Macroeconomic Dynamics, **25**, 2021, 1126–1143. Printed in the United States of America.
doi:10.1017/S1365100519000592

A NOTE ON MEASURING US TIME SERIES VOLATILITY DURING THE GREAT MODERATION

Isaiah Hull
Research Division, Sveriges Riksbank

We identify volatility breaks in all testable series in the FRED database over the 1957–2013 period. This yields 17,681 breaks, which we categorize using text analysis. We show that 70.5% of series categories experienced a decline in volatility over the 1985–1999 period, suggesting that the Great Moderation was far broader in scope than the literature has documented. We also show that this decline reversed in 2000, leading to a sharp increase in volatility for most time series categories; however, this did not fully materialize in GDP volatility until the Great Recession. Finally, we identify labor markets, demographics, finance, and government debt as potential drivers of low-frequency shifts in volatility over the 1957–2013 period.

Keywords: Great Moderation, Macroeconomics, Business Cycles, Volatility

1. INTRODUCTION

Kim and Nelson (1999), McConnell and Perez-Quiros (2000), and Blanchard and Simon (2001) first document the post-1984 reduction in US GDP volatility, which was later termed the "Great Moderation" by Stock and Watson (2002). Three hypotheses were proposed for this decline: (1) better policy; (2) improved business practices; and (3) smaller shocks.[1] Additional work generalized this decline to many macrovariables (Stock and Watson (2002), Sensier and van Dijk (2004)) and many countries (Blanchard and Simon (2001), Stock and Watson (2002)).

The 2007–2009 Great Recession, which was accompanied by an increase in the volatility of GDP and its components, renewed interest in the Great Moderation. Recent work explores whether the Great Moderation ended, when it ended, and why it ended.[2] While there is broad agreement that the Great Moderation paused during the Great Recession, disagreement remains over whether it paused before 2007 and whether it resumed after 2009. Some (e.g. Clark (2009)) argue that the economy's default regime is characterized by modest volatility, occasionally

I would like to thank Gaetano Antinolfi, Martin D. Evans, Edward Kung, Andres Liberman, Jesper Lindé, Alexander Ludwig, George Skiadopoulos, Hiroatsu Tanaka, Robert Vigfusson, and Xin Zhang, as well as participants at the GSMG, CEF, and SNDE for their comments and suggestions. The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank. Address correspondence to: Isaiah Hull, Research Division, Sveriges Riksbank, SE-103 37, Stockholm, Sweden. e-mail: isaiah.hull@riksbank.se. Phone: +46 076 589 0661. Fax: +46 80 821 0531.

 1469-8056/19

interrupted by periods of high volatility, such as the Great Recession. Others (e.g. Stock and Watson (2012)) extend this claim by arguing that the temporary departure from the Great Moderation in 2007–2009 was brought on by the realization of larger versions of the same types of shocks that were also present during the Great Moderation. Still, others (e.g. Carvalho and Gabaix (2013)) argue that a shift in GDP's composition toward higher volatility sectors, such as finance, reduced stability prior to the onset of the Great Recession. Relatedly, Grydaki and Bezemer (2013) suggest that output volatility may have fallen during the Great Moderation precisely because of a credit expansion that enabled spending out of wealth. They also suggest that the same expansion may explain the increase in volatility during the Great Recession.[3]

Most empirical work on the Great Moderation examines volatility breaks in selected time series, such as GDP, GDP components, oil prices, and interest rates. Some work also applies more systematic analysis to a larger number of series (Stock and Watson (2002), Sensier and van Dijk (2004)). And Carvalho and Gabaix (2013) examine the Great Moderation through a different lens by constructing a measure of "fundamental volatility" with deep theoretical foundations. This paper contributes to the literature by taking a different methodological approach to testing for volatility breaks. Rather than examining only GDP components or a selected set of time series with ex ante assumed relevance, we instead start with the universe of aggregate US time series, drawn from the St. Louis FRED's database. We then apply selection criteria to identify a testable subset of 9,355 series, perform Bai and Perron (2003) break tests on their volatilities, and then use text analysis to characterize the nature of the breaks.

Our primary measure of volatility is the proportion of error variance breaks that occurred in a given period. We evaluate the validity of this approach by reproducing the findings that established the Great Moderation literature. In particular, we show that the frequency of volatility breaks declines for GDP and GDP component series during both the 1985–2013 and 1985–1999 periods, relative to the 1957–1983 period. This includes the most frequently tested series in the Great Moderation literature:[4] government spending, investment, inventories, and durable goods. In addition to this, we reconstruct the measures of GDP volatility introduced in Stock and Watson (2002), Davis and Kahn (2008), and Carvalho and Gabaix (2013) and demonstrate consistency with the GDP volatility measure produced in this paper. This suggests that our approach captures volatility breaks well; and may reasonably be applied to describe the behavior of a much larger range of series for which we have no direct comparison.

We next document structural change in the volatility of GDP and its components over several subperiods of the Great Moderation. The decline in break frequency is most pronounced during the 1985–1999 period, which we will henceforth refer to as the "narrow" Great Moderation period. To the contrary, the period that immediately follows (2000–2013) actually witnesses a mild, but statistically insignificant increase in the intensity of GDP and GDP component breaks. This, however, is primarily generated by an increase in break intensity during and after the Great Recession. The period between 2000 and 2006 also experiences a decline in intensity relative to 1957–1983, but an increase relative to the narrow Great Moderation period. This finding complements Carvalho and Gabaix (2013), which suggests that the expansion of the financial sector began to push up fundamental volatility in 2000. Importantly, Carvalho and Gabaix (2013) assume fixed within-sector volatility and demonstrate that the share of high-volatility sectors in GDP grew; whereas, we show that within-sector volatility itself grew.

Next, we demonstrate that the Great Moderation was a more general phenomenon than the literature has shown. We do this by dividing the set of structural breaks into categories by title similarity. We show that 70.5% of series categories experienced a decline in the intensity of structural change during the 1985–1999 period. This, however, reverses during the 2000–2006 period, immediately preceding the Great Recession. Here, the average series category actually sees an increase in volatility, which intensifies further during the Great Recession, leading to an increase in the frequency of structural breaks for 71% of series categories between 2000 and 2013. The increase in volatility over the 2000–2006 period is slightly larger than the corresponding decline during the narrow Great Moderation period. Furthermore, if we include the Great Recession and post-Great Recession period, this flips the usual Great Moderation result: for the average volatility series category, the intensity of structural change increases slightly over the 1985–2013 period, relative to the 1957–1983 period. We also show that breaks in the persistence of volatility tended to fall during the Great Moderation and rise during the period that followed.

We add to our core findings by examining the z-scores of volatility changes in the period immediately following each structural break. This allows us to identify the direction of each break. We again divide the data into categories, but this time use the FRED system to identify a smaller set of groups for the purpose of interpretability. We find statistically significant declines in the intensity of breaks in 65% of series categories over the broad Great Moderation period. Furthermore, we show that only two broad series categories experienced statistically significant increases in volatility over the broad Great Moderation period: federal government debt and finance companies. The significance and magnitude of these results suggest that these series categories may have been the underlying drivers of the broader increase in volatility that began in 2000 and spilled over into GDP in 2007; however, we refrain from making strong statements about causality, since our analysis is restricted to the examination of univariate time series volatilities.

Moreover, we show that some of the largest declines in volatility come from demographic and employment series categories. These changes persist even during the Great Recession, which suggests that structural change in demographics and labor markets may be long-term sources of volatility reduction. This complements the work by Jaimovich and Siu (2009), Lugaeur and Redmond (2012), and Lugaeur (2012b), which finds that an increase in the volatile-age share of the labor force partially explains the decline in output volatility during the Great Moderation.[5] Importantly, however, we do not repeat the empirical

https://doi.org/10.1017/S1365100519000592 Published online by Cambridge University Press

exercise performed in this literature—and, thus, are not subject to the criticism in Everaert and Vierke (2016)—but instead demonstrate that there were broad, sharp declines in volatility across demographic and labor market time series between 1957 and 2013.

Finally, while we do not explicitly test the three proposed explanations for the Great Moderation, our findings suggest that good luck (i.e. smaller shocks) in the form of demographic change may have reduced volatility during the Great Moderation. Furthermore, we do not find evidence for the traditional "good monetary policy" story, but our results do suggest that fiscal and financial regulatory policy may have contributed to the post-2000 volatility increase. We also show that inventory series tended to experience volatility reductions during the Great Moderation, which supports the improved business practices narrative; however, we do not quantify its importance relative to other potential causes.

The remainder of the paper is structured as follows: Section 2 describes the data which were obtained from the St. Louis Fed's FRED system. Section 3 describes the text analysis and econometric methods that were used to perform tests and organize results. Section 4 provides a description and discussion of the results. Section 5 concludes.

2. DATA

As of August 2016, the St. Louis Fed's FRED database contained 391,000 time series, which were divided into eight broad series categories: (1) US regional data; (2) international data; (3) population, employment, and labor market data; (4) national accounts; (5) academic data; (6) price data; (7) production and business activity data; and (8) money, banking, and finance data. While some series are constructed by the St. Louis Fed itself, most are contributed by 78 institutional partners, including the US Census Bureau, the US Bureau of Labor Statistics, the Organization for Economic Co-operation and Development, the World Bank, and the Federal Reserve Board.

2.1. Series

For the purposes of feasibility and interpretability, we limit the set of series included in our analysis. We start with the entire St. Louis Fed's FRED database, which contained 391,000 series at the date of collection, and then remove the following categories of series: (1) local, regional, and international series;[6] (2) high-frequency series; (3) categorical or descriptive series; and (4) incomplete series. Our intent is to exclusively retain aggregate US time series that are suitable for break testing and are not available at a lower frequency.

Of the original 391,000 series, we identify 9,355 that are suitable for break testing. Within the 9,355 series, 4,195 are annual, 3,212 are quarterly, and 1,948 are monthly. We organize series into groups using a three-step procedure. First, we construct a text corpus from the titles. We do this by transforming all letters to lowercase, and removing all punctuation and special characters. We then filter common words out of the corpus, such as articles and prepositions. We also remove terms related to measurement and economics, such as "gross," "real," "percent," "change," and "contributions."[7] Finally, we segment series into groups using three topic modeling algorithms: bag-of-words (BoW), latent Dirichlet allocation (LDA), and non-negative matrix factorization (NMF). Most exercises use the BoW algorithm, but the LDA and NMF algorithms are applied in a set of robustness tests.

The BoW algorithm is a simple model that identifies topics by computing the frequency with which each word is used in the title corpus. If, for instance, investment appears more frequently than inflation, then investment will appear higher on the list of topics identified using the BoW model. LDA is a clustering algorithm that uncovers a latent distribution of topics in a text corpus. It does this through the use of an iterative probabilistic model, where each word in each title is assigned a probability of belonging to a latent topic, and each topic is assigned a probability of describing a given title. Finally, NMF is a linear algebraic method for topic modeling that relies on the factorization of the document-term matrix to generate clusters.[8]

Figure A1, which is located in Appendix, shows the 25 most commonly used words in series titles, as identified by the BoW algorithm. Note that the best represented series are related to GDP and its components, including consumption, government expenditures, and investment. Price indices, real estate and construction series, population series, interest rate series, demographic series, labor market series, commodities, and goods series are also among the set of most common topics.

2.2. Breaks

We next describe how series were transformed and how breaks were measured, but defer a more detailed discussion of break testing to Section 3. Across all 9,355 series, our tests returned 17,681 breaks, yielding an average of 1.89 per series. This number does not appear to increase with series frequency, as quarterly series have fewer breaks (1.70) on average than annual series (1.82); however, the proportion of series with no breaks is highest for annual series and lowest for monthly series. Only 10.37% of series had no breaks in error variance over the 1957–2013 period. Overall, this suggests that breaks are likely to be well identified, since there is no systematic relationship with series frequency.

Each break test allows for up to 10 breaks, which appears to be sufficient, since only 2 series have 10 breaks. Most series have at least one break and very few series have more than five breaks. The maximum break date for all break order groups is 2013,[9] but the mean and minimum increase monotonically in the break order. The first break for most series predates the start of the Great Moderation.

3. METHODS

Since break tests are performed on 9,355 series, we take a conservative approach to transforming the data to ensure stationarity and interpretability. We follow a procedure that is similar to the one outlined in Stock and Watson (2002), which is specified below:

1. **Percentage change series.** If a series title contains the words "percentage change," we do not make any transformations to it in this stage.
2. **Percent series.** If a series title contains the word "percent," but not "percentage change," we take the first difference of the series.
3. **Positive, non-percent series.** If a series is always positive and does not contain the word "percent," we take the log of the series and then the first difference.
4. **Negative, non-percent series.** All remaining series are differenced until they are stationary.

This procedure leaves us with three types of series: (1) growth rates; (2) changes in percentage points; and (3) n-differenced, stationary series. Finally, we deseasonalize all of the transformed series to remove recurring components that might generate spurious breaks.

We follow McConnell and Perez-Quiros (2000) and assume that the transformed series follow autoregressive processes. To account for changes in the mean and persistence over time, we employ a time-varying parameters regression. Following the literature,[10] we focus on breaks in the error variance, rather than on breaks in the autoregressive parameters; however, we do briefly consider breaks in the persistence of volatility in an extension.

For each series, we run the following time-varying parameters regression:

$$y_t = \mu(t) + \alpha(t) y_{t-1} + \epsilon_t \tag{1}$$

In equation (1), y_t is the post-transformation, stationary series. Notice that the mean, $\mu(t)$, and autoregressive parameter, $\alpha(t)$, are permitted to vary over time. We recover estimates of both using Kalman filtering and smoothing. We then use the estimated values to recover the residual series:

$$\hat{\epsilon}_t = y_t - \mu(t) - \alpha(t) y_{t-1} \tag{2}$$

As in McConnell and Perez-Quiros (2000) and Davidian and Carroll (1987), we use the unbiased absolute value estimator for the standard deviation of ϵ_t:

$$\hat{\sigma}_{\epsilon_t} = \sqrt{\frac{\pi}{2}} |\hat{\epsilon}_t| \tag{3}$$

We next perform tests on a regression of the following form:[11]

$$\hat{\sigma}_{\epsilon_t} = \gamma + u_t \tag{4}$$

https://doi.org/10.1017/S1365100519000592 Published online by Cambridge University Press

We also extend our results by adopting a specification that allows for persistence in volatility over time. Here, we test for breaks in the autoregressive parameter:

$$\hat{\sigma}_{\epsilon_t} = \gamma + \rho \hat{\sigma}_{\epsilon_{t-1}} + u_t \tag{5}$$

Since we do not have a prior about any of the break dates and cannot inspect all series visually, we use the Bai and Perron (2003) test, as implemented by Hornik et al. (2003), which does not require a known break date. This test also permits the identification of multiple breaks in a single series. As indicated in the previous section, some series contain as many as 10 structural breaks. Forcing structural breaks to compete would leave critical periods of structural change undetected.

The output from this exercise is a set of 0–10 break dates for the 1957–2013 period for each of the 9,355 series tested. Each is best interpreted as a break in the volatility of the series. In the following section, we will perform tests on these break series to obtain our main results.

We first use the approach developed in this paper to verify the original Great Moderation results for GDP, GDP's components, and the most frequently tested series in the literature. There are two purposes behind this exercise. First, it provides further confirmation of the original Great Moderation results; and second, it demonstrates that the approach used in this paper yields results that are consistent with established methods.

In Section 3, we described the procedure used to compute volatility breaks. We now use text analysis to categorize those breaks. We will start by identifying series that (1) have a title that contains the words "gross," "real," "domestic," and "product"; and (2) do not have a title that contains the word "price." Note that the first criterion identifies all real GDP and real GDP component series. The second criterion filters out all price index series, leaving only breaks that are related to quantities.

Next, we use the bag-of-words algorithm to identify the 50 most commonly used words in the titles of remaining series (e.g. investment, consumption, etc.). We then create a series of breaks for each such word; and normalize each series by the total number of breaks in the series. This gives us the proportion of all breaks associated with a frequently used word that occurred in a given year between 1957 and 2013.

Figure A2, which is located in Appendix, shows the series we identify as real GDP volatility breaks, along with three comparison measures for the 1960–2013 period. The other measures are as follows: (1) fundamental volatility (Carvalho and Gabaix (2013)); (2) 5-year rolling standard deviations (e.g. Davis and Kahn (2008)); and (3) autoregressive process residuals (e.g. Stock and Watson (2002)). Each series is demeaned and standardized. Note that all four measures—including the one we introduce—closely comove over the sample period. Each sees a decline from 1960 to 1970, an increase from 1970 to 1984, and a persistent decrease during the narrow Great Moderation period. Note that the Carvalho and Gabaix (2013) series ends in 2008. Other measures of GDP volatility, including the break proportion method that we introduce, increased during this period.

4. RESULTS

In addition to recovering series categories via topic modeling, we also separately perform substring searches on titles to identify series that belong to commonly tested groups. In particular, we try to identify the sets of breaks associated with the following GDP components, which are the most frequently tested in the Great Moderation literature: (1) government spending; (2) consumption; (3) investment; (4) inventories; and (5) durable goods. We repeat the same normalization procedure and use the same sample period.

After we reestablish the results from the Great Moderation literature, we perform the same procedure, but for an expanded sample that contains series of all category types. That is, we do not limit ourselves to GDP and GDP component series, but instead use the BoW, LDA, and NMF algorithms to extract the 400 most common topics from the series titles. We then use the same normalization procedure as we did for the GDP and GDP component breaks. Note that using text analysis algorithms allows us to divide the universe of available time series into exhaustive and arbitrarily narrow groupings. This is particularly useful in exercises where we wish to measure changes in the distribution of all time series volatilities, but do not necessarily want to identify individual series shifts in volatility.

We now have the following: (1) a real GDP series; (2) GDP component series identified via text analysis; (3) targeted series of interest that have been frequently tested in the literature; (4) series groups identified via the BoW, LDA, and NMF algorithms; and (5) raw break series. In each case, the time series spans the 1957–2013 period and consists of volatility break proportions. For instance, an observation of 4.2 for real GDP in 1982 would indicate that 4.2% of all real GDP volatility breaks occurred in 1982.

Finally, we perform the following regression for series constructed using the procedure above:

$$z_t = \alpha + \beta D_t + \epsilon_t \tag{6}$$

Note that z_t is the proportion of breaks in series of a given type that happened in year t, and D_t is a dummy variable for whether year t is between years τ_0 and τ_1:

$$D_t = \begin{cases} 1 \text{ if } \tau_0 \leq t \leq \tau_1 \\ 0 \text{ if else} \end{cases} \tag{7}$$

We perform the regression above for four different (τ_0, τ_1) pairs: (1) $\tau_0 = 1985$ and $\tau_1 = 2013$; (2) $\tau_0 = 2000$ and $\tau_1 = 2013$; (3) $\tau_0 = 1985$, $\tau_1 = 1999$; and (4) $\tau_0 = 2000$, $\tau_1 = 2006$. Note that the dummies overlap and that we run a separate regression for each (τ_0, τ_1) pair. We select 1985 as the baseline cutoff for τ_0 because 1984 is most frequently identified as the start of the Great Moderation. We use the 1985–1999 and 2000–2006 sub-windows to measure the change in the intensity of breaks during the narrow Great Moderation period and the period immediately prior to the Great Recession. This segmentation follows Carvalho and Gabaix (2013), which finds that fundamental volatility increased prior to the Great Recession, starting as early as the late 1990s.

Note that $\beta < 0$ indicates that the frequency of breaks in a given series type decreased between dates τ_0 and τ_1; whereas, $\beta > 0$ indicates an increase in the frequency of breaks. Figure A3, which is located in Appendix, shows the kernel density plots of β estimates for all 400 BoW-generated series categories over the 1985–1999 period and 2000–2013 periods. Note that the 1985–1999 distribution of β indicates that series tended to experience a reduction in break frequency during the Great Moderation, which accords well with the findings in Stock and Watson (2002) and Sensier and van Dijk (2004); and suggests that the Great Moderation was far more expansive than had previously been demonstrated. To the contrary, we see a rightward shift in the distribution over the 2000–2013 period, leading to an increase in volatility breaks for 71% of series.

Table 1 formalizes these results by computing means and standard errors for the coefficient distributions generated from estimating (6)–(7). It also expands the set of results to include alternative topic modeling algorithms and tests on the autoregressive coefficients. We will start with column 1, which provides the mean estimates for breaks in GDP volatility. Here, we find that volatility break frequency tended to decrease over the 1985–2013, but experienced a substantially sharper decrease during the narrow Great Moderation period (1985–1999). Furthermore, while the decrease in volatility breaks slowed during the 2000–2006 period, the reversal did not fully materialize until the Great Recession, which generated a sufficiently large increase in volatility breaks to yield no statistically significant decline over the entire 2000–2013 period. Note that the "YES" in the "Mean" row indicates that all breaks considered in this column are breaks in the mean of volatility, rather than in its autoregressive parameter. Furthermore, BoW, LDA, and NMF refer to the algorithms used to extract topics from titles. Note that we used substring search, coupled with the BoW algorithm, to identify GDP and its components.

We next analyze the results for volatility mean breaks for an exhaustive division of the universe of time series, given in columns 2–4. We will use the BoW specification, given in column 2, as our baseline. Here, we find mixed evidence for the broad Great Moderation period. The BoW algorithm yields no statistically significant result over the 1985–2013 period, which suggests that the volatility increases over the 2000–2013 period were as large as the decreases during the narrow Great Moderation. The LDA and NMF specifications, however, find statistically significant decreases, indicating that the volatility reduction during the Great Moderation may have dominated. Similar to the results for GDP, we find that the mean series, identified using either BoW, LDA, or NMF, experienced a statistically significant decrease in volatility breaks over the 1985–1999 period and an increase over the 2000–2013 period. Furthermore, the BoW and LDA algorithm groupings yield statistically significant increases in volatility over the 2000–2006 period, which suggests that volatility may have increased elsewhere first before spilling over into GDP.

TABLE 1. Mean of β: all series and GDP series

	(1) (GDP)	(2) (ALL)	(3) (ALL)	(4) (ALL)	(5) (GDP)	(6) (ALL)	(7) (ALL)	(8) (ALL)
1985–2013	−0.955***	0.095	−0.275***	−0.233*	−1.549***	−0.737***	−1.05***	−0.845***
	(0.14)	(0.075)	(0.092)	(0.124)	(0.177)	(0.1)	(0.125)	(0.231)
1985–1999	−1.816***	−0.746***	−1.24***	−1.264***	−2.137***	−1.301***	−1.899***	−1.694***
	(0.135)	(0.105)	(0.11)	(0.13)	(0.183)	(0.123)	(0.12)	(0.217)
2000–2013	0.027	1.103***	0.795***	0.845***	−0.574***	0.261**	0.088	0.25
	(0.139)	(0.092)	(0.104)	(0.145)	(0.15)	(0.102)	(0.134)	(0.239)
2000–2006	−0.866***	0.499***	0.27*	0.223	−1.186***	−0.398***	−0.572***	−0.439
	(0.108)	(0.136)	(0.139)	(0.185)	(0.141)	(0.114)	(0.144)	(0.268)
Mean	YES	YES	YES	YES	NO	NO	NO	NO
BoW	YES	YES	NO	NO	YES	YES	NO	NO
LDA	NO	NO	YES	NO	NO	NO	YES	NO
NMF	NO	NO	NO	YES	NO	NO	NO	YES

Notes: This table provides the means of the β coefficients from equations (6)–(7). Standard errors are shown in parentheses. Columns 1 and 5 show results for GDP and GDP components exclusively. Columns 2–4 and 6–8 show results for series categories generated by topic modeling algorithms: BoW, LDA, and NMF. Columns 1–4 show results for breaks in the mean of volatility ("Mean" = YES) and columns 5–8 show results for breaks in the persistence of volatility ("MEAN" = NO). Standard errors are given in parentheses. Significance levels: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

TABLE 2. Regression results for GDP, GDP components, and commonly tested series

	(1) 1985–2013	(2) 2000–2013
GDP	-1.36^{***}	-0.59
	(0.50)	(0.39)
Government spending	-1.76^{***}	-0.92^{***}
	(0.30)	(0.26)
Consumption	-1.40^{***}	-1.00^{***}
	(0.35)	(0.31)
Investment	-1.65^{***}	-0.93^{**}
	(0.42)	(0.42)
Inventories	-2.33^{**}	-2.33^{***}
	(0.93)	(0.61)
Durable goods	-2.02^{***}	-1.80^{***}
	(0.71)	(0.54)

Notes: This table provides the means of the β coefficients from equations (6)–(7). Results are shown for the series categories most frequently tested in the literature. Standard errors are given in parentheses. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

In columns 5–8, we repeat the same exercise, but instead consider breaks in the persistence of volatility. Here, we find that the average series experienced a statistically significant reduction in volatility persistence breaks over both the broad and narrow Great Moderation periods. As with the breaks in mean volatility, breaks in persistence tended to increase over the 2000–2013 period; however, this does not appear to materialize over the 2000–2006 period, which suggests that this could be more closely related to the Great Recession.

Table 2 shows the regression results for selected GDP and GDP component series, identified via a substring search, rather than topic modeling. This table provides further confirmation for the core results in the Great Moderation literature. Again, in column 1, we find a decrease in break frequency for GDP series in the post-1984 period, but this is moderated when we limit the sample to the post-1999 period in column 2, which includes the Great Recession. We also find statistically significant decreases for frequently tested series, including investment, inventories, durable goods, and government spending.

Next, we consider an alternative division of the breaks. Rather than using topic modeling, we perform a basic substring search on titles to identify series that fall into the groupings provided by the St. Louis Fed's FRED system. In particular, we use broad and exhaustive divisions that categorize the data into 20 series groups with clear interpretations. In addition to this, we focus on the direction, rather than frequency, of the breaks. We do this by computing the z-score associated each with break,

$$z_{it} = \frac{x_{it} - \mu_i}{\sigma_i} \tag{8}$$

https://doi.org/10.1017/S1365100519000592 Published online by Cambridge University Press

TABLE 3. *Z*-score regression results for FRED tag groups

	β
Banking	-0.046^{**}
	(0.022)
Current employment statistics establishment survey	-0.040^{***}
	(0.007)
Current population survey household survey	-0.055^{***}
	(0.003)
Federal government debt	0.043^{*}
	(0.031)
Finance companies	0.049^{**}
	(0.025)
Income distribution	-0.056^{**}
	(0.025)
Industrial production capacity utilization	-0.033^{***}
	(0.004)
National accounts	-0.008^{***}
	(0.001)
National income product accounts	-0.008^{***}
	(0.001)
Population employment labor markets	-0.05^{***}
	(0.003)
Production business activity	-0.026^{***}
	(0.004)
Productivity costs	-0.043^{***}
	(0.012)
US trade international transactions	-0.041^{**}
	(0.021)

Notes : We performed a break test on all series within each broad FRED tag category. We then computed the size of the change in each series at the break in standard deviations (z-score), and regressed the z-score on a time trend and constant. We did not find significant results for "prices," "countries," "consumer price indexes cpi and pce," "producer price indices ppi," "indicators," "international data," and "population." Standard errors are given in parentheses. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < .01$.

where z_{it} is the z-score for series i at time t, x_{it} is the series value at time t, μ_i is the series mean, and σ_i is the series standard deviation.

In each case, we estimate the coefficient on a linear time trend over the entire sample period. A negative coefficient estimate indicates that the series tended to experience reductions in volatility over the 1957–2013 period; and, thus, may be a long-term source of downward pressure on macroeconomic volatility. Of the 20 series categories, 13 yield at least marginally significant coefficient estimates for β, which are shown in Table 3. Of these, only "finance companies" and "federal government debt" have positive coefficient estimates. The finding on "finance companies" complements Carvalho and Gabaix (2013): they show that growth

https://doi.org/10.1017/S1365100519000592 Published online by Cambridge University Press

in the financial sector's share of the economy increased fundamental volatility; whereas, we show that volatility in series associated with the sector itself grew.

The largest reductions in volatility appear to come from demographic and labor market sources: (1) current population survey (CPS) series; (2) current employment statistics establishment survey series; and (3) population employment labor markets series. For each of these categories, the average z-score reduction was an order of magnitude larger than for the NIPA or National Accounts categories. This supports the literature initiated by Jaimovich and Siu (2009), who find that reductions in the volatile-age share of labor partially account for the Great Moderation. Note that we do not test this same hypothesis, but instead show that the volatilities of demographic and labor market series themselves broadly declined over the entire 1957–2013 period. Thus, the evidence we provide is not subject to the critique in Everaert and Vierke (2016).[12]

Finally, Figure A4, which is located in Appendix, plots the raw z-scores for breaks associated with two categories of demographic and labor market series. In both panels, we see large, positive z-scores clustered in the 1960s and 1970s, followed by a decline in both frequency and z-score magnitude that continues through the broad Great Moderation period. In both cases, z-scores continue to fall through the Great Recession and remain low thereafter, suggesting that demographic and labor market stability may play a continued role in dampening GDP volatility.

5. CONCLUSION

From the universe of aggregate US time series in the St. Louis Fed's FRED database, we identified a subset of 9,355 series that each span the 1955–2015 period and were suitable for break testing. We then transformed each series, performed a time-varying parameters regression, extracted the regression residuals, and then subjected a transformation of their absolute values to Bai and Perron (2003) tests for up to 10 unknown structural breaks, which yielded a total of 17,681 breaks. Each break is best interpreted as a shift in the volatility of a series. We then applied text analysis to the break series titles, categorized them into groups based on title similarity, and then used the resulting data to revisit open hypotheses in the Great Moderation literature, extending most closely from Stock and Watson (2002) and Sensier and van Dijk (2004); and complementing Jaimovich and Siu (2009) and Carvalho and Gabaix (2013).

We first replicate the original Great Moderation findings from Kim and Nelson (1999) and McConnell and Perez-Quiros (2000): GDP and GDP component series experience fewer breaks in volatility after 1984, which is conventionally dated as the start of the Great Moderation. This result also extends to the list of series that are most frequently tested series in the Great Moderation literature: (1) inventories; (2) durable goods; (3) consumption; (4) government spending; and (5) investment. Furthermore, when we use text analysis to sort the 17,681 breaks into series categories according to title similarity, we find that 70.5% experience fewer breaks in volatility during the 1985–1999 period. This result updates and expands the original findings of Stock and Watson (2002) and Sensier and van Dijk (2004); and suggests that the scope of the Great Moderation was far broader than the literature had previously identified.

Next, we repeat the same exercise, but use 2000–2013, rather than 1985–2013, as the window of interest. This range of dates was selected because it is most closely associated with the Great Recession and the era surrounding it, including the rise in indebtedness and house prices, and the ensuing crash and deleveraging. It is also the period during which Carvalho and Gabaix (2013) identify an increase in fundamental volatility. We find a slightly higher frequency of volatility breaks for GDP series and some of its components, relative to the narrow Great Moderation period. Furthermore, when we extend the analysis using all series categories, which we identify via topic modeling of series titles, the results are far more pronounced, leading to a sign flip. In particular, the frequency of volatility breaks for all series categories declines during the narrowly defined Great Moderation period (1985–1999), but increases in both the 2000–2013 and 2000–2006 periods. This suggests that broad, sharp volatility increases prior to the Great Recession were not well captured by changes in GDP and GDP component volatility.

Finally, we use a separate division of the break series to identify two sources of increased volatility during the Great Moderation: (1) finance companies, and (2) federal government debt. Of the 20 broad categories identified by the St. Louis Fed's FRED database, these were the only two that experienced significant increases in volatility. Using the same division, we show that demographic factors and labor markets have been a consistent source of stability since 1960s; and have witnessed persistent declines in volatility, even during the Great Recession, complementing the hypothesis in Jaimovich and Siu (2009). Furthermore, the sizes of the decreases were among the largest of all series categories; and were an order of magnitude greater than the NIPA and National Accounts declines, which form the original basis for the Great Moderation literature.

NOTES

1. See Stock and Watson (2002), Ahmed et al. (2004), Primiceri (2006), Sims and Zha (2006) for work on the better policy hypothesis; Bils and Kahn (2000), Kahn (2008), Maccini and Pagan (2013), and Sarte et al. (2015) for work on improved business practices; and Clarida et al. (2000), Lubik and Schorfheide (2004), Boivin and Giannoni (2006), Herrera and Pesavento (2009), and Coibion and Gorodnichenko (2011) for work on the smaller shocks hypothesis.

2. See Clark, 2009, Coibion and Gorodnichenko (2011), Gabaix (2011), Acemoglu et al. (2012), Stock and Watson (2012), Carvalho and Gabaix (2013), and Burren and Neusser (2013).

3. Specifically, they argue that increased credit availability induced a wealth effect by enabling homeowners to withdraw equity and businesses to engage in share buybacks. They also argue that hypotheses related to the Great Moderation should be tested in variances, rather than first moments, since the Great Moderation is primarily a story about reduced volatilities.

4. See Stock and Watson (2002) for the original set of series that was identified and then retested by many papers in the Great Moderation literature.

5. For a theoretical examination of the relationship between demographic change and the Great Moderation, see Lugaeur (2012a) and Heer et al. (2017).

6. We retain bilateral series that reference the US, but otherwise discard all international series.

7. Note that we remove terms like “percent” and “change” because we ultimately use a procedure similar to the one in Stock and Watson (2002) to transform all series. We remove common terms, such as “gross” and “contributions” because our intent is to describe broad groups that contain multiple time series. This approach also avoids the creation of unintended and uninterpretable groupings across categories of series (e.g. gross interest rates and gross domestic product).

8. The document-term matrix stores the frequency with which each word is used in each title in the corpus.

9. This is determined by the Bai and Perron (2003) test, which excludes data from the endpoints, where breaks are difficult to identify.

10. McConnell and Perez-Quiros (2000), Blanchard and Simon (2001), and Stock and Watson (2002) find that the Great Moderation is mostly explained by breaks in volatility, rather than breaks in means.

11. We use a Monte Carlo simulation to test whether persistence in $\hat{\sigma}_{\epsilon_t}$ would confound our break tests. Specifically, for a given autoregressive parameter, we simulate 1000 series, each of which is 56 periods in length. We insert a structural break in the 28th period by increasing the mean by 3 standard deviations of the error variance. For an autoregressive parameter of 0.80, 95% of identified break dates are within three periods of the true date. The number of periods is reduced to 2 for autoregressive parameters of 0.90, 0.95, and 0.99. This suggests that our results should be robust to the presence of persistence in volatility.

12. Jaimovich and Siu (2009) regress the volatility of output on the volatile share of the labor force (e.g. workers aged 15–29 or 60–64) in G7 countries. They find that a decrease in the volatile share of labor may have accounted for 1/5 to 1/3 of the volatility reduction during the Great Moderation. Everaert and Vierke (2016) suggests that the regression may be spurious, since the series exhibit non-stationarity, but are not cointegrated.

REFERENCES

Acemoglu, D., V. Carvalho, A. Ozdaglar and A. Tahbaz-Salehi (2012) The network origins of aggregate fluctuations. *Econometrica* 80(5), 1977–2016.

Ahmed, S., A. Levin and B. Wilson (2004) Recent U.S. macroeconomic stability: Good policies, good practices, or good luck? *The Review of Economics and Statistics* 86(3), 824–832.

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

Bils, M. and J. Kahn (2000) What inventory behavior tells us about business cycles. *American Economic Review* 90(3), 458–481.

Blanchard, O. and J. Simon (2001) The long and large decline in U.S. output volatility. *Brookings Papers on Economic Activity* 32(1), 135–164.

Boivin, J. and M. Giannoni (2006) Has monetary policy become more effective? *The Review of Economics and Statistics* 88(3), 445–462.

Burren, D. and K. Neusser (2013) The role of sectoral shifts in the decline of real GDP Volatility. *Macroeconomic Dynamics* 17(3), 477–500.

Carvalho, V. and X. Gabaix (2013) The great diversification and its undoing. *American Economic Review* 103(5), 1697–1727.

Clarida, R., J. Gali and M. Gertler (2000) Monetary policy rules and macroeconomic stability: Evidence and some theory. *The Quarterly Journal of Economics* 115(1), 147–180.

Clark, T. (2009) Is the great moderation over? An empirical analysis. Federal Reserve Bank of Kansas City, *Economic Review* 4, 5–42.

Coibion, O. and Y. Gorodnichenko (2011) Monetary policy, trend inflation, and the great moderation: An alternative interpretation. *American Economic Review* 101(1), 341–370.

Davidian, M. and R. Carroll (1987) Variance function estimation. *Journal of the American Statistical Association* 400(82), 1079–1091.

Davis, S. and J. Kahn (2008) Interpreting the great moderation: Changes in the volatility of economic activity at the macro and micro levels. *Journal of Economic Perspectives* 22(4), 155–180.

Everaert, G. and H. Vierke (2016) Demographics and business cycle volatility: A spurious relationship? *Journal of Applied Econometrics* 31, 1467–1477.

Gabaix, X. (2011) The granular origins of aggregate fluctuations. *Econometrica* 79(3), 733–772.

Grydaki, M. and D. Bezemer (2013) The role of credit in the great moderation: A multivariate GARCH approach. *Journal of Banking and Finance*, 37, 4615–4626.

Heer, B., S. Rohrbacher and C. Scharrer (2017) Aging, the great moderation, and business-cycle volatility in a life-cycle model. *Macroeconomic Dynamics* 21(2), 362–383.

Herrera, A.M. and E. Pesavento (2009) Oil price shocks, systematic monetary policy, and the "Great Moderation." *Macroeconomic Dynamics* 13(1), 107–137.

Hornik, K., C. Kleiber, W. Krämer and A. Zeileis (2003) Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis* 44(1–2), 109–123.

Jaimovich, N. and H. Siu (2009) The young, the old, and the restless: Demo- graphics and business cycle volatility. *American Economic Review* 99(3), 804–829.

Kahn, J. (2008) Durable goods inventories and the great moderation. *Federal Reserve Bank of New York Staff Report No. 325*.

Kim, C. and C. Nelson (1999) Has The U.S. economy become more stable? A bayesian approach based on a Markov-Switching model of the business cycle. *The Review of Economics and Statistics* 81(4), 608–616.

Lubik, T. and F. Schorfheide (2004) Testing for indeterminacy: An application to U.S. monetary policy. *American Economic Review* 94(1), 190–217.

Lugaeur, S. (2012a) Demographic change and the great moderation in an overlapping generations model with matching frictions. *Macroeconomic Dynamics*, 16(5), 706–731.

Lugaeur, S. (2012b) Estimating the effect of the age distribution on cyclical output volatility across the United States. *Review of Economics and Statistics* 94(4), 896–902.

Lugaeur, S. and M. Redmond (2012) The age distribution and business cycle volatility: International evidence. *Economics Letters* 117(3), 694–696.

McConnell, M. and G. Perez-Quiros (2000) Output fluctuations in the United States: What has changed since the early 1980s. *The American Economic Review* 90(5), 1464–1476.

Maccini, L.J. and A. Pagan (2013) Inventories, fluctuations, and goods sector cycles. *Macroeconomic Dynamics* 17(1), 89–122.

Primiceri, G. E. (2006) Why inflation rose and fell: Policy-Makers beliefs and U.S. postwar stabilization policy. *The Quarterly Journal of Economics* 121(3), 867–901.

Sarte, P., F. Schwartzman and T. Lubik (2015) What inventory behavior tells us about how business cycles have changed. *Journal of Monetary Economics* 76, 264–283.

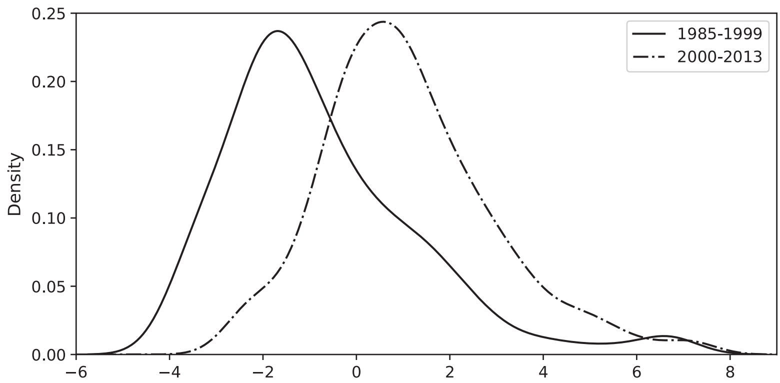
Sensier, M. and D. van Dijk (2004) Testing for volatility changes in U.S. macroeconomic time series. *The Review of Economics and Statistics* 86(3), 833–839.

Sims, C. and T. Zha (2006) Were there regime switches in U.S. monetary policy. *American Economic Review* 96(1), 54–81.

Stock, J. and M. Watson (2012) Disentangling the Channels of the 2007–2009 Recession, NBER Working Papers: No. 18094.

Stock, J. H. and M. W. Watson (2002) Has the business cycle changed and why? *NBER Macroeconomics Annual* 17, 160–230.

APPENDIX



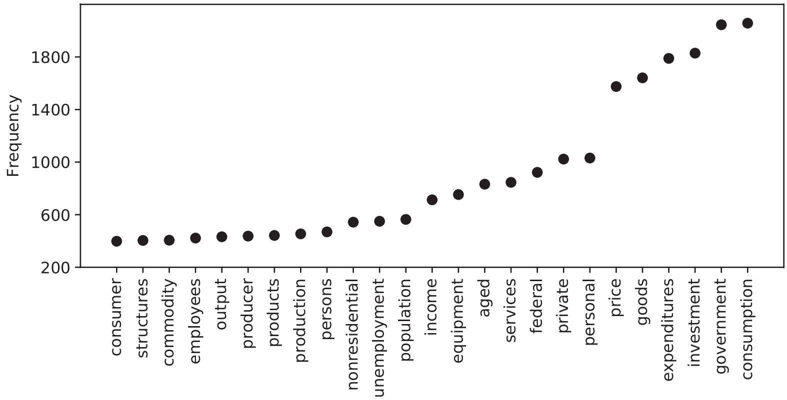
Notes: The figure above shows the distributions of β estimates across all series for (τ_0,τ_1) = (1985–1999) and (2000, 2013). A negative value of β indicates that the frequency of breaks declined.

FIGURE A1. Distribution of regression coefficients.

Notes: The figure above shows GDP volatility computed using four different methods: (1) the method introduced in this paper (break proportion); (2) fundamental volatility from Carvalho and Gabaix (2013); (3) five-year rolling standard deviations (e.g. Davis and Kahn (2008)); and (4) the residual from an autoregressive process (e.g. Stock and Watson (2002)).

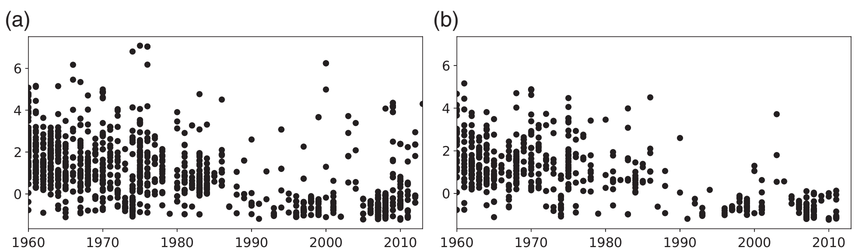
FIGURE A2. GDP volatility: method comparison.

https://doi.org/10.1017/S1365100519000592 Published online by Cambridge University Press



Notes: The unigrams in the figure above were computed using annual, quarterly, and monthly series. Titles were stripped of common terms and the bag-of-words algorithm was applied to recover the frequency of word use.

FIGURE A3. Most common topics for annual, quarterly, and monthly series titles.



Notes: We performed a break test on all series within each broad FRED tag category. We then computed the size of the change in each series at the break in standard deviations (z-score). The plot above shows a time series of the z-scores for employment series and current population survey series.

FIGURE A4. Demographic and labor market breaks. (a) Employment series. (b) Current population survey.

https://doi.org/10.1017/S1365100519000592 Published online by Cambridge University Press