BANK FOR INTERNATIONAL SETTLEMENTS

BIS Working Papers
No 407

Do economies stall? The international evidence

by Wai-Yip Alex Ho and James Yetman

Monetary and Economic Department

March 2013

JEL classification: C22, E32, E37

Keywords: Business cycles, stall speed, Markov switching

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)

ISBN 1682-7678 (online)

Do economies stall? The international evidence

Wai-Yip Alex Ho and James Yetman[1]

Abstract

A "stalling" economy has been defined as one that experiences a discrete deterioration in economic performance following a decline in its growth rate to below some threshold level. Previous efforts to identify stalls have focused primarily on the US economy, with the threshold level being chosen endogenously, and have suggested that the concept of a stall may be useful for macroeconomic forecasting.

We examine the international evidence for stalling in a panel of 51 economies using two different definitions of a stall threshold (time-invariant and related to lagged average growth rates) and two complementary empirical approaches (in-sample statistical significance and out-of-sample forecast performance). We find that the evidence for stalling based on time-invariant thresholds is limited: only 12 of the 51 economies in our sample experience statistically significant stalls, and including a stall threshold generally results in only modest improvements to out-of-sample forecast performance. When we instead model the stall threshold as varying with average growth rates, the number of economies with statistically-significant stalls actually declines (to nine), but in 71% of the cases we examine, including a stall threshold results in an improvement in out-of-sample forecast performance.

Keywords: Business cycles, stall speed, Markov switching

JEL classifications: C22, E32, E37

[1] Ho is Manager at the Hong Kong Monetary Authority (HKMA). Yetman is Senior Economist at the Bank for International Settlements (BIS). The views expressed here are the authors' and do not necessarily reflect those of the BIS or the HKMA. We thank Russell Davidson for advice on bootstraps and Nathan Sheets and seminar participants at the BIS, the Bank of Korea and the Hong Kong Institute for Monetary Research for comments. Any remaining errors are ours alone.

Introduction

Recently, some market analysts and central bank researchers have suggested that economies, like aircraft, can stall. While different precise definitions of macroeconomic stalling have been considered, the central idea is that a slow rate of economic growth will tend to be followed by a discrete deterioration in economic performance. In one form of the stalling hypothesis, such a slowdown will lead to a recession. If economies stall, this characteristic would be helpful for modelling and forecasting macroeconomic growth.

For example, Nalewaik (2011) modelled a stalling economy as one in which the growth rate is too low to sustain normal growth and which therefore slips into recession. He illustrated the quantitative value of including a low-growth state consistent with this in a Markov switching model of the US economy. He found that hitting the low-growth state significantly increased the likelihood of entering the recession state in the following period, although this appears to be sensitive to how GDP growth is defined and the inclusion of additional variables in the empirical model.[2]

Sheets (2011) and Sheets and Sockin (2012) [hereafter *Sheets*] considered an alternative definition of stalling as a decline in the year-on-year growth rate of real GDP to below some threshold. They showed that, thus defined, stalling appears to play an important role in predicting a future slowdown, not just in the United States but in other economies as well, based on graphical analysis and regression results.

Ho and Yetman (2012) considered both definitions of stalling in US GDP data using kernel density estimates, probit estimates and Markov regime switching models. They found that if a stall is defined as a low but positive quarter-on-quarter growth rate of real GDP, as in Nalewaik (2011), there is no evidence of stalling in US GDP data. In contrast, if a stall is defined as a decline in the year-on-year growth rate of the economy to below some threshold, as in Sheets, then stalling appears to significantly increase the likelihood of a future recession, at least in-sample, for the United States.

There are two important differences between these tests for stalling. First, the variable used to define a stall – quarter-on-quarter growth versus year-on-year growth at quarterly frequency – turns out to be important to the US results. Below, we will illustrate how the evidence of stalling is greater based on year-on-year growth rates than quarter-on-quarter growth rates.

Second is the way in which the different definitions select different episodes as potential "stalls". Real GDP growth rates are persistent in the short run, and have a tendency to mean-revert. As a result, low-growth periods generally occur at two points during the business cycle: i) shortly before the economy enters a recession, and ii) as the economy exits a recession. If one defines stalling as simply "experiencing a low level of growth", as in Nalewaik (2011), both types of low-growth episodes are considered together: those that tend to be followed by higher growth, and those that tend to be followed by recessions. In contrast, if one only considers low-growth episodes that occur when the economy is slowing down, as in

2 In a model based on quarter-on-quarter growth rates, as in Nalewaik (2011), stalling is evident when both GDI and GDP growth are included, but not when GDP growth is considered alone (see Ho and Yetman (2012) for a discussion).

Sheets, this may help to single out episodes that precede recessions. Thus it is not surprising that the latter definition of stalling will generally result in greater evidence of stalling than the former.

But there is a potential problem with looking for evidence of stalling. Stalling in macroeconomics is an empirical concept, and it is not clear ex ante which level of growth should be identified with stalling. One tempting approach is to examine the data and then choose the stall speed that appears to fit best, based either on casual empiricism or formal estimation. However, depending on how the stall threshold is selected, this may exaggerate the evidence in favour of stalling.[3]

Here, we examine the evidence for stalling in a panel of 51 economies, taking care to avoid overstating the fit. We first define a stall as a decline in the growth rate of the economy, to below some threshold level, as in Sheets, and select the threshold level as the one that makes the model fit the data best in terms of minimising the sum of squared residuals. Based on standard critical values, our initial results suggest moderate evidence for stalling: in 14 (out of 51) economies, we find that stalling is a statistically significant phenomenon at the 5% level. But, given that our panel includes a wide variety of economies, including many emerging market economies that have seen significant declines in their trend growth rate over the sample, we also consider a second definition of stalling where the stall threshold is characterised by the growth rate relative to its recent trend, where the trend is defined as a backward-looking 40-quarter moving average. By this alternative definition, only 11 economies appear to stall, based on standard critical values.

Statistical inference based on our estimates is complicated by the fact that the stall threshold is unidentified under the null hypothesis that the economy does not stall.[4] Thus statistical tests of the importance of stalling do not have the standard asymptotic distributions and inference based on standard critical values may be misleading. Using an approach that parallels Hansen (1996, 1997), we correct for this problem using a careful bootstrap exercise. Resampling the data using a sieve bootstrap based on a data generating process that, by construction, does not include a role for stalling, we calculate corrected p-values. We find that the evidence for stalling is then a little weaker. In the case of a fixed stall threshold, instead of 14 economies, nine to 12 stall at the 5% level, depending on the number of lags we include in our sieve bootstrap. And in the case of a time-varying threshold, where the stall threshold depends on the current growth rate relative to its 40-quarter moving average, 9 out of the 11 economies that appeared to stall based on standard critical values remain significant stallers based on the bootstrap results.

We also assess the effect of including stall dynamics on out-of-sample forecast performance. Given that the concept of stalling may potentially be useful for forecasting, it is natural to consider the effect of including a stall threshold on forecast performance. We find that modelling the stall phenomena as occurring at a fixed threshold results on average in a modest improvement in out-of-sample forecast performance. In contrast, assuming that the threshold varies with the

[3] In practical terms, naïve statistical tests suggest that there is much greater evidence for stalling if the stall threshold is selected so as to maximise the statistical significance in favour of stalling (ie the test statistic on the stall variables) instead of maximising the explanatory power of the model (ie minimising the residual sum of squares of the estimated model). The former approach may be thought of variously as a form of selection bias, pre-test bias or data mining.

[4] See, for example, the discussion in Hansen (1996, 1997, 2000).

average growth rate offers much greater evidence in favour of stalling: based on the Clark and West (2007) test statistic, forecast performance improves in 71% of the cases that we examine. Further, our results suggest that alternative threshold effects, which are not consistent with the idea of stalling but are more akin to a rebound from low growth rates, also significantly improve out-of-sample forecast performance in this case.

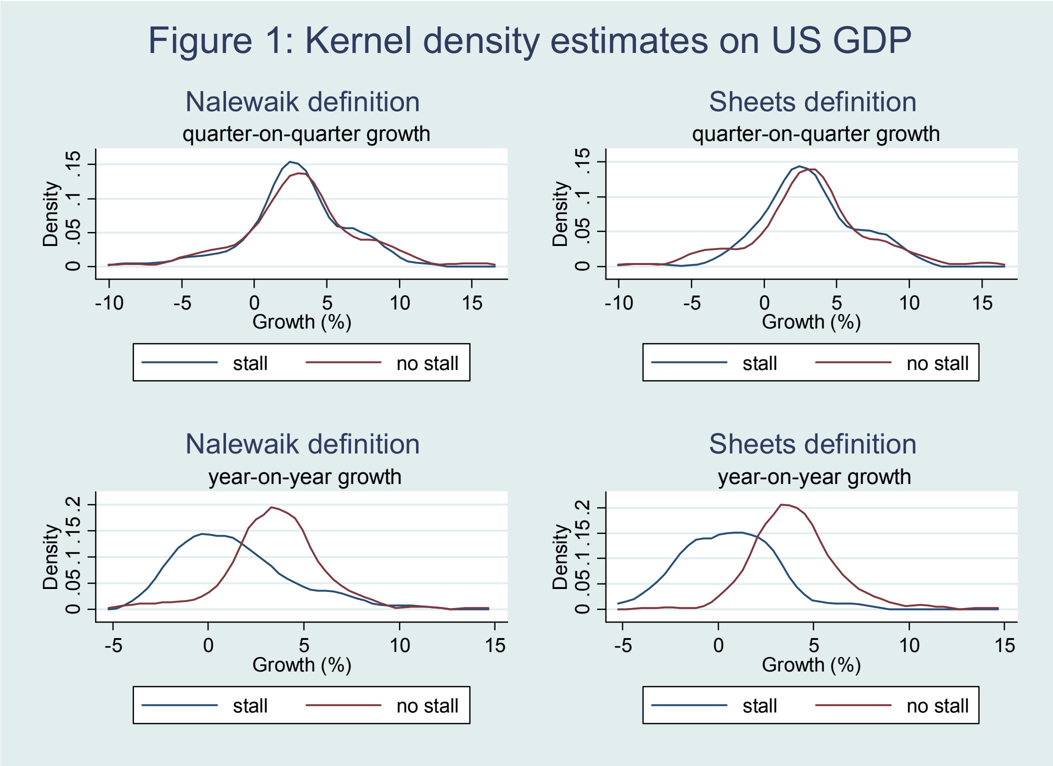
In the next section, we outline the intuition for our arguments based on kernel density estimates using US data. We then examine the evidence for stalling based on a stall threshold that is assumed to be time-invariant in Section 2. In Section 3, we repeat the analysis under the assumption that the stall threshold varies with the trend growth rate, defined as a backward-looking 40-quarter moving average. Finally, we conclude.

1. Stalling: the intuitive evidence

To illustrate the idea of stalling, we first provide some simple, graphical evidence based on kernel density forecasts on US data. Figure 1 displays four graphs based on different definitions of stalling. Each graph contains two kernel density estimates,[5] one based on observations one to four periods following a "stall", and one based on all other observations. In the left-hand graphs the classification of stalling is based on Nalewaik (2011): real GDP growing at 0–1%. In the right-hand panels, it is based on Sheets: real GDP growth slowing from above 1.5% to between 0 and 1.5%. Vertically, in the top row a stall is based on quarter-on-quarter GDP growth rates (annualised), as in Nalewaik (2011), and the bottom row on year-on-year growth rates at quarterly frequency, as in Sheets.[6]

5 These estimates are based on the Epanechnikov kernel and the default bandwidth setting in Stata.

6 In Sheets, a stall is defined in terms of year-on-year growth rates, but the variable being explained by a stall is quarter-on-quarter growth rates (annualised). Here we use year-on-year growth rates for both variables.



Comparing the four panels, there is little evidence of stalling, and little difference between the two measures, based on quarter-on-quarter growth (top row). However, based on year-on-year growth there is considerable evidence of stalling, and more so using the Sheets definition (right-hand-lower) than the Nalewaik definition (left-hand-lower): the decline in mean growth rates in the four quarters following a stall compared with those following other growth periods is 2.17% using the Nalewaik (2011) definition, and 3.52% using the Sheets definition.

In what follows, we will focus on the case presented here with the greatest evidence of stalling in the case of the United States: the Sheets definition of a stall, based on year-on-year growth rates in real GDP. This offers the added advantage that we are using data for a large number of economies, drawn from national sources, and year-on-year data allow us to work with non-seasonally adjusted data, increasing the comparability of our results across economies.[7]

We will answer the following questions: is stalling a common characteristic across economies? Is it statistically significant? And can it be used to improve forecast performance?

[7] Given that methods of seasonal adjustment are likely to vary by country, we use non-seasonally adjusted data wherever possible. Exceptions to this are Great Britain, Israel, Lithuania, Portugal, South Africa and the United States.

2. Empirical evidence for stalling: fixed threshold

2.1 Naive estimation

Our sample is real GDP drawn from publicly available national sources. We include a large panel of advanced and emerging market economies for which we can obtain at least 10 years of data, and construct year-on-year real GDP growth, at quarterly frequency. The full sample, along with summary statistics, is given in Table 1 in order of the name of the economy.

To estimate a model of threshold effects for each economy in our sample, we take the following steps:

1. Regress the year-on-year growth rate of real GDP, y_t, on up to four lags, selecting the number of lags L using the Akaike information criterion (AIC)
2. Consider every possible stall threshold level γ between zero and the mean growth rate in steps of 0.01. Then, at each possible stall level:
   a. Construct a stall dummy D_t that takes the value 1 if growth passed from above the stall level to below the stall level in period t, and 0 otherwise
   b. Regress the growth rate on a constant, the number of lags determined by the AIC in step 1, and four lags of the stall dummy, as in Sheets:
   $$y_t = \alpha + \sum_{l=1}^{L} \beta_l y_{t-l} + \sum_{m=1}^{4} \delta_m D_{t-m} + \varepsilon_t$$
   c. Calculate the sum of squared residuals (SSR):
   $$SSR(\gamma) = \sum_t \left(\hat{\varepsilon}_t\right)^2$$
3. Then compare the test results for all possible stall levels, and the threshold is the one for which $\gamma^* = \operatorname{argmin} SSR(\gamma)$. The economy is identified as a "staller" if $\bar{\delta}\left(\gamma^*\right) = \sum_{m=1}^{4} \delta_m$ is statistically significantly negative, based on a one-sided test.

The approach taken here is very similar to a self-exciting threshold auto-regressive (SETAR) model, in which the model parameters are regime-dependent and the regime itself is a function of the lags of the dependent variable. Our model can be viewed as a version of the SETAR model where only the constant term varies across regimes, but where the effect of crossing the threshold has an effect for multiple periods.

We impose one additional restriction when selecting stall levels. Since we are interested in identifying a regularity in the data that is useful for modelling and forecasting growth, we restrict ourselves to stall levels that are triggered at least three times over the sample period.[8] The results from applying this process across all the economies in our sample are given in Table 2, with stallers indicated in red.

[8] Results are similar for a range of alternative choices on how many times a stall threshold must be triggered to be considered, from just once to a minimum of five times.

GDP growth data summary

Table 1

Country	Code	Start	End	Mean	Standard deviation
Argentina	AR	1981Q1	2011Q4	2.91	6.40
Australia	AU	1960Q3	2012Q1	3.57	2.37
Austria	AT	1961Q1	2012Q1	2.95	2.29
Belgium	BE	1973Q1	2012Q1	2.07	2.57
Brazil	BR	1981Q1	2012Q1	3.15	4.69
Canada	CA	1962Q1	2012Q1	3.38	2.55
Chile	CL	1980Q1	2012Q1	4.79	5.17
China	CN	1993Q1	2012Q1	10.28	2.12
Chinese Taipei	TW	1962Q1	2012Q1	7.35	4.17
Colombia	CO	1995Q1	2012Q1	3.36	2.96
Croatia	HR	1998Q1	2011Q4	3.41	5.54
Czech Republic	CZ	1997Q1	2012Q1	2.68	3.13
Denmark	DK	1978Q1	2012Q1	2.02	3.14
Estonia	EE	1994Q1	2012Q1	4.68	6.54
Finland	FI	1971Q1	2012Q1	2.65	3.44
France	FR	1964Q1	2012Q1	2.69	2.28
Germany	DE	1961Q1	2012Q1	2.54	2.53
Great Britain	GB	1956Q1	2012Q1	2.36	2.34
Greece	GR	1996Q1	2012Q1	1.83	3.89
Hong Kong SAR	HK	1967Q1	2012Q1	6.02	5.09
Hungary	HU	1996Q1	2012Q1	2.24	2.95
India	IN	1961Q1	2011Q4	5.32	3.97
Indonesia	ID	1981Q1	2012Q1	5.54	4.24
Ireland	IE	1977Q1	2011Q4	4.53	4.08
Israel	IL	1991Q1	2012Q1	5.06	4.89
Italy	IT	1961Q1	2012Q1	2.68	2.95
Japan	JP	1956Q2	2012Q1	4.45	4.24
Korea	KR	1961Q1	2012Q1	7.63	4.91
Latvia	LV	1997Q1	2012Q1	4.64	7.65
Lithuania	LT	1996Q1	2012Q1	5.21	6.29
Malaysia	MY	1989Q1	2012Q1	6.09	4.53
Mexico	MX	1981Q1	2012Q1	2.49	3.80
Netherlands	NL	1961Q1	2012Q1	2.91	2.54
New Zealand	NZ	1978Q2	2012Q1	2.61	3.01
Norway	NO	1979Q1	2012Q1	2.69	2.63
Peru	PE	1980Q1	2012Q1	3.35	7.42
Philippines	PH	1974Q1	2011Q4	3.80	3.81
Poland	PL	1996Q1	2012Q1	4.41	2.28
Portugal	PT	1961Q1	2012Q1	3.49	3.50
Romania	RO	2001Q1	2012Q1	3.90	4.51
Singapore	SG	1976Q1	2012Q1	6.98	4.54
Slovakia	SK	1991Q1	2012Q1	2.81	6.24
Slovenia	SI	1993Q1	2012Q1	3.48	3.63
South Africa	ZA	1961Q1	2012Q1	3.20	2.74
Spain	ES	1971Q1	2012Q1	2.79	2.42
Sweden	SE	1971Q1	2012Q1	2.07	2.54
Switzerland	CH	1966Q1	2012Q1	1.98	2.68
Thailand	TH	1994Q1	2012Q1	3.61	5.33
Turkey	TR	1988Q1	2011Q4	4.20	5.86
United States	US	1956Q1	2012Q1	2.77	2.34
Venezuela	VE	1994Q1	2012Q1	2.47	8.10

Results from estimated stall model – fixed thresholds

Table 2

Code	Lags (L)	Stall Level γ^* Minimum	Stall Level γ^* Maximum	Size of stall ($\bar{\delta}$)	t- test statistic	p-value	Significance
AR	4	2.54	2.65	-2.03	-0.78	0.219	
AU	4	0.00	0.12	-3.99	-1.94	0.027	**
AT	4	0.88	1.00	-0.23	-0.16	0.435	
BE	4	0.94	0.96	1.64	1.19	0.883	
BR	4	2.26	2.31	3.43	1.40	0.917	
CA	4	3.27	3.30	0.71	0.86	0.804	
CL	4	2.22	2.58	0.95	0.34	0.633	
CN	1	8.89	8.90	0.41	0.40	0.654	
TW	4	5.47	5.47	-4.91	-2.93	0.002	***
CO	4	2.28	2.41	-1.45	-0.79	0.218	
HR	3	0.00	0.05	-10.57	-2.45	0.009	***
CZ	4	0.02	0.33	-12.10	-2.29	0.013	**
DK	4	1.93	1.93	2.84	2.14	0.983	
EE	4	2.67	3.11	-0.74	-0.18	0.429	
FI	4	1.27	1.28	0.23	0.19	0.576	
FR	4	0.00	0.37	2.71	2.26	0.988	
DE	4	2.12	2.16	-0.64	-0.82	0.207	
GB	4	0.00	0.00	-1.29	-1.13	0.130	
GR	1	0.17	0.37	-6.54	-2.81	0.003	***
HK	4	4.40	4.56	-3.31	-1.43	0.078	*
HU	2	1.69	2.20	-6.17	-2.95	0.002	***
IN	4	5.25	5.29	0.93	0.76	0.776	
ID	3	2.07	3.67	-10.60	-3.34	0.001	***
IE	4	1.93	2.17	4.09	2.17	0.984	
IL	2	0.84	0.90	3.22	1.27	0.896	
IT	4	2.65	2.67	3.10	3.39	1.000	
JP	3	0.75	0.87	-1.19	-0.83	0.205	
KR	4	7.53	7.53	6.15	2.93	0.998	
LV	4	1.65	2.53	-4.09	-0.97	0.169	
LT	4	3.03	3.58	-9.29	-1.72	0.045	**
MY	3	2.90	4.09	-8.05	-1.85	0.034	**
MX	3	0.00	0.09	-13.32	-4.26	0.000	***
NL	4	2.17	2.22	0.71	0.79	0.785	
NZ	4	2.21	2.24	1.00	0.64	0.739	
NO	4	0.93	1.41	-2.25	-1.59	0.057	*
PE	2	0.00	0.14	2.87	0.81	0.790	
PH	2	1.64	1.77	-2.52	-1.01	0.156	
PL	4	4.38	4.40	2.34	1.86	0.966	
PT	4	0.51	0.53	-3.52	-2.36	0.010	***
RO	2	1.49	1.79	-2.72	-0.55	0.295	
SG	4	4.25	4.30	-6.97	-2.28	0.012	**
SK	4	1.84	1.99	-5.10	-1.54	0.064	*
SI	3	2.11	2.21	0.40	0.15	0.559	
ZA	4	3.17	3.20	1.20	1.44	0.924	
ES	4	1.88	2.03	-0.99	-1.42	0.078	*
SE	4	1.63	1.70	-1.00	-1.13	0.131	
CH	4	0.00	0.00	-0.62	-0.45	0.325	
TH	4	3.06	3.16	-2.58	-0.85	0.198	
TR	4	2.63	2.66	-9.40	-2.11	0.019	**
US	3	0.99	1.00	-1.28	-1.44	0.075	*
VE	4	1.74	2.13	-18.33	-3.20	0.001	***

We report two threshold levels: a minimum and a maximum. This is because there is no unique stall level identified by this approach for many economies. For example, for Argentina, all threshold levels from 2.54% to 2.65% yield the same sum of squared residuals as there are no observations in the Argentinian dataset where the growth rate of the economy fell from between 2.54% and 2.65% in one period to a level of growth below 2.54% in the following. Thus, any threshold level from 2.54% to 2.65% fits the data identically well. The range of identified stall levels varies widely, from 0.00% (Chinese Taipei, Denmark, Great Britain, Korea and Switzerland) to more than 1.00% (Indonesia and Malaysia).

Similarly, the size of the identified subsequent cumulative four-quarter fall in GDP following a stall, given by $\bar{\delta}$, varies widely from more than 10% (Croatia, the Czech Republic, Indonesia, Mexico and Venezuela) to an increase of more than 3% (Brazil, Ireland, Israel, Italy and Korea). One could think of these latter cases, where a slowdown to below some threshold is associated with higher future growth, as an economic rebound rather than a stall. Further, for some of these economies, the associated p-value from our one-sided test is very close to 1, indicating that this rebound threshold is statistically significant.

Finally, the p-values of the test $\mathrm{H_0}: \bar{\delta} < 0$ are given for each economy, along with an indication of significance at the 10% (*), 5% (**) or 1% (***) level based on conventional critical values. The results suggest that 19 of the 51 economies in our sample have statistically significant stalls at the 10% level, 14 at the 5% level and 8 at the 1% level. At face value, these results suggest that stalling may be a useful concept for modelling and forecasting GDP growth, for at least a portion of the economies in our sample.

2.2 Bootstrap correction

One problem with the results outlined in the above section is that their statistical significance was assessed in terms of standard critical values. However, the distribution of the test statistics is non-standard, since the stall threshold is not defined under the null hypothesis, and inference based on standard critical values may therefore be biased (see Hansen (1996, 1997) for details).

To correct for this, we take an approach that parallels Hansen (1996, 1997) and construct a sieve bootstrap as outlined in MacKinnon (2006) to generate corrected critical values. The sieve bootstrap is a semi-parametric bootstrap that is intended to approximate time-dependent data well. We take the estimated residuals from the estimated GDP process without the stall dummies (see step 1 in the last section; the *null model* hereafter) and fit them with an auto-regressive process, as follows:

$$\hat{\varepsilon}_t = \sum_{i=1}^{p} \rho_i \hat{\varepsilon}_{t-i} + u_t \ ,$$

where p is chosen using the Akaike information criterion (AIC). To check the robustness of our bootstrap results, we consider maximum lag lengths (p) of 4 and 8.

Using our estimates of the autoregressive parameters $\hat{\rho}_i$ from this equation, we can then compute artificial residuals recursively using

$$\varepsilon_t^* = \sum_{i=1}^{p} \hat{\rho}_i \varepsilon_{t-i}^* + u_t^*,$$

where u_t^* are resampled from $\hat{u}_t$ (the residual of residual in the previous equation) with replacement, rescaled by the factor $\sqrt{n/(n-p)}$ to correct for degrees of freedom.[9] We then generate artificial data on real GDP growth recursively using the equation

$$y_t^* = \hat{\alpha} + \sum_{l=1}^{L} \hat{\beta}_l y_{t-l}^* + \varepsilon_t^* ,$$

where $\hat{\alpha}$, $\hat{\beta}_l$ and L are drawn from the null model. Our simulated data should then have similar properties to the actual real GDP growth rate data, with the notable exception that, by construction, there are no stalls in the data generating process.

We construct 1000 such artificial samples for each economy and then repeat the same estimation process outlined in Section 2 on each sample. We then determine the proportion of test statistics in our artificial samples that are larger than we obtained in our data. For every such test statistic across our artificial sample, the p-value from our bootstrap exercise increases (from zero) by 0.001.

The p-values based on both standard critical values and the bootstraps are contained in Table 3. These are colour-coded: yellow indicates significance at the 10% level, orange at the 5% and red at the 1%.

The bootstrap exercise suggests that there is less evidence that economies stall than standard critical values would imply. Instead of 19 economies with statistically significant stalls at the 10% level, there are now 12. At the 5% level of significance, 14 drop to 9. And at the 1% level, the number of economies with statistically significant stalls falls from 8 to 3. If the number of lags used in the autoregressive representation of the residuals in the bootstrap is increased to a maximum of up to 8, then the number of economies with statistically significant stalls at the 10%, 5% and 1% are 12, 12 and 6, respectively.[10]

We also considered a number of robustness checks: allowing up to 8 lags in the AR representation of the economies (L), and alternative sample start points (1960, 1970 and 1980). In all these cases, the results are similar to those reported above.

Overall, based on a constant stalling threshold, these results suggest some evidence that GDP stalls, although this appears to be concentrated in emerging market economies. Further, we find that the evidence in favour of stalling for the US economy is statistically insignificant using this approach.

9 To deal with the initial value problem, we generate much longer series than we require to replicate our sample and then discard sufficient initial observations to match our sample size.

10 The full set of economies that exhibit statistically significant stalls at the 10% level based on our bootstrap results are Australia, Chinese Taipei, Croatia, the Czech Republic, Greece, Hungary, Indonesia, Mexico, Portugal, Singapore, Turkey and Venezuela.

p-values from standard distribution and bootstrap – fixed thresholds

Table 3

Code	Standard	4 lags	8 lags
AR	0.219	0.386	0.311
AU	0.027	0.071	0.037
AT	0.435	0.480	0.255
BE	0.883	0.665	0.472
BR	0.917	0.595	0.448
CA	0.804	0.720	0.604
CL	0.633	0.495	0.492
CN	0.654	0.604	0.583
TW	0.002	0.025	0.008
CO	0.218	0.347	0.225
HR	0.013	0.055	0.038
CZ	0.009	0.066	0.048
DK	0.983	0.925	0.785
EE	0.429	0.467	0.491
FI	0.576	0.557	0.532
FR	0.988	0.864	0.880
DE	0.207	0.335	0.256
GB	0.130	0.155	0.120
GR	0.003	0.014	0.002
HK	0.078	0.229	0.232
HU	0.002	0.021	0.009
IN	0.776	0.474	0.384
ID	0.001	0.009	0.007
IE	0.984	0.808	0.803
IL	0.896	0.625	0.396
IT	1.000	0.954	0.953
JP	0.205	0.218	0.232
KR	0.998	0.977	0.935
LV	0.169	0.279	0.318
LT	0.045	0.133	0.140
MY	0.034	0.152	0.173
MX	0.000	0.000	0.000
NL	0.785	0.530	0.525
NZ	0.739	0.647	0.658
NO	0.057	0.155	0.139
PE	0.790	0.685	0.453
PH	0.156	0.326	0.514
PL	0.966	0.837	0.876
PT	0.010	0.032	0.036
RO	0.295	0.417	0.443
SG	0.012	0.073	0.045
SK	0.064	0.214	0.216
SI	0.559	0.504	0.476
ZA	0.924	0.484	0.497
ES	0.078	0.159	0.137
SE	0.131	0.293	0.208
CH	0.325	0.432	0.402
TH	0.198	0.370	0.385
TR	0.019	0.092	0.041
US	0.075	0.126	0.133
VE	0.001	0.005	0.007

2.3 Out-of-sample forecast performance

Given that the concept of stalling is potentially useful for forecasting, we next compare the out-of-sample forecast performance of models with and without stall dummies.[11] We first estimate our stall model on half of the available data, following the procedure outlined in section 2.1. We then construct out-of-sample forecasts at horizons of one to four quarters. Adding one more observation to our sample, we then repeat the process, until the full sample is included.[12]

Clark and West (2007) suggest that an appropriate test statistic to test the null hypothesis of equal out-of-sample predictive power among nested models against the alternative hypothesis that the larger model has greater predictive power is given by

$$MSPE(\text{adjusted}) = P^{-1}\sum(y_{t+\tau} - \hat{y}_{1t,t+\tau})^2 - [P^{-1}\sum(y_{t+\tau} - \hat{y}_{2t,t+\tau})^2 - P^{-1}\sum(\hat{y}_{2t,t+\tau} - \hat{y}_{1t,t+\tau})^2],$$

where $y_{t+\tau}$ is the observation at time $t+\tau$, $\hat{y}_{1t,t+\tau}$ is the τ period ahead forecast made by the parsimonious model (the pure AR(L) model in our case) at time t, $\hat{y}_{2t,t+\tau}$ is the τ period ahead forecast made by the larger model (the AR(L) model plus stall dummies), and P is the number of predictions being examined. To implement this test, we define

$$f_{t+\tau} = (y_{t+\tau} - \hat{y}_{1t,t+\tau})^2 - [(y_{t+\tau} - \hat{y}_{2t,t+\tau})^2 - (\hat{y}_{2t,t+\tau} - \hat{y}_{1t,t+\tau})^2],$$

and then regress $f_{t+\tau}$ on a constant. Clark and West (2007) show that the *t*-test statistic for a constant of zero is approximately normal. They recommend rejecting the null hypothesis of no difference in out-of-sample forecast performance when the test statistic is greater than 1.282 for a one-sided 10% test, and 1.645 for a one-sided 5% test.[13]

[11] These results are based on the maximum stall thresholds; almost identical results are obtained if the minimum stall thresholds are used instead.

[12] In addition to starting with half the sample and then adding one observation at a time until the full sample is included, we considered a number of robustness checks (results available upon request) including varying the initial sample (from first available data point to 1995Q4 or 1980Q1–1995Q4) and excluding the forecast performance during the most volatile periods associated with the International Financial Crisis (by dropping all forecasts for periods after 2007Q4). Results are robust to these alternative specifications.

[13] Our test statistic is based on Newey-West standard errors with $\max\{P/3,1\}$ lags.

Out-of-sample forecast performance – fixed thresholds

Clark and West (2007) MSPE(adjusted) test statistics   Table 4

Code	1-quarter	2-quarters	3-quarters	4-quarters
AR	0.376	0.690	0.237	-0.344
AU	-0.281	0.474	0.109	-0.471
AT	1.710	0.901	-0.139	-0.886
BE	0.321	2.097	1.713	2.192
BR	-0.346	-1.001	-1.112	-0.670
CA	-0.747	1.014	0.360	-0.307
CL	-1.242	-2.070	-2.329	-2.283
CN	0.644	1.306	0.471	1.105
TW	0.258	-0.613	-0.447	-1.470
CO	-0.358	-0.361	0.541	0.662
HR	0.409	-1.398	-1.463	-1.502
CZ	1.200	1.382	1.151	0.410
DK	1.338	1.874	1.307	1.290
EE	-1.306	-1.284	-1.611	-1.978
FI	-0.713	-1.082	-0.488	-0.110
FR	1.007	1.606	1.597	2.299
DE	1.110	0.913	0.252	-0.399
GB	1.049	1.447	1.326	1.366
GR	0.523	1.639	1.027	1.145
HK	1.231	1.231	-0.012	-1.837
HU	-0.218	-0.939	-1.279	-0.194
IN	0.092	1.087	0.399	0.963
ID	-0.658	-0.967	-1.256	-1.150
IE	-0.897	-0.388	-1.225	-1.798
IL	-0.359	-0.018	-0.332	-0.235
IT	1.124	0.131	0.283	0.380
JP	0.818	1.192	1.553	1.928
KR	-0.447	0.275	1.180	1.550
LV	-0.431	-0.411	-0.216	-0.276
LT	0.689	0.785	1.023	0.404
MY	1.541	1.517	1.200	0.955
MX	2.226	2.482	2.652	1.612
NL	0.748	0.739	0.114	1.282
NZ	1.631	1.381	1.634	1.697
NO	-0.176	-0.088	-0.167	-1.130
PE	-0.572	-0.991	-0.689	-0.561
PH	-0.615	-1.320	-1.018	-0.727
PL	0.872	0.673	1.191	0.798
PT	2.044	1.880	1.693	0.546
RO	-0.904	-0.864	-1.290	1.202
SG	1.559	0.900	0.459	0.591
SK	-0.455	0.621	0.700	-0.629
SI	-0.572	-0.843	-0.552	-0.511
ZA	1.249	2.479	1.930	1.291
ES	1.465	-1.388	-0.959	-1.092
SE	0.913	-0.327	-1.713	-1.535
CH	2.242	2.146	2.039	-0.034
TH	-0.775	-0.718	-0.856	-0.656
TR	2.593	2.278	1.632	0.401
US	-0.071	0.297	-0.284	-0.752
VE	-0.724	-0.435	-0.760	-0.806

Out-of-sample forecast performance – fixed thresholds

Clark and West (2007) MSPE(adjusted) test statistics | Table 5

	Stalls				Rebounds			
Code	1-quarter	2-quarters	3-quarters	4-quarters	1-quarter	2-quarters	3-quarters	4-quarters
AR	0.521	1.247	1.356	0.366	1.358	0.652	-0.659	-0.718
AU	-0.281	0.474	0.109	-0.471	-1.088	-1.425	-1.212	-0.884
AT	1.742	1.579	0.702	-0.602	2.249	1.955	0.688	-0.263
BE	2.169	2.157	2.516	2.266	1.055	2.175	1.752	2.083
BR	0.641	0.735	-0.629	-0.699	-0.346	-1.001	-1.112	-0.670
CA	-0.779	0.106	0.022	-1.324	-0.326	2.325	0.692	0.696
CL	0.146	-1.146	-1.398	-1.656	0.630	-0.561	-0.871	-1.602
CN	0.763	1.376	0.512	1.266	-0.076	-0.266	-0.733	-1.909
TW	0.708	-0.367	-0.294	-1.095	0.268	0.199	1.018	0.851
CO	1.111	0.962	1.110	0.847	-0.611	-0.131	0.102	1.484
HR	0.294	-1.698	-1.641	-1.476	-1.348	-1.364	-1.248	-0.605
CZ	1.200	1.382	1.151	0.410	-0.721	-0.267	-0.281	-0.866
DK	1.513	1.146	0.408	-0.451	1.338	1.874	1.307	1.290
EE	-1.130	-1.095	-1.237	-1.659	-1.285	-1.428	-1.702	-1.796
FI	-0.026	0.248	0.412	0.963	0.475	-0.019	0.047	0.574
FR	-1.245	-1.367	-1.040	-0.859	1.007	1.606	1.597	2.299
DE	0.875	0.871	0.432	0.104	1.064	1.451	1.323	-1.048
GB	1.530	1.648	1.644	1.546	-2.276	-1.970	-1.658	-1.860
GR	-1.080	0.536	0.047	-0.519	1.182	1.648	1.317	1.783
HK	1.231	1.231	-0.012	-1.837	1.716	1.659	4.042	4.502
HU	0.217	-0.635	-0.891	-0.219	-0.497	-0.681	-1.001	-1.093
IN	-0.987	0.779	3.277	2.197	0.092	1.087	0.399	0.963
ID	-0.998	-1.242	-0.323	1.127	-0.643	-1.153	-1.319	-1.268
IE	-0.330	-0.699	-1.570	-1.098	-0.376	-0.147	-0.131	-0.723
IL	0.771	-0.422	-0.886	-0.899	-0.359	-0.018	-0.332	-0.235
IT	-0.912	-2.250	-2.992	-4.580	1.124	0.131	0.283	0.380
JP	1.415	0.891	1.296	1.404	-0.473	-0.068	0.238	0.530
KR	0.421	0.420	-0.998	-1.060	-0.447	0.275	1.180	1.550
LV	-1.076	-0.466	-0.006	0.611	0.981	0.934	-0.209	-0.147
LT	0.689	0.785	1.023	0.404	0.395	1.264	2.727	0.638
MY	1.541	1.517	1.200	0.955	0.500	-0.350	-1.382	-1.582
MX	2.226	2.482	2.652	1.612				
NL					0.748	0.739	0.114	1.282
NZ	0.967	1.022	-0.474	-0.469	1.631	1.381	1.634	1.697
NO	0.100	0.298	0.136	-0.106	0.017	0.516	-0.160	0.385
PE	-1.135	-0.144	0.289	0.757	-0.572	-0.991	-0.689	-0.561
PH	-0.615	-1.320	-1.018	-0.727				
PL	0.829	-0.059	-0.177	0.533	0.872	0.673	1.191	0.798
PT	2.085	2.044	2.137	1.536	-0.421	-0.736	-1.135	-1.268
RO	-1.087	-1.069	-1.153	0.893	1.314	1.288	1.066	1.253
SG	1.909	1.162	0.574	0.659	-0.196	-0.604	-0.127	-0.058
SK	-0.276	1.307	1.247	-0.398	-0.558	0.504	-0.286	-0.876
SI	-0.537	-1.160	-0.927	-1.754	-0.572	-0.843	-0.552	-0.511
ZA	-0.370	-0.519	-0.530	-0.655	2.192	2.676	1.973	1.399
ES	0.647	1.341	1.915	0.999	2.327	-1.219	-0.728	-0.860
SE	1.675	1.116	0.070	-0.104	0.741	-0.797	-2.455	-1.850
CH	2.242	2.146	2.039	-0.034	2.010	1.774	1.298	0.624
TH	-0.683	-0.648	-1.244	-1.000	-1.159	-0.721	-0.243	0.124
TR	2.928	2.257	1.610	0.054	-0.808	0.403	-1.915	-1.766
US	-0.071	0.297	-0.284	-0.752	1.452	1.227	1.340	1.223
VE	-0.276	1.319	2.078	2.298	-1.338	-1.882	-1.309	-1.329

Table 4 displays the *MSPE*(adjusted) test statistic for each country. A cursory examination of the table indicates that adding stall dynamics does not generally result in an improvement in out-of-sample forecast performance. At the 5% significance level, in only 21 out of 204 cases considered (shaded in orange in the table) does the model with stall dummy perform better in out-of-sample forecasts. At 10% of the number of cells in the table, this is only twice the number that we would expect to find by chance if stalling merely added random noise to the model.[14] Additionally, in only 110 of the 204 cases (54%) does the addition of the stall threshold improve out-of-sample performance at all. In the other 94 cases (46%), adding a stall threshold results in a deterioration in forecast performance, based on the Clark and West test.

2.4 Separating stalls from rebounds

The results reported in Table 4 are based on selecting the threshold model that, when the growth rate crosses from above, minimises the *unconditional* sum of squared residuals. This includes cases where the effect of crossing the threshold is to increase the future growth rate (a rebound), as well as to decrease the future growth rate (a stall).

To distinguish between these two possible types of threshold effects, we repeat the out-of-sample forecast performance exercise twice more with additional restrictions imposed. We first identify a *stall* threshold as the threshold that minimises the sum of squared residuals among those that are associated with a *decline* in future growth (ie, conditional on the sign of the sum of coefficients on the threshold dummies being *negative*). We then repeat the exercise to identify a *rebound* threshold as the threshold that minimises the sum of squared residuals among those that are associated with an *increase* in future growth (ie, conditional on the sign of the sum of coefficients on the threshold dummies being *positive*).

The results for these exercises are given in Table 5. The first thing to note is that there are a number of gaps in the tables. For the empty cells in the left-hand (right-hand) half of the table, there are insufficient samples for which a stall (rebound) threshold could be identified to yield a meaningful test statistic.[15]

Second, for many economies, the most empirically important type of threshold – in terms of lowering the residual sum of squares by the most – varies over the sample. Table 4 contains 204 elements; 84 (or 41%) of them coincide with the contents of the equivalent cell in Table 5 that is associated with either a stall or a rebound. For the remaining 59% of cases, there is no match between the tables, indicating that the type of threshold that reduces the residual sum of squares by the greatest amount switches at least once over the out-of-sample forecast exercise, as we add additional observations, from stalling to rebounding or vice versa.

Third, stall (rebound) thresholds are associated with an improvement in forecast performance in 55% (48%) of cases based on generating a positive Clark and West (2007) measure, and the remainder with no change (in the case that no stall

[14] At the 10% level, 44 cells (or 22% of the total, shaded in yellow) of the 204 cases are statistically significant.

[15] In the case of some of the missing cells, no stall (rebound) threshold that lowered the residual sum of squares could be identified in all of the samples. In some other cases, a stall (rebound) threshold could be identified for only one sample, resulting in an infinite *MSPE*(adjusted) statistic.

(rebound) thresholds can be identified) or a deterioration. And while the critical values suggested by Clark and West should be interpreted with caution in this non-standard application, they suggest that including a stall (rebound) threshold improves forecast performance at the 10% level in 21% (21%) of cases and at the 5% level in 12% (11%) of cases.

2.5 Summary

Taken together, all these results suggest that there is some evidence that economies stall when we define a stall threshold as time-invariant. For the majority of economies, however, the in-sample evidence is statistically insignificant. Further, a model of rebounding, where a decline in growth rates to below some threshold leads to an increase in future growth, has almost as much empirical support as a model of stalling.

3. Empirical evidence for stalling: time-varying threshold

3.1 Naive estimation

So far, we have assumed that the threshold for stalling is constant over time at some level between 0% and the long-run sample mean of an economy, and have reported weak evidence for stalling. Alternatively, one could think of the stall threshold as a variable that is time-varying and related to the underlying growth rate of potential output in the economy. To explore this possibility, we repeat the above analysis but with the threshold assumed to vary one-for-one with the 40-quarter backward-looking moving average of the growth rate of the economy. Specifically, we assume

$$\gamma_t = \bar{y}_t + x ,$$

where x is a constant between -3% and +3% and $\bar{y}_t = \frac{1}{40}\sum_{i=1}^{40} y_{t-i}$. We then define the stall dummy at time t equal to 1 if the actual growth rate falls from above γ_{t-1} at time $t-1$ to below γ_{t-1} at time t. x is chosen based on a grid search in steps of 0.01 as the value that minimises the sum of squared residuals from the regression, ie $x^* = \operatorname{argmin} SSR(x)$.

The results are given in Table 6. As with the previous section, there is a lot of diversity in the results across countries. For a small set of countries (Croatia, Greece, Romania and Venezuela), we cannot identify any thresholds at all.[16]

[16] Effectively, the available sample is shortened by 10 years relative to that examined in the preceding section by the need to construct the 40-quarter backward-looking moving average of growth rates. For these four economies, no single threshold is crossed at least three times.

Results from estimated stall model – time-varying thresholds

Table 6

Code	Lags (L)	Stall Level x^* Minimum	Stall Level x^* Maximum	Size of stall ($\bar{\delta}$)	T- Test statistic	p-value	Significance
AR	4	2.78	3.00	-1.31	-0.57	0.286	
AU	4	1.12	1.15	1.38	1.46	0.926	
AT	4	-0.52	-0.49	0.25	0.37	0.644	
BE	4	-0.15	-0.15	0.24	0.31	0.623	
BR	4	1.68	1.73	1.57	0.85	0.801	
CA	4	1.44	1.61	2.55	2.45	0.992	
CL	4	-0.76	-0.60	1.84	1.09	0.859	
CN	1	1.23	1.27	-2.57	-1.74	0.047	**
TW	4	1.32	1.36	0.00	0.00	0.499	
CO	4	1.56	2.04	-0.46	-0.18	0.428	
HR	4						
CZ	3	-0.14	0.17	1.04	0.25	0.597	
DK	4	0.21	0.48	6.11	3.46	1.000	
EE	4	1.26	1.31	-0.51	-0.08	0.468	
FI	4	-1.79	-1.69	0.39	0.27	0.604	
FR	4	-2.12	-2.12	-0.12	-0.14	0.446	
DE	4	2.23	2.27	3.53	3.27	0.999	
GB	4	1.17	1.18	1.81	2.17	0.984	
GR	1						
HK	4	-3.00	-2.83	-1.20	-0.45	0.326	
HU	2	-0.56	-0.47	6.49	2.31	0.981	
IN	4	0.43	0.43	1.24	0.93	0.822	
ID	3	-2.68	-2.54	-4.47	-2.04	0.022	**
IE	4	2.05	2.08	3.28	1.88	0.968	
IL	2	0.07	0.16	-6.23	-3.04	0.002	***
IT	4	-2.86	-2.76	2.67	2.48	0.993	
JP	3	-2.68	-2.68	-0.55	-0.45	0.325	
KR	4	1.91	2.02	2.95	1.58	0.941	
LV	4	2.53	2.72	3.93	0.32	0.620	
LT	4	1.04	1.67	-16.89	-2.18	0.026	**
MY	3	1.85	3.00	0.88	0.43	0.663	
MX	3	-3.00	-2.67	-16.07	-3.80	0.000	***
NL	4	-1.96	-1.94	-0.95	-0.73	0.233	
NZ	4	-1.90	-0.98	4.09	2.36	0.990	
NO	4	1.93	1.93	3.36	1.75	0.958	
PE	2	1.54	2.09	-10.16	-3.07	0.001	***
PH	2	-2.07	-1.86	0.02	0.01	0.503	
PL	4	0.35	0.55	0.80	0.38	0.644	
PT	4	1.31	1.40	2.44	2.70	0.996	
RO	2						
SG	4	2.27	2.38	0.11	0.05	0.521	
SK	4	-0.48	-0.16	-2.36	-0.79	0.219	
SI	3	-0.69	-0.65	-1.13	-0.34	0.369	
ZA	4	-0.31	-0.28	0.78	0.89	0.814	
ES	4	-1.90	-1.71	-6.03	-4.87	0.000	***
SE	4	-2.60	-2.56	-8.58	-6.68	0.000	***
CH	4	2.92	2.97	-0.55	-0.42	0.339	
TH	4	-1.62	-1.27	-29.29	-4.82	0.000	***
TR	4	-2.57	-1.75	-35.83	-5.09	0.000	***
US	3	-0.87	-0.87	-1.86	-2.85	0.002	***
VE	4						

p-values from standard distribution and bootstrap – time-varying thresholds

Table 7

Country	Standard	4 lags	8 lags
AR	0.286	0.431	0.366
AU	0.926	0.559	0.381
AT	0.644	0.519	0.275
BE	0.623	0.412	0.322
BR	0.801	0.474	0.329
CA	0.992	0.911	0.861
CL	0.859	0.633	0.598
CN	0.047	0.229	0.205
TW	0.499	0.315	0.220
CO	0.428	0.414	0.314
CZ	0.597	0.342	0.322
DK	1.000	0.983	0.920
EE	0.468	0.390	0.381
FI	0.604	0.561	0.497
FR	0.446	0.357	0.323
DE	0.999	0.986	0.939
GB	0.984	0.642	0.622
HK	0.326	0.441	0.429
HU	0.981	0.751	0.750
IN	0.822	0.438	0.359
ID	0.022	0.132	0.169
IE	0.968	0.696	0.649
IL	0.002	0.012	0.001
IT	0.993	0.666	0.680
JP	0.325	0.314	0.283
KR	0.941	0.811	0.664
LV	0.620	0.305	0.308
LT	0.026	0.046	0.031
MY	0.663	0.685	0.659
MX	0.000	0.007	0.003
NL	0.233	0.235	0.173
NZ	0.990	0.913	0.901
NO	0.958	0.771	0.644
PE	0.001	0.019	0.007
PH	0.503	0.508	0.718
PL	0.644	0.519	0.517
PT	0.996	0.734	0.757
SG	0.521	0.340	0.244
SK	0.219	0.382	0.402
SI	0.369	0.442	0.378
ZA	0.814	0.283	0.306
ES	0.000	0.000	0.000
SE	0.000	0.000	0.000
CH	0.339	0.499	0.449
TH	0.000	0.002	0.001
TR	0.000	0.000	0.000
US	0.002	0.029	0.028

In approximately half of the remaining cases, the threshold that reduces the sum of squared residuals by the most is associated with a positive (rebound), rather than negative (stall), change in the growth rate in future periods: $\bar{\delta} > 0$. For eight economies the increase is statistically significant at standard critical values.[17] And the threshold that appears to trigger this increase in growth varies from being well above the 40-quarter moving average growth rate (Canada, Germany, Great Britain and Portugal), close to the moving average growth rate (Denmark and Hungary) and well below it (Italy and New Zealand). For these economies, threshold effects may be useful for modelling GDP, but they are not necessarily consistent with the idea of stalling.

Moving on to the economies where a threshold is associated with a decline in growth rates, there are 11 such economies with statistically significant stalls at the 5% level (or 8 at the 1% level). For the majority (Indonesia, Mexico, Spain, Sweden, Thailand, Turkey and the United States), the associated stall threshold is well below the moving average growth rate. The size of the subsequent cumulative four-quarter fall in GDP following a stall, given by $\overline{\delta}$, varies between 1.9% for the United States and 10–20% for Lithuania, Mexico and Peru. For Thailand and Turkey, the subsequent decline in growth rates is very large, as the threshold dynamics for these economies capture the historical decent into crises.

Comparing our results with those obtained earlier using a fixed threshold, only four economies stall significantly (at the 5% level) in both cases: Indonesia, Lithuania, Mexico and Turkey. Clearly evidence for stalling is sensitive to the precise definition used and the data sample.[18]

3.2 Bootstrap correction

We again correct the standard errors based on a sieve bootstrap, using the same procedure as before, and report the results in Table 7.[19] Here the results are relatively robust: with the exception of China and Indonesia, stallers identified by the standard critical values are also identified as stallers based on the bootstrap exercise. Further, there is generally less discrepancy between the bootstrap and the standard p-values based on the time-varying threshold than the fixed threshold models.

3.3 Out-of-sample forecast performance

Next, we examine the out-of-sample forecast performance, and report the results for the MSPE(adjusted) test statistics in Table 8 at horizons of one to four quarters.[20]

[17] The p-values in Table 5 are for a one-sided test of $\overline{\delta} > 0$. A sufficiently large p-value in the table (eg >0.95) indicates significant evidence of an increase in growth rates if the economy passes through the threshold.

[18] The discrepancy between the two sets of results is partly down to the different sample size, as our definition of stalling here effectively shortens the sample size by 10 years. To show the effect of this, Table A1 in the appendix repeats the analysis reported in Table 2 above, but with the first 40 observations dropped, so that the sample matches that used to generate Table 6. Focusing on this shortened sample, approximately half of the economies that experience a significant stall by each definition also stall according to the alternative. This intersecting set of stalling economies by both definitions (at the 5% level) is made up of Mexico, Peru, Spain, Sweden, Thailand and Turkey.

[19] Economies for which no stalls were identified are excluded from Tables 7 and 8.

[20] In Table 8 we additionally exclude Hungary, Latvia and Poland due to their short samples and the consequent lack of out-of-sample periods to examine.

Here the evidence suggests that including thresholds can generally improve the out-of sample forecast performance. In 139 out of 176 cells in the table (79%), a positive Clark and West (2007) test statistic suggests that the threshold model results in better out-of-sample forecast performance than the simple autoregressive model. And in 27% (41%) of all cases, this difference is statistically significant at the 5% (10%) level, shaded in orange (yellow). These percentages are all higher than the equivalent numbers reported for fixed thresholds in Section 2.3.

Out-of-sample forecast performance – time-varying thresholds

Clark and West (2007) MSPE(adjusted) test statistics

Table 8

Code	1-quarter	2-quarters	3-quarters	4-quarters
AR	2.354	2.622	1.200	0.173
AU	0.091	0.706	1.045	0.737
AT	2.193	2.893	2.630	2.873
BE	2.562	2.817	2.530	2.444
BR	-0.684	-1.715	-1.736	-1.578
CA	2.377	2.218	2.353	2.316
CL	-0.020	-0.401	-0.282	-0.638
CN	1.013	-0.945	1.941	1.407
TW	1.967	3.257	3.939	5.467
CO	-0.161	-1.855	0.805	-1.129
CZ	3.938	3.052	2.711	3.273
DK	-0.732	0.350	1.563	1.150
EE	1.078	1.231	0.881	1.485
FI	-0.566	-0.797	-2.275	-1.802
FR	3.138	2.568	2.558	3.476
DE	1.236	0.862	1.103	1.374
GB	-0.412	1.306	2.927	2.178
HK	1.602	1.931	1.658	2.799
IN	0.816	0.713	1.345	1.964
ID	0.444	0.083	-0.318	-0.905
IE	0.810	0.492	0.177	-0.167
IL	-0.211	-1.069	-1.623	-1.159
IT	0.593	0.709	2.769	2.185
JP	2.113	1.076	1.381	1.554
KR	2.247	1.495	1.395	1.019
LT	1.745	0.939	0.066	0.477
MY	0.795	-0.467	1.474	2.160
MX	0.306	0.027	-0.183	-0.422
NL	0.376	0.294	1.176	1.556
NZ	1.339	0.499	0.822	1.430
NO	0.430	0.328	-0.617	0.895
PE	0.917	2.920	2.322	3.077
PH	1.331	0.953	2.349	0.253
PT	1.329	0.487	0.273	0.569
SG	-0.813	0.434	-0.133	0.112
SK	1.404	0.816	0.600	-0.217
SI	-0.882	-1.055	1.055	-1.055
ZA	2.854	1.531	1.276	1.303
ES	1.270	1.818	1.642	1.968
SE	1.389	1.258	1.315	1.468
CH	1.491	1.539	0.532	-0.336
TH	0.334	0.736	0.953	1.055
TR	0.412	0.490	0.485	0.911
US	-0.823	0.051	0.774	1.886

3.4 Separating stalls from rebounds

On the face of it, the results reported in Table 8 look much more favourable to the time-varying threshold model than those reported in Table 4 based on fixed thresholds. However, they are subject to the same concern outlined earlier that they include cases where slowing growth is followed by either an increase or a decrease in future growth.

Out-of-sample forecast performance – time-varying thresholds

Clark and West (2007) MSPE(adjusted) test statistics — Table 9

	Stalls				Rebounds			
Code	1-quarter	2-quarters	3-quarters	4-quarters	1-quarter	2-quarters	3-quarters	4-quarters
AR	2.948	2.672	1.305	0.352	1.801	1.944	1.460	1.123
AU	-0.415	0.373	0.786	0.256	0.812	1.661	2.043	2.054
AT	2.588	2.864	2.729	3.443	1.784	2.612	2.646	2.937
BE	2.308	2.268	2.578	2.506	2.878	2.169	2.541	2.261
BR	-1.782	-1.816	-1.279	-0.323	-0.304	-1.221	-1.283	-1.122
CA	0.468	0.633	1.623	1.719	2.377	2.218	2.353	2.316
CL	0.170	-0.205	-0.152	-0.337	-0.301	-0.655	-0.758	-0.925
CN	1.054	-1.064	1.137	-0.768	1.057	-0.690	1.383	1.716
TW	1.606	1.968	1.779	2.329	2.244	3.484	4.007	4.965
CO	0.493	-1.174	1.246	-1.150	-0.227	-0.434	-0.513	0.947
CZ					6.615	2.399	2.147	3.061
DK	0.298	-0.370	-0.806	-1.673	-0.732	0.350	1.563	1.150
EE					1.136	1.230	1.185	1.194
FI	0.279	0.237	-1.418	-1.865	0.073	-0.625	-2.284	-1.905
FR	2.650	2.234	1.988	2.934	4.127	3.689	3.619	4.176
DE	1.028	1.041	0.766	1.044	1.006	0.600	0.865	0.911
GB	0.126	-0.612	-0.519	-0.279	0.388	1.636	2.978	2.348
HK	2.177	2.447	2.576	2.618	1.617	1.835	1.559	2.775
IN	0.015	0.091	2.328	2.693	0.816	0.713	1.345	1.964
ID	-1.543	-1.493	-1.384	-1.271	0.428	0.125	-0.306	-0.883
IE	-0.491	-0.770	-0.583	0.269	0.810	0.492	0.177	-0.167
IL	-0.020	-1.238	-0.690	1.815	1.039	1.039	1.039	-0.765
IT	0.771	2.191	3.762	4.361	1.166	1.224	3.049	2.885
JP	2.623	4.540	3.391	4.236	0.823	-0.506	1.371	1.109
KR	1.947	0.870	0.719	0.213	2.433	1.761	1.026	0.150
LT	-1.659	-1.341	-1.227	-1.232	2.939	3.410	0.639	0.887
MY	1.069	1.063	1.688	2.269	-1.085	-1.226	1.119	1.857
MX	1.202	1.035	0.542	-0.129	-1.016	-1.157	-1.181	-0.886
NL	-0.175	1.257	1.391	1.464	0.122	0.214	1.199	1.466
NZ	1.362	0.489	-1.074	-1.383	1.339	0.499	0.822	1.430
NO	0.245	0.737	-0.537	0.360	0.133	-0.244	-0.093	0.359
PE	2.143	3.355	2.761	3.212	0.697	2.563	2.965	3.423
PH	2.159	1.743	2.176	1.568	1.350	0.527	2.561	0.104
PT	1.121	1.154	1.198	0.868	2.000	1.629	1.771	3.181
SG	0.247	1.039	1.270	0.917	-1.029	-0.895	-0.473	0.019
SK	1.512	1.024	0.705	-0.145	1.605	1.265	0.905	-0.133
SI	0.826	0.945	-0.290	-0.814	-0.854	-1.054	1.055	-1.055
ZA	1.988	0.958	0.914	0.888	3.480	2.187	1.868	1.928
ES	1.397	2.061	1.639	2.019	-0.981	0.206	0.064	0.692
SE	1.389	1.258	1.315	1.468	2.208	2.008	0.335	-0.502
CH	1.491	1.539	0.532	-0.336	1.219	1.095	1.117	1.254
TH	0.401	0.663	0.891	1.007	-0.021	-0.898	-0.032	0.253
TR	0.566	0.775	0.646	1.361	-1.154	-2.691	-3.051	-1.928
US	0.226	0.657	0.709	0.985	0.163	1.170	1.075	1.046

We next distinguish between the two possible types of thresholds, and repeat the analysis on out-of-sample forecasts using the same methodology outlined in Section 2.4. The results are given in Table 9. As before, there are a number of cases where cells in the left-hand (right-hand) panels are blank, where no stall (rebound) thresholds are identified.

More importantly, for nearly all economies, the type of threshold varies over the sample, even more so than when we examined fixed thresholds. Table 8 contains 176 elements. Only 28 (or 16%) of them coincide with the contents of either the relevant stall or rebound cells in Table 9. In the other 84% of cases, the type of threshold that minimises the unconditional residual sum of squares switches at least once over the out-of-sample forecast exercise, as we add additional observations, from stalling to rebounding or vice versa.

However, in general, adding either type of threshold results in an improvement to out-of-sample forecast performance. Stall (rebound) thresholds are associated with an improvement in forecast performance in 71% (74%) of cases, based on generating a positive Clark and West (2007) measure. Again, taking the Clark and West critical values as indicative, a stall (rebound) threshold improves forecast performance at the 10% level in around 35% (39%) of cases, shaded in yellow, and at the 5% level in 26% (31%) of cases, shaded in orange.

3.5 Summary

Taken together, all these results suggest that there are important threshold effects in the behaviour of GDP relative to lagged growth rates, and that these may be useful for improving out-of-sample forecast performance. However, the idea of a stall – a slowdown in growth to below some threshold level – is not the only relevant threshold for helping to forecast GDP growth. Based on our analysis, for many economies, an equally important regularity appears to be that a slowdown to below some threshold is followed by higher growth rates in the coming quarters, something we have labelled a rebound.

Curiously, these two phenomena, of stalls and rebounds, are not mutually exclusive. For 14 economies, there is at least one forecast horizon for which including either a stall or a rebound threshold results in a significant improvement in out-of-sample forecast performance at the 5% level. Further, focusing on the US economy, we find stronger evidence for rebounds than for stalls based on our out-of-sample forecast performance. We leave a more careful examination of the different types of possible threshold effects for modelling and forecasting GDP for future study.

Conclusions

A "stalling" economy has been defined as one that experiences a discrete deterioration in economic performance if its growth rate slows to below some threshold value. Conceptually, the idea of a stall could be very useful for modelling and forecasting purposes, but it lacks a theoretical foundation, or any basis on which to determine the growth rate at which the economy stalls ex ante.

Given the lack of theoretical foundation, we consider two different definitions of a stall threshold: a time-invariant, fixed level, or one that varies with the

40-quarter backward-looking moving average growth rate. Based on these measures, we estimate models that incorporate stalling dynamics on a panel of 51 economies.

To assess the importance of stalling, we then use a bootstrap procedure to estimate the statistical significance of stalling thresholds in-sample, and forecasts for one to four quarters ahead, to assess the ability of models incorporating stalling to predict growth rates out-of-sample.

Overall we find limited evidence in favour of fixed stalling thresholds. In contrast, models that incorporate time-varying threshold effects, with the threshold level varying with the moving average growth rate, generally forecast future growth rates better than models that exclude threshold effects. However, stalls do not appear to be the only such threshold effects at work in our panel. Rebounds – higher growth rates that follow a slowdown in growth rates – generally improve out-of-sample forecast performance in our panel to an even higher degree.

References

Clark, Todd E and Kenneth West (2007): "Approximately normal tests for equal predictive accuracy in nested models", *Journal of Econometrics*, 138(1), 291–311.

Hansen, Bruce E (1996): "Inference when a nuisance parameter is not identified under the null hypothesis", *Econometrica*, 64(1), 413–430.

Hansen, Bruce E (1997): "Inference in TAR models", *Studies in Nonlinear Dynamics and Econometrics*, 2(1), 1–14.

Hansen, Bruce E (2000): "Sample splitting and threshold estimation", *Econometrica*, 68(3), 575–603.

Ho, Wai-Yip Alex and James Yetman (2012): "Does US GDP stall?", *BIS Working Papers*, no 387, September.

MacKinnon, James G (2006): "Bootstrap methods in econometrics", *The Economic Record*, 82 (special issue), S2–S18.

Nalewaik, Jeremy J (2011): "Forecasting recessions using stall speed", *Finance and Economics Discussion Series*, Federal Reserve Board, 2011-24.

Sheets, Nathan and Robert A Sockin (2012): "Stall speeds and spillovers: some new evidence for the global economy", *Citi Empirical and Thematic Perspectives*, 6 September.

Sheets, Nathan (2011): "Is the US economy approaching 'stall speed'?", *Citi Empirical and Thematic Perspectives*, 11 November.

Results from estimated stall model – fixed thresholds

(first 40 observations dropped)

Table A1

Code	Lags (L)	Stall Level γ^* Minimum	Maximum	Size of stall ($\bar{\delta}$)	t- test statistic	p-value	Significance
AR	2	2.16	2.65	0.14	0.04	0.518	
AU	4	0.00	0.12	-5.39	-2.76	0.003	***
AT	4	0.86	0.87	0.18	0.15	0.559	
BE	3	1.17	1.17	0.41	0.52	0.700	
BR	4	3.32	3.40	0.59	0.35	0.637	
CA	4	1.42	1.45	1.53	1.65	0.949	
CL	2	0.86	1.38	-0.69	-0.21	0.418	
CN	4	8.88	9.03	2.94	1.05	0.847	
TW	4	6.53	6.54	-4.00	-2.98	0.002	***
CO	1	4.77	4.78	-3.49	-1.41	0.088	**
HR	1	0.00	0.00	-1.31	-0.33	0.376	
CZ	2						
DK	4	1.93	1.93	3.40	2.01	0.976	
EE	3						
FI	4	0.00	0.07	-10.33	-3.83	0.000	***
FR	4	0.00	0.37	0.69	0.82	0.794	
DE	4	2.06	2.07	-0.86	-1.08	0.141	
GB	4	0.00	0.22	-1.88	-1.44	0.075	**
GR	4						
HK	4	4.40	4.56	-1.83	-0.75	0.227	
HU	2	0.00	0.15	0.96	0.24	0.593	
IN	4	4.99	5.03	0.77	0.52	0.700	
ID	2	4.11	4.13	-2.61	-1.24	0.109	
IE	4	2.21	2.87	3.72	1.59	0.942	
IL	3	3.29	3.29	-1.52	-0.57	0.287	
IT	2	1.06	1.16	0.48	0.71	0.760	
JP	3	0.92	1.02	-0.93	-0.73	0.234	
KR	4	0.00	1.69	2.12	0.64	0.740	
LV	3						
LT	1						
MY	2	4.10	4.32	-3.10	-1.04	0.152	
MX	3	0.00	0.16	-14.19	-3.31	0.001	***
NL	4	2.17	2.22	0.23	0.28	0.612	
NZ	1	2.25	2.27	2.81	1.95	0.973	
NO	4	1.80	1.81	2.27	1.60	0.944	
PE	3	4.62	4.62	-11.79	-2.86	0.003	***
PH	3	1.64	2.21	-1.64	-0.76	0.225	
PL	1	4.18	4.60	-0.24	-0.11	0.455	
PT	4	0.51	0.53	-3.13	-2.05	0.021	**
RO	3						
SG	4	4.25	4.30	-5.38	-1.42	0.080	*
SK	4	4.03	4.25	-1.43	-0.55	0.295	
SI	2	2.22	2.25	-5.27	-1.59	0.062	*
ZA	4	1.24	1.26	0.53	0.69	0.755	
ES	4	1.06	1.14	-6.02	-4.69	0.000	***
SE	4	0.52	0.52	-3.85	-2.97	0.002	***
CH	3	1.76	1.78	0.76	0.81	0.791	
TH	1	3.06	3.16	-24.73	-3.29	0.002	***
TR	2	2.63	2.66	-26.93	-4.23	0.000	**
US	3	0.12	0.18	-1.26	-1.28	0.101	
VE	4						