of $\bar{x}_{h,t\notin S_0}$.

This result motivates the following test procedure: Calculate x_{t_0+h,t_0} and $\bar{x}_{h,t\notin S_0}$ for some h which is taken to be larger than $\bar{h}$ – the horizon above which all temporary shocks have dissipated. Test the null hypothesis that x_{t_0+h,t_0} is significantly smaller (or larger) than $\bar{x}_{h,t\notin S_0}$. If the null is rejected, the shock is taken to be permanent.

Our suggested test procedure does not make use of the $h-1$ observations for which $t \in [t_0-h+1, t_0-1] \subset S_0$. The reason is that these observations are not necessarily informative about the permanency of the shock. To see this, suppose that $\beta = 0$ and $\varepsilon_{t_0+1} \geq -d$ so that the shock at t_0+1 is large but the process $y_{t+1} = \alpha + \mu(t+1) + \varepsilon_{t+1}$ will have recovered by time t_0+2. However, the 20-period ahead growth rate, say, at time $t_0 - 19 \in S_0$ has an expected value larger than $\mu - d/h$ whereas the same growth rate at time t_0 has the expected value μ. Hence, only the latter growth rate reveals the temporary nature of the shock. More generally, if we knew $\bar{h}$ all observations on $x_{t+h,t}$ for $h \geq \bar{h}$ and $t \in [t_0 - (h - \bar{h}), t_0]$ would be informative of permanent effects. But since we typically don't know $\bar{h}$ our best strategy is to pick some horizon, h, for which we think $h \geq \bar{h}$ holds and only focus on the observations in $t = t_0$.

It is interesting to note what happens under the alternative if we use $\bar{x}_h$ instead of $\bar{x}_{h,t \notin S_0}$. In this case, $\mu_h = \left(\frac{h}{T-h}\right)(\mu - d/h) + \left(\frac{T-2h}{T-h}\right)\mu = \mu - \frac{d}{T-h}$ which implies that the distance $\mu_h - \mu_{t_0,h}$ is reduced by the factor $(T-2h)/(T-h)$. This means that we can no longer be assured that we would be able to distinguish between μ_h and $\mu_{t_0,h}$. Nevertheless, finding that x_{t_0+h,t_0} is an outlier with respect to $\bar{x}_h$ also implies that it is one with respect to $\bar{x}_{h,t \notin S_0}$ and, hence, provides a stronger result.

A.3 Recurring permanent shocks

A more interesting case arises when there are several outlier observations, $t_{0j} + 1$ for $j = 1, ..., J$, that might have permanent effects, $d_j \geq \lambda \sigma_1$. Assuming for simplicity that $|t_{0j} - t_{0k}| \geq h$ for all $j \neq k$, we could still form $\bar{x}_{h,t \notin \cup_j S_{0j}}$ provided that $T - (J+1)h \geq h$. In this case, our test procedure becomes: Calculate $\bar{x}_{h,t \in \cup_j t_{0j}}$ and $\bar{x}_{h,t \notin \cup_j S_{0j}}$ for some h which is taken to be larger than $\bar{h}$. Test the null hypothesis that $\bar{x}_{h,t \in \cup_j t_{0j}}$ is significantly smaller than $\bar{x}_{h,t \notin \cup_j S_{0j}}$. If the null is rejected, the shocks are taken to be permanent.

Note that $|t_{0j} - t_{0k}| \geq h$ for all $j \neq k$ is a worst case scenario from the perspective of the power of the test. If $|t_{0j} - t_{0k}| < h$, there are both fewer elements in $\cup_j D_{0j}$ and for some horizons consecutive d_j's will pile up and thereby increase the distance between the observations at t_{0j} and $\bar{x}_{h,t \notin \cup_j S_{0j}}$.

What if the arrival of permanent shocks is random, for example following a Bernoulli distribution? In this case, the variance of $\bar{x}_h$ would be larger than (A.5) under the alternative and therefore make it harder to detect a difference in the means all else equal.

A.4 Issues

We end this section by discussing different issues that arise with the above approach. So far we have assumed that the t_0 observations occur during normal periods. But it may be more realistic to assume that large economic contractions are preceded by booms. If so, this would bias our result toward finding significant permanent effects even if there are none. We deal with this issue in the spirit of Blanchard et al. (2015) by using the points $\tilde{t}_0 = t_0 - 8$ in place of our identified t_0 as a robustness exercise. This implicitly assumes that booms last for at most two years, and that $\tilde{h} = 40 - 8 \geq \bar{h}$ if the maximum growth horizon is 10 years (ie 40 quarters).

Another issue is that our test cannot distinguish between the case where both expansions and contractions have permanent effects since it is one-sided. Of course, one can equally well find the $t_{0,pos}$ associated with large expansions and redo the tests to see if they have permanent effects. But this raises a thorny issue: what happens if we tests for positive permanent effects and there are, in truth, only negative ones? In this case, there will be a $S_{0,neg}$ set associated with each negative event where the mean is reduced by d/h. Since there are, in truth, no permanent positive effects, the expected value of $x_{t_{0,pos}+h,t_{0,pos}}$ is μ, but even so, this growth rate is higher than the growth rate outside of $S_{0,pos}$ due to the presence of the $S_{0,neg}$ sets. Hence, we are likely to reject the null hypothesis of identical means and erroneously conclude expansions have permanent effects. Indeed, if either expansions or contractions have permanent effects, we are likely to find symmetric results w.r.t. both. How can we then be certain that finding significant negative permanent effects, say, is not just reflective of permanent positive effects? One possibility is to simultaneously identify $t_{0,neg}$ and $t_{0,pos}$ and then remove $S_{0,pos}$ from the sample before testing if the mean of the $t_{0,neg}$ observations is different from the mean of the remaining observations (and vice versa for the mean of the $t_{0,neg}$ observations). While we do implement this approach, it is costly in terms of removed observations as we need to remove 39 quarters before each identified event.

An alternative, less costly, approach is to first exclude the possibility that there are both positive and negative permanent effects. This can be done by looking at the unconditional growth distribution at horizon h. If this distribution is bimodal, the permanent effects are likely to be one-sided (either positive or negative), and if it is trimodal the effects are likely to be two-sided. The reason is that the unconditional mean is different for $x_{t+h,t} \in \cup_j S_{0j,neg}$, $x_{t+h,t} \in \cup_j S_{0j,pos}$ and $x_{t+h,t} \notin (\cup_j S_{0j,neg}) \cup (\cup_j S_{0j,pos})$ in the permanent cases. Once, two-sided permanent effects are excluded, one can simply look at the behavior of the output series around $t_{0,pos}$ and $t_{0,neg}$ to determine which case is likely to prevail. We also note two further cases are of interest: when no disruptions have permanent effects and when all disruptions have permanent effects. In both of these cases, we should see a unimodal unconditional growth distribution. But $\bar{x}_{h,t\in\cup_j t_{0j}} \approx \bar{x}_{h,t\notin\cup_j S_{0j}}$ is the former case so that the null of equal means cannot be rejected, whereas the distance between the two means should increase evenly and symmetrically with the distance from the median in the latter case.

The last issue that we need to address is the possibility that the underlying growth rate, μ, of the y_t process is not fixed over time. For instance, evidence suggest that growth has slowed down in many economies over the past decades (e.g., Antolin-Diaz et al., 2017). Hence, it might be better to replace $\mu(t+1)$ with μ_{t+1} in (A.1). For instance, setting $\mu_{t+1} = \mu + \mu_t + v_{t+1}$ would lead to a difference stationary specification for y_{t+1}. However, from the perspective of detecting permanent outliers, such a growth slowdown is problematic as it would bias x_{t_0+h,t_0} downward as h increases, making it more likely to misinterpret the outlier as permanent even when it is not. In our application, we deal with this issue by removing slow-moving trends from our growth rates prior to the analysis.

B Additional tables and figures

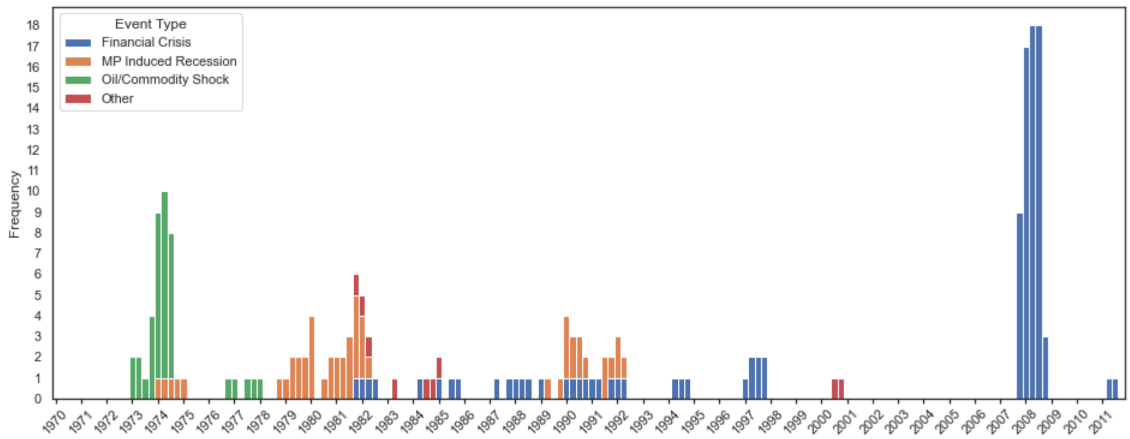

Figure B.1: Event type frequencies over time.

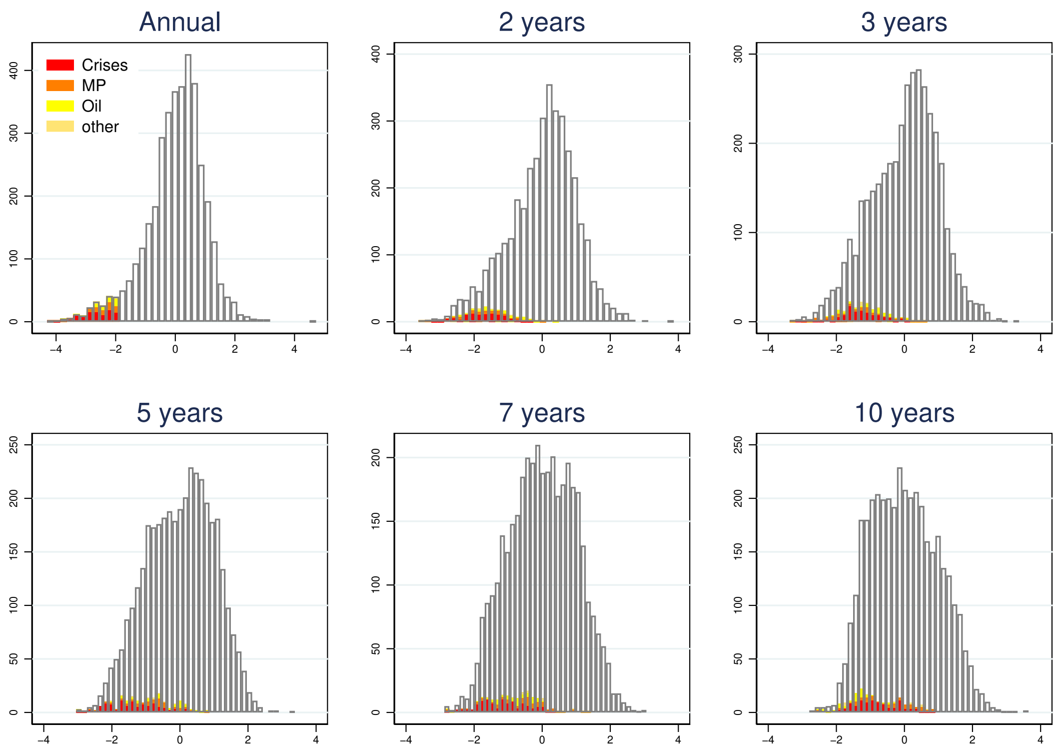

Figure B.2: Distribution of annual , 3-year, 5-year and 10-year real growth rates (normalised) around severe contractions at or below the 5th percentile of the annual growth distribution.

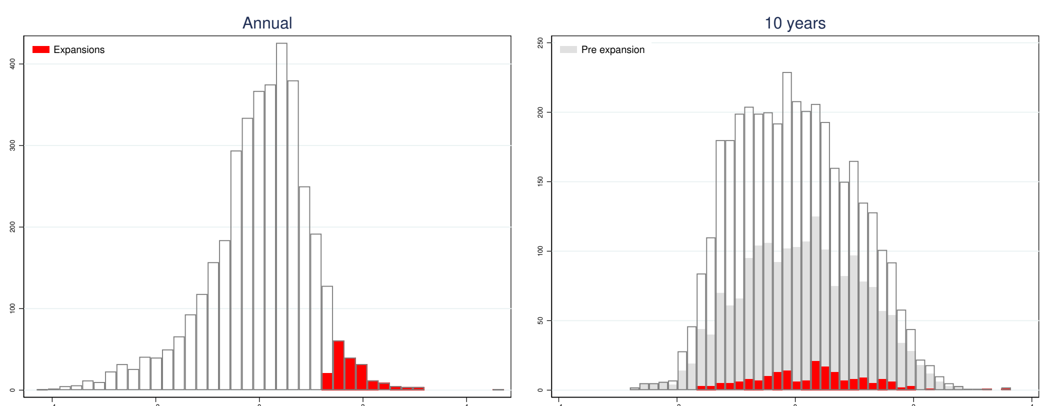

Figure B.3: Distribution of annual and 10-year real growth rates (normalised) around large expansions above the 95th percentile of the annual real GDP growth distribution.

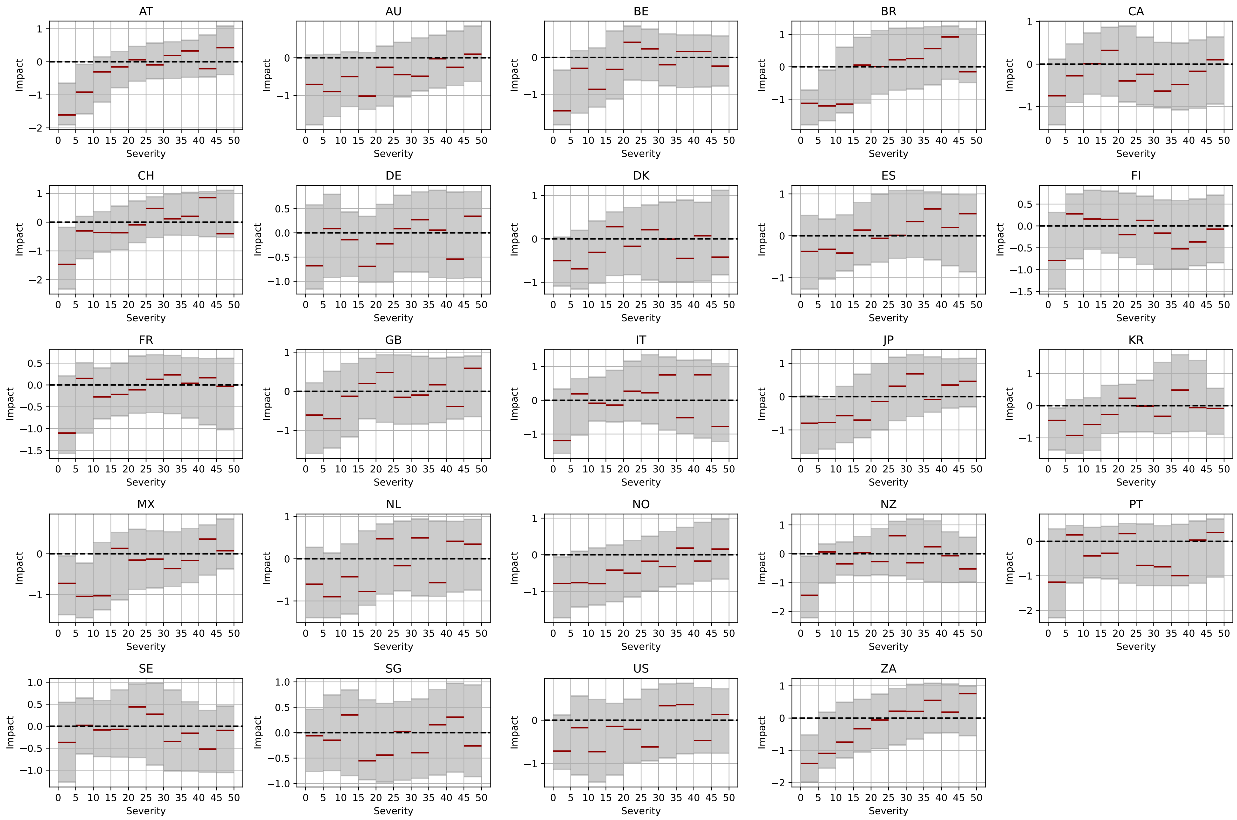

Figure B.4: Mean differences in 10-year real growth rates at the country level following severe contractions at different percentiles. Shaded areas are the 95% confidence intervals. The y-axis is in standard deviations of 10-year real GDP growth.

	Annual				10 year			
	Raw		Detrended		Raw		Detrended	
Country	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
AT	2.31	1.93	0.10	1.82	2.18	0.59	-0.04	0.36
AU	3.01	1.74	-0.05	1.71	3.08	0.39	-0.03	0.28
BE	2.16	1.83	0.09	1.76	2.05	0.46	-0.03	0.33
BR	2.15	3.76	-0.35	3.71	2.55	0.77	-0.01	0.68
CA	2.67	2.16	0.09	2.08	2.55	0.64	-0.03	0.49
CH	1.67	2.25	0.01	2.22	1.69	0.50	0.01	0.42
DE	1.93	2.13	0.06	2.06	1.85	0.61	-0.02	0.37
DK	1.85	2.17	0.07	2.14	1.75	0.60	-0.05	0.44
ES	2.50	2.39	0.15	2.24	2.30	1.04	-0.12	0.74
FI	2.32	3.21	0.03	3.09	2.30	1.14	-0.03	0.88
FR	2.15	1.72	0.12	1.56	2.01	0.64	-0.03	0.33
GB	2.17	2.24	0.02	2.20	2.14	0.62	-0.06	0.46
IT	1.63	2.46	0.05	2.14	1.55	1.19	-0.03	0.38
JP	2.33	2.70	0.07	2.15	2.23	1.56	0.00	0.46
KR	6.63	4.28	-0.06	3.48	6.68	2.20	-0.07	0.61
MX	3.10	3.49	0.15	3.24	2.89	1.27	0.04	0.87
NL	2.25	2.09	0.07	1.99	2.16	0.79	-0.08	0.50
NO	2.77	1.99	0.09	1.88	2.64	0.69	-0.01	0.56
NZ	2.59	3.86	0.12	3.82	2.44	0.76	-0.06	0.53
PT	2.46	3.20	0.20	2.90	2.23	1.37	-0.07	0.60
SE	2.13	2.16	0.00	2.14	2.13	0.55	-0.02	0.48
SG	6.23	4.12	-0.14	3.90	6.37	1.17	-0.01	0.74
US	2.75	2.09	0.01	2.01	2.73	0.68	-0.06	0.41
ZA	2.31	2.36	-0.02	2.28	2.35	0.90	0.03	0.70
Average	**2.67**	**2.60**	**0.04**	**2.44**	**2.62**	**0.88**	**-0.03**	**0.53**

Table B.1: Descriptive statistics.

Country	Fin Crisis	MP	Oil Shock
AT	08q1, 08q2, 08q3	80q1	74q1, 74q2, 74q3, 77q4
AU		81q3, 81q4, 82q1, 82q2, 90q1, 90q2, 90q3, 90q4	73q4
BE	07q4, 08q1, 08q2, 08q3	80q1, 92q1	74q1, 74q2, 74q3
BR		80q3, 80q4, 89q2, 89q4, 90q1, 91q3	
CA	08q1, 08q2, 08q3	81q1, 81q2, 81q3, 81q4, 82q1, 90q1, 90q2, 90q3	
CH	08q1, 08q2, 08q3	74q1, 74q2, 74q3, 74q4, 75q1	
DE	08q1, 08q2, 08q3, 08q4		74q1, 74q2,
DK	89q1, 07q4, 08q1, 08q2, 08q3,	80q1	74q1
ES	08q1, 08q2, 08q3, 11q2, 11q3	92q1, 92q2,	
FI	90q1, 90q2, 90q3, 90q4, 91q1, 91q2, 08q1, 08q2, 08q3, 08q4		
FR	07q4, 08q1, 08q2, 08q3		74q1, 74q2, 74q3
GB	07q4, 08q1, 08q2, 08q3	79q2, 79q4, 80q1	73q1, 73q2, 74q2, 74q3,
IT	08q1, 08q2, 08q3		74q1, 74q2, 74q3,
JP	07q4, 08q1, 08q2, 08q3		73q1, 73q2, 73q3, 73q4
KR	97q1, 97q2, 97q3, 97q4,	78q4, 79q1, 79q2, 79q3, 79q4	
MX	81q4, 82q1, 82q2, 82q3, 85q3, 85q4, 94q2, 94q3, 94q4, 08q1, 08q2, 08q3		
NL	08q1, 08q2, 08q3, 08q4	80q4, 81q2, 81q4, 82q1	78q1,
NO	87q2, 87q4, 88q1, 88q2, 88q3, 07q4, 08q2, 08q3		
NZ	85q1		74q2, 74q3, 76q4, 77q, 177q3
PT			73q4, 74q1, 74q2, 74q3
SE	91q4, 92q1, 92q2, 07q4, 08q1, 08q2, 08q3		
SG	97q2, 97q3, 97q4, 07q4, 08q1		
US	07q4, 08q1, 08q2, 08q3	79q3, 81q1, 81q3, 81q4	73q4, 74q1, 74q2
ZA	84q2, 08q2, 08q3	91q3, 91q4	

Table B.2: Classification of severe contractions. The dates for the "other" category are: PT (83q2), SG (84q3, 84q4, 85q1, 00q3, 00q4), ZA (81q4, 82q1, 82q2).