Economic Quarterly—Volume 100, Number 1—First Quarter 2014—Pages 1–21

Measuring Resource Utilization in the Labor Market

Andreas Hornstein, Marianna Kudlyak, and Fabian Lange

The U.S. unemployment rate increased substantially following the Great Recession, reaching close to 10 percent in the fourth quarter of 2009. As of December 2014, the unemployment rate has declined by more than 4 percentage points, faster than many policymakers forecasted at the time. As unemployment rates declined, labor force participation rates also declined by about 2 percentage points. This has raised doubts on the ability of the unemployment rate alone to accurately represent the state of resource utilization in the labor market.[1] Broader measures than the standard unemployment rate may therefore be needed to indicate resource utilization in the labor market.

In this article, we briefly review the extended unemployment measures of the Bureau of Labor Statistics (BLS), which capture individuals not usually counted as unemployed. Importantly, these measures of unemployment assign the same weight to all nonemployed individuals included in the measures despite there being substantial differences in labor force attachment among the nonemployed. For example, those nonemployed who are actively searching for work usually have a higher transition rate to employment than those who express a desire to work but do not actively engage in job search activities. Presumably these

Andreas Hornstein is a senior advisor and Marianna Kudlyak is an economist in the Research Department at the Federal Reserve Bank of Richmond. Fabian Lange is an associate professor of economics at McGill University. This article is a substantially revised version of Hornstein, Kudlyak, and Lange (2014) and provides the background for Hornstein et al. (2014). The authors thank Marisa Reed for excellent research assistance. The views expressed in this article are those of the authors and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: Andreas.Hornstein@rich.frb.org; Marianna.Kudlyak@rich.frb.org; Fabian.Lange@mcgill.ca.

[1] See, for example, Appelbaum (2014) or Yellen (2014).

persistent differences in transition rates reflect differences in the degree of labor force attachment.

We therefore proceed to construct an alternative measure of labor utilization—a nonemployment index—that accounts for differences in labor market attachment among nonemployed individuals. Our approach builds on recent advances in our understanding of how individuals transition between labor market states, identifying labor market attachment with observed average transition rates to employment. Since we weight nonemployed individuals by their relative transition rates to employment, our measure can cover all nonemployed individuals, and we are not forced to draw arbitrary distinctions on who is to be included in the set of nonemployed individuals as is necessary even for the usual BLS extended measures of unemployment.

Even though broader measures of resource utilization, that is, the extended BLS measures and our nonemployment index, may better reflect the "true" state of the labor market, the standard unemployment rate may still represent a valid signal of the cyclical state of the labor market.[2] We find that prior to the Great Recession the standard unemployment rate and broader measures of unemployment are indeed moving closely together. Thus, the broader measures of resource utilization and the more narrow standard unemployment rate provide the same signal about the labor market prior to 2007.

After the Great Recession, however, there appears to be a break in the relationship between the standard unemployment rate and the broader measures of resource underutilization. Whether this break implies that the standard unemployment rate understates or overstates the true degree of resource underutilization in the labor market after the Great Recession does however depend on the measure of "true" resource underutilization. If one believes that the BLS measure—the extended unemployment rate U6, which includes the marginally attached and those working part time for economic reasons—best reflects the true state of the labor market, then the standard unemployment rate understates how much labor in the labor market is idle after 2007. If, however, we believe that the nonemployed should be weighted by their workforce attachment, then the standard unemployment rate overstated true resource underutilization for most of the post-2007 period and provides a more or less accurate representation of labor resource underutilization as of 2014.

[2] For instance, the extended unemployment rate U6, which includes the marginally attached and those working part time for economic reasons, is by construction always greater than the standard unemployment rate (U3). Even if U6 more accurately captures the totality of all labor resources that are underutilized in the labor market, it is possible that U3 provides a good indication of the state of the business cycle in the labor market.

Our analysis thus shows that the standard unemployment rate will not always accurately reflect "true" underlying resource underutilization. In particular, taking the nonemployment index as a "true" measure of labor resource underutilization, the discrepancy (or lack thereof) between the signal and the true measure depends on the composition of the nonemployed population by their degree of work attachment.

More than 30 years ago, Flinn and Heckman (1983) pointed out that the distinction between those being unemployed and those being out of the labor force is not clear cut but a matter of degree. Recently, and mostly in the context of estimating matching efficiency of the labor market, Veracierto (2011), Diamond (2013), Elsby, Hobijn, and Şahin (2013), and Hall and Schulhofer-Wohl (2013) have argued that it is important to account for the job seekers out of the labor force in addition to the unemployed. Furthermore, Hornstein (2012) and Krueger, Cramer, and Cho (2014) have argued that even within the group of unemployed the pattern of long-term unemployment suggests significant differences in employability.[3] Kroft et al. (2013) explore how differences in transition rates to employment across unemployed with different unemployment duration and those out of the labor force (OLF) shaped the evolution of the U.S. labor market over the Great Recession. To our knowledge, our nonemployment index is the first measure that consistently aggregates different categories of the nonemployed using observed differences in employability. Similar measures of labor market resource utilization were constructed for the United Kingdom (see Jones, Joyce, and Thomas [2003]; and Schweitzer [2003]).

This article is structured as follows. We first characterize differences in workforce attachment among the nonemployed in terms of their average transition rates to employment. We then review the various (extended) unemployment rates constructed by the BLS and construct an alternative index of nonemployment that weights its components according to their workforce attachment. Finally, we evaluate the quality of the standard unemployment rate as a signal for broader measures of nonemployment.

1. HETEROGENEITY OF NONEMPLOYMENT

The BLS Classification Scheme

Among the most widely reported statistics from the BLS are the shares of the working-age population who are currently employed, unemployed,

[3] Recent resume audit studies (Kroft, Lange, and Notowidigdo 2013; Erikson and Rooth 2014) confirm differences in employability between the short-term and long-term unemployed.

Table 1 Nonemployment by BLS Categories

	1	2	3	4	5	6
	Share of Working-Age Population			Employment Probability		
	1994–2013	2007	2010	1994–2013	2007	2010
	Unemployed					
Short-term	3.0	2.5	3.5	28.0	29.7	21.8
Long-term	1.0	0.5	2.7	14.4	15.5	10.3
	OLF, Want a Job					
Marginally attached, discouraged	0.2	0.2	0.5	13.1	16.5	10.7
Marginally attached, other	0.4	0.3	0.3	12.7	14.9	10.2
Other	1.8	1.5	1.7	14.5	15.7	12.1
	OLF, Do Not Want a Job					
Other, in school	4.1	4.5	5.0	8.5	8.2	6.2
Other, not in school	7.4	7.2	7.0	7.5	8.1	6.9
Disabled	4.6	4.8	5.2	1.7	1.7	1.4
Retired	15.4	15.2	15.4	1.4	1.5	1.4

Notes: Share of working-age population and employment transition probability in percent.

and OLF. These shares are estimated using responses from the monthly Current Population Survey (CPS). A nonemployed respondent is counted as unemployed if she has been actively looking for work in the month preceding the survey week. Those neither employed nor actively looking for work are classified as OLF. Starting with the comprehensive revision of the CPS in 1994, the BLS provides additional detail on the labor market attachment of the nonemployed based on survey responses as to why an individual is not actively looking for work. The average population shares for the different nonemployment categories in the CPS are listed in Table 1, columns 1 through 3. We report these shares for the period 1994–2014 and the years 2007 and 2010, that is, the year prior to the Great Recession and the year when unemployment reached its peak.

The unemployed can be subdivided based on their reported length of unemployment. Short-term unemployment (STU) covers those who have been unemployed for 26 or fewer weeks, while long-term unemployment (LTU) encompasses those who have been unemployed for more than 26 weeks. On average, only one-fourth of all unemployed report

more than 26 weeks of unemployment in any one month, but the share of LTU increased to close to one-half following the Great Recession.[4]

The unemployed represent only one-tenth of those without employment. The remaining nine-tenths are OLF.

Over nine-tenths of those OLF do not want a job. Among these individuals we can distinguish between those who are retired, disabled, currently in school, and the remainder. On average, the retired and disabled account for about two-thirds of those who do not want work. Following the Great Recession we saw a noticeable increase in the disabled and those attending school.

While most OLF do not want a job, a little less than one-tenth declare that they do want to work, even though they did not actively look for work in the previous month. Those in this group who want a job, are available for work, and searched for work within the last year (not the last month) are classified as marginally attached. On average, about one-fourth of those who want work are marginally attached, and there are twice as many unemployed as there are marginally attached respondents. Those marginally attached who did not search for a job during the last month because they were discouraged over job prospects are classified as discouraged. On average, discouraged individuals make up about one-third of the marginally attached, but following the recession their share increased noticeably.

Transition Rates to Employment

We are motivated to examine broader unemployment concepts since the distinction between unemployment and OLF is not as sharp as one would think. In fact, from month to month, roughly twice as many individuals transition from OLF as opposed to unemployment to employment. We now show that the transition rates to employment are indeed positive for all nonemployed, but that there is also substantial heterogeneity in transition rates among the nonemployed. We also show that the pattern of average transition rates to employment among the nonemployed seems to be consistent with the self-reported labor market attachment.

We first use the CPS microdata to construct transition probabilities from nonemployment to employment using the short rotating four-month panels in the CPS. In any month we observe the labor market status in the current and following month for three-fourths of

[4] That the share of LTU has been exceptionally high since 2007 is also evident from the fact that the average share of LTU for the period from 1948–2007 was a mere 15 percent.

the sample. Based on the responses to the CPS questions, we group the nonemployed into the nine nonemployment segments discussed above: the two duration segments of the unemployed, the three segments of OLF who want a job (marginally attached, discouraged, other), and the four segments of OLF who do not want a job (retired, disabled, in school, not in school). We then construct the transition probabilities into employment for each segment by matching the individual records from the CPS microdata month to month.[5] The transition probability from a particular segment of nonemployment to employment is the fraction of that segment that exits to employment from one month to the next.

Table 1, columns 4 through 6, show annual averages of the monthly probabilities of becoming employed for the two unemployment segments and seven OLF segments averaged across 1994–2014, and for the years 2007 and 2010. The probability of becoming employed differs substantially among these groups. The probability is highest for the short-term unemployed: On average they have a 30 percent chance of finding a job within a month. Next are the long-term unemployed and those OLF individuals who want a job: They are about half as likely to become employed as are the STU.[6] Then there is the group of those who do not want a job but who are neither retired nor disabled: They are only one-fourth as likely to become employed as are the STU. Finally, there is the group of retired and disabled who are less than one-tenth as likely to become employed as are the STU.[7]

In recessions the employment probabilities tend to fall for all groups, but the ranking of the different groups in terms of their transition probabilities to employment remains the same.[8] Furthermore, the ranking of employment probabilities coincides with the desire to work as stated in the survey: Those who actively search tend to have higher transition rates to employment than those who want to work but do not actively look for work, and those who want to work have higher transition rates than those who do not want to work.

---

[5] Our matching procedure follows the algorithms described in Madrian and Lefgren (1999) and Shimer (2012) The CPS microdata fields are available at http://thedataweb.rm.census.gov/ftp/cps_ftp.html#cpsbasic.

[6] Note that the employment transition rates among the marginally attached OLF do not differ much. In particular, there is no reason to single out discouraged workers based on the likelihood of becoming employed again.

[7] See also Fujita (2014).

[8] See Kudlyak and Lange (2014) for graphs of annual averages of monthly job finding rates for the years 1994 to 2013.

Classification by Labor Force Status Histories

The decomposition of the OLF nonemployed as to why they are not actively looking for work is only available since 1994.[9] This is unfortunate since the Great Recession is an exceptional event for the period since 1994, and we therefore cannot tell whether broader measures of labor market resource utilization performed differently during the Great Recession than at other times of stress in the labor market. We therefore consider an alternative measure of the labor force attachment of the nonemployed that is based on individuals' observed labor market histories and that can be constructed for the time period since 1976. This longer time period contains the recessions of the early 1980s when standard measures of unemployment were of a magnitude similar to the Great Recession.

For the period since 1976, Kudlyak and Lange (2014) use the panel feature of the CPS to construct labor market segments based on respondents' labor force status (LFS) histories, that is, their status as employed, unemployed, or OLF in the current month and the preceding two months. They define classes of LFS histories based on the status in the current month, and whether the current status of a nonemployed individual differs from the status in the preceding two months in particular, if the nonemployed was employed (see Table 2). Conditional on this decomposition of the nonemployed for each segment, Kudlyak and Lange (2014) calculate the probability of being employed in the next month. They find significant and persistent differences in the employment probabilities for these segments.

In Table 2 we report the average population shares and employment transition probabilities of the nonemployed for the Kudlyak and Lange (2014) decomposition.[10] The population shares of the nonemployed segments with different LFS histories for the full sample period, 1976–2014, and the post-1994 subsample are very similar. Nonemployed individuals who were employed in at least one of the previous two months have the highest chance of being employed again. For this group, active search increases the probability of reemployment somewhat but not much. Next are the nonemployed who have no recent

[9] Prior to 1994, only individuals who were about to exit the sample were asked about their desire to work. Thus, the job-finding probabilities for the OLF segments by desire to work cannot be constructed prior to 1994.

[10] We should note that there is month-to-month attrition in the CPS sample that is in addition to the outgoing rotation groups. Since the population shares of currently unemployed and OLF in the subsample with complete three-month LFS histories are not the same as the population shares in the full sample, cf Tables 1 and 2, this attrition does not appear to be completely random.

Table 2 Nonemployment by Labor Force Status Histories

	1	2	3	4	5	6	7	8
	Share of Working-Age Population				Employment Probability			
	1976–2014	1994–2014	2007	2010	1976–2014	1994–2014	2007	2010
	Currently Unemployed							
Recent employment	1.3	1.2	1.1	1.4	38.8	39.2	40.7	34.2
No recent employment	1.1	1.1	0.8	1.5	17.1	16.0	17.2	9.6
Continuously unemployed	1.4	1.3	0.8	2.8	17.7	17.2	19.0	11.0
	Currently OLF							
Recent employment	2.9	2.8	3.0	2.6	27.7	27.1	27.8	27.6
No recent employment	1.3	1.3	1.0	1.9	9.6	9.5	9.6	7.1
Continuously OLF	30.9	30.2	30.4	31.1	2.0	1.8	1.8	1.5

Notes: The first set of rows covers those nonemployed who are unemployed in the current month and the second set covers those nonemployed who are OLF in the current month. For each group, the first row (Recent employment) denotes those who have been employed at least once in the previous two months; the second row denotes those who have not been employed in any of the previous two months but also not unemployed/OLF in both months; and the last row denotes those who have been unemployed/OLF in both of the previous two months. The share of working-age population and the employment probability are in percent.

employment experience but are actively looking for work: Having no recent work experience reduces the employment probability by more than half. Finally, there are the nonemployed who are not actively looking for work and have no recent employment experience: They are less than one-fourth as likely to find work. Similar to the BLS classification by reason of nonemployment, the employment transition rates decline significantly in a recession, for example from 2007 to 2010 following the Great Recession, but the relative rankings remain constant.[11]

Our evidence from employment transition rates suggests that clear distinctions between being in and out of the labor force are not possible and might not be useful for determining the degree of labor utilization. This conclusion emerges for both methods of measuring labor force attachment. For example, for the BLS classification by reason of nonemployment, those who are OLF but want to work have essentially the same employment probabilities as the long-term unemployed, yet only the latter are included in the standard unemployment rate. Similarly for the Kudlyak and Lange (2014) classification based on LFS histories, even though those nonemployed who are OLF with some recent employment experience are more likely to become employed than those who are unemployed with no recent employment experience, the latter and not the former are included in the standard definition of the unemployment rate.

2. MEASURES OF RESOURCE UTILIZATION

The most widely used measure of resource utilization in the labor market is the unemployment rate, U3 to be precise. The unemployment rate is defined as the share of the unemployed, that is, those nonemployed who are actively looking for work, in the labor force where the latter is the sum of the employed and unemployed. We now briefly review the BLS extended measures of unemployment that broaden the set of the potentially employable working-age population, but weight all of these potentially employable equally. Since we have argued above that labor force attachment for the nonemployed is a matter of degree rather than satisfying a simple in or out criteria, we then propose two alternative indices of nonemployment that quantify the degree of labor force attachment. These indices include all nonemployed members of the working age population but weight the nonemployed according to their average employment transition rate.

[11] Again, see Kudlyak and Lange (2014) for time series of annual averages of the transition rates.

Extended Unemployment Rates from the BLS

The BLS constructs extended measures of unemployment that move subgroups from OLF to unemployed. In particular, the U4 rate adds discouraged workers from the marginally attached, and the U5 rate includes all marginally attached. The corresponding unemployment rates are defined as before with appropriately adjusted labor force measures. In addition, the BLS publishes the U6 rate, which includes those employed who are working part time for economic reasons (PTfER) in the unemployment rate.[12] These individuals, sometimes referred to as involuntary part-time workers, would have preferred to work full time but had to work part time because they did not find full-time work or because their hours had been reduced to part-time work. Including these employed among the unemployed is usually motivated by the argument that, like the unemployed, they are not employed as much as they would like to be. For each of these extended measures of unemployment, the group that is added receives the same weight as the unemployed who are part of U3.[13]

Nonemployment Rates Adjusted for Labor Market Attachment

We now construct a nonemployment index (NEI) that is more comprehensive than the unemployment rate but also accounts for the fact that not all nonemployed are equally attached to the labor market. Our proposed NEI is a weighted average of the population shares of the various subgroups among the unemployed and OLF, where the weight for each subgroup is given by the sample average of its employment transition rate relative to the group with the highest transition rate. Our index thus measures the effectively available labor resources in units of the group with the strongest labor market attachment.[14] We use sample

[12] Unlike for U4 and U5, adding those working PTfER does not increase the labor force in the definition of the unemployment rate.

[13] Bregger and Haugen (1995) provide a short history of the BLS extended measures of unemployment.

[14] Our procedure to adjust available nonemployed for their effective labor market attachment is similar to the quality adjustment of employment, where one uses relative wages as measures of relative labor efficiency. These quality-adjusted employment measures have a long tradition in labor economics. For example Katz and Murphy (1992) use this method to generate efficiency units of labor supply by education group. In addition to weighting the nonemployed by their relative job finding rate, one can consider the quality of jobs that different segments of the nonemployed find. This investigation is beyond the scope of the article.

averages of the transition rates to ensure that the variation in the index over time is not driven by cyclical changes in relative transition rates.

We construct two versions of the NEI. The first version uses the BLS classifications of nonemployment for the period from 1994 on, NEI1 for short, and the second version uses the Kudlyak and Lange (2014) classification scheme based on LFS histories from 1976 on, NEI2 for short. Employment transition rates are defined relative to the short-term unemployed for the BLS classification and relative to the unemployed with some employment in the previous two months for the LFS history classification.

For each NEI we also construct a version that incorporates those working part time for economic reasons. We weight this group by the product of its relative transition probability to full-time employment and its "underutilization" rate. Analogous to the weighting of the nonemployed, we normalize the transition rate relative to the highest employment transition rate among the group of the nonemployed. The underutilization rate is defined as the ratio of the difference of the average weekly hours worked by those working full time and the average weekly hours worked by those working part time for economic reasons to the average weekly hours worked by those working full time.

Using the CPS microdata from January 1994 to December 2013, we find that the average monthly transition probability from involuntary part-time work to full-time work is 0.30, about the same as the employment transition rate of the short-term unemployed. The average work week of those working PTfER is 22.9 hours, about half of the work week of those working full time, which is 44.5 hours.[15] Those working part time for economic reasons therefore receive a weight of about one-half in the nonemployment index.[16]

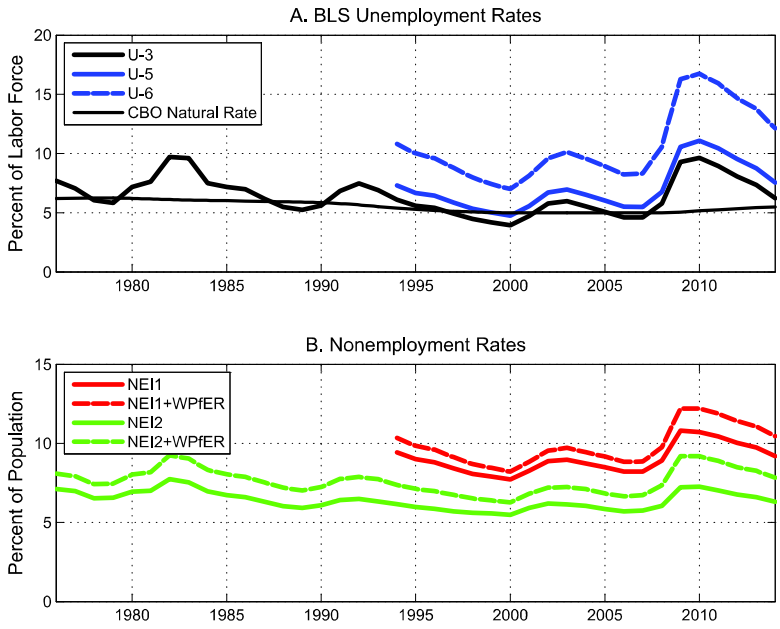
A First Look at Resource Utilization, 1976–2014

The qualitative features of the standard unemployment rate, the extended unemployment rates, and the nonemployment rates are essentially the same: They rise and fall together and all increase more

[15] For these calculations we use reported "actual total" hours worked. Alternatively, we could use "usual total" hours worked, or "total" or "usual" hours worked at the primary job. For these various hours measures, the implied weight on those working part time for economic reasons in the nonemployment index then ranges from 0.133 to 0.145. Thus our choice of "hours worked" definition maximizes the weight for those working part time for economic reasons.

[16] Hornstein, Kudlyak, and Lange (2014) and Hornstein et al. (2014) use an ad hoc weight of 0.5 for those working part time for economic reasons. This weighting choice also follows the pre-1994 BLS definition for U6, Bregger and Haugen (1995).

Figure 1 Measures of Labor Market Resource Utilization



Notes: The series are annual averages of monthly unemployment rates and nonemployment rates. The BLS unemployment rates in Panel A are the standard U3 unemployment rate for the period 1976–2014, black line, and the extended unemployment rates, U5 (solid blue) and U6 (dashed blue) for the period 1994–2014. The extended rate U5 includes unemployed and marginally attached workers, and U6 includes unemployed and marginally attached and those working part time for economic reasons. The thin black line is the CBO natural rate of unemployment. The nonemployment rates in Panel B are our alternative measures based on BLS nonemployment categories for 1994–2014, red solid line, and LFS histories for 1976–2014, green solid line. The corresponding dashed lines include weighted employed who are working part time for economic reasons.

following the Great Recession than they did during the 2001 recession. The standard unemployment rate U3 and the two extended unemployment rates U5 and U6 are displayed in the top panel of Figure 1, and the two nonemployment indices, with and without PTfER, are displayed in the bottom panel of Figure 1. The rates differ in their levels and to some extent in their volatility.

It is common to assume that because of frictions in the labor market there will always be some unemployment in the economy. In other words, there is a natural rate of unemployment and policy should only be concerned with deviations from that natural rate. For the standard U3 unemployment rate, the most frequently referenced estimate of the natural rate is provided by the Congressional Budget Office (CBO), the thin black line in the top panel of Figure 1. The CBO has the natural rate increasing from about 5.2 percent in 1950, to 6.2 percent in the late 1970s, from where it declines to 5 percent by 2000, and then increases again to 5.5 following the Great Recession. According to the CBO, the natural rate is essentially 5 percent with some upward allowance made when actual unemployment is high.

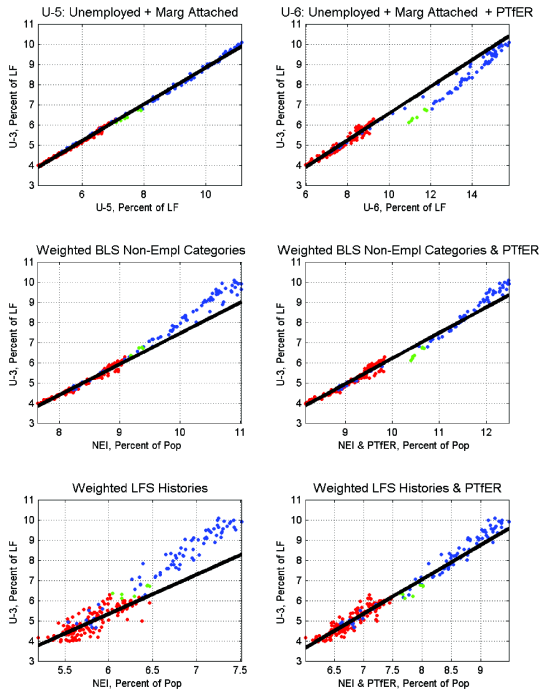
By construction, the extended unemployment and nonemployment rates are higher than the standard unemployment rate, but similar to the standard unemployment rate, therefore one could define natural rates that stay close to the respective lower bounds of these broader utilization measures. Rather than constructing these alternative natural rates, in the following we will study how well the standard unemployment rate does as a signal for the broader utilization measures. This approach is motivated by the fact that prior to the Great Recession the standard unemployment rate was widely accepted as the relevant measure of labor market utilization. If, following the Great Recession, we now believe that a broader utilization measure is more appropriate, we would like to know how closely the standard unemployment rate was correlated with the broader measure prior to 2007 and in what way the relation between the standard unemployment rate and the broader measure broke down after 2007.

3. NARROW AND BROAD MEASURES OF UNEMPLOYMENT AFTER 2007

Pointing to the exceptionally large increase of discouraged workers and those working PTfER after the Great Recession, it is often argued that the standard unemployment rate understates the degree of resource underutilization for this period. We now argue that while this may be true for the BLS measure U6, for nonemployment measures that account for differences in workforce attachment the standard unemployment rate actually overstates "true" unemployment for this period.

In Figure 2 we plot monthly data of the standard unemployment rate U3 against various broader measures of unemployment for the

Figure 2 The Unemployment Rate as a Signal of Labor Market Utilization, 1994–2014



Notes: All panels plot the standard unemployment rate U3 on the vertical axis against alternative measures of labor market utilization on the horizontal axis. In the first column the alternative measures are on the first row the extended BLS unemployment rate U5, on the second row the NEI based on weighted BLS nonemployment categories, and on the third row the NEI based on weighted LFS histories. The second column adds those working part time for economic reasons, unweighted in the first row (U6) and weighted for the NEIs in the second and third rows. The sample period is 1994 to 2014 for monthly data.

period 1994 to 2014.[17] The rows represent our different broad measures of unemployment, U5, NEI1, and NEI2, and the right columns add

[17] Scatterplots for annual averages of the monthly unemployment and nonemployment rates have the same qualitative features, but the structural breaks estimated in Table 3 are no longer statistically significant.

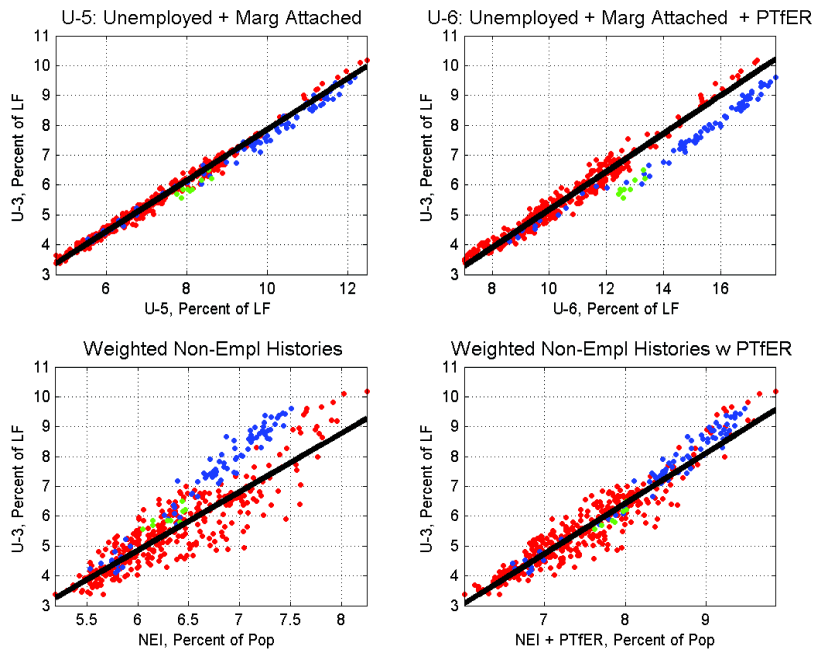
those working PTfER to these broader measures. For each panel we plot the fitted line for a regression of U3 on the relevant broad measure of unemployment for the sample period 1994 to June 2007, represented by the red dots in the different panels. This sample represents the period when presumably there was a close relationship between the standard unemployment rate U3 and the alternative broader measures of unemployment. If the actual U3 unemployment rate for the period after June 2007 is consistently below (above) the fitted line for the pre-2007 sample, then we would say that U3 understates (overstates) true unemployment relative to the pre-2007 relation. For the post-2007 period, we distinguish between the months from July 2007 to December 2013, blue dots, and the year 2014, green dots, the most recent period.

A close relationship between U3 and the extended BLS unemployment rates for the time prior to June 2007 is apparent in the top row of Figure 2, somewhat less so for U6 than for U5. However, for most of the period after June 2007, U3 is consistently below what would have been predicted based on U6 for the pre-2007 period but not so much for U5. Given that including marginally attached workers in U5 does not have much of an impact, the break in U6 is indeed almost exclusively attributable to the exceptional increase of those working PTfER. Since the increase of those working PTfER has persisted into 2014, U3 continues to understate unemployment relative to pre-2007.

Proceeding now to our nonemployment indices we also find a close relationship between them and U3 for the pre-2007 period, somewhat less so for NEI2 based on LFS histories than for NEI1 based on BLS nonemployment categories. Contrary to the extended BLS unemployment rates, we find that for the post-2007 period U3 actually overstates unemployment relative to the NEIs that exclude those working PTfER. This break relative to the pre-2007 relation is due to the exceptionally large increase of long-term unemployment following the Great Recession. Since our NEIs down-weight long-term unemployed significantly relative to short-term unemployed, the NEIs increase less than U3 after the Great Recession. Including those working PTfERs in the NEIs then reduces the overstatement of U3 after 2007, since the exceptional increase in those working PTfER compensates for the exceptional increase in long-term unemployment. As of 2014, however, observations on U3 appear to be consistent with the pre-2007 relationship between U3 and any of our NEI.

The magnitude of nonemployment after 2007 for any of our measures is exceptional relative to the time period from 1994 to 2007. It is therefore not obvious that the relationship between U3 and broader measures of unemployment can be extrapolated from the pre-2007 period. While the extended BLS measures of unemployment and the NEI

Figure 3 The Unemployment Rate as a Signal of Labor Market Utilization, 1976–2014



Notes: All panels plot the standard unemployment rate U3 on the vertical axis against alternative measures of labor market utilization on the horizontal axis. In the first column the alternative measures are on the first row—our estimate of the extended BLS unemployment rate (U5), and on the second row the NEI based on weighted LFS histories. The second column adds those working part time for economic reasons, unweighted for U6 and weighted for the NEI. The sample period is 1976 to 2014 for monthly data.

that is based on BLS nonemployment categories are only available from 1994 on, we can construct the NEI that is based on LFS histories for the years from 1976 on, a period that contains unemployment rates that are comparable to the unemployment rates following the Great Recession. In Figure 3 we plot the standard U3 unemployment rate

Table 3 Post-2007 Bias of the U3 Unemployment Rate

BLS Extended Unemployment Rates, 1994–2014	
U5	0.02 (0.02)
U6	−0.28 (0.05)
NEI Based on BLS Nonemployment Categories, 1994–2014	
Without WPfER	0.31 (0.05)
With WPfER	0.02 (0.05)
NEI Based on LFS Histories, 1994–2014	
Without WPfER	0.47 (0.09)
With WPfER	−0.15 (0.07)
NEI Based on LFS Histories, 1976–2014	
Without WPfER	0.96 (0.07)
With WPfER	0.15 (0.05)

Notes: Coefficients c for a structural break in June 2007 in the OLS regression $U3(t) = a + b * X(t) + c * B(t)$ where $B(t)$ is 1 after June 2007 and 0 before, and $X(t)$ is a broad measure of nonemployment as indicated in the subheaders and row titles. The regression is performed on monthly data. The break coefficients are in percentage points with standard error in parentheses. NEI = nonemployment index as described in the article. WPfER = working part time for economic reasons.

against our versions of the extended BLS unemployment rates and the NEI based on LFS histories for the sample period from 1976 to 2014.[18]

The qualitative features of Figure 3 for the period following the Great Recession are the same as in Figure 2. Relative to the pre-2007 period, the standard unemployment rate U3 understates "true" unemployment for the BLS extended unemployment rates and overstates "true" unemployment for the nonemployment index from 2007 to 2013. More recently, in 2014 U3 has been well in line with the NEIs but it continues to understate unemployment relative to U6.

We can formalize our discussion by simply running a linear regression of the standard unemployment rate U3 on the various broader measures of unemployment for the full sample while allowing for a structural break in the middle of 2007. In Table 3 we report the coefficient of the parallel shift term of the relationship between U3 and the broader measures of unemployment. Relative to the pre-2007 period,

[18] Since information on marginally attached OLF is not available prior to the 1994 comprehensive revision of the CPS, we approximate the marginally attached nonemployed with the LFS history group that is currently OLF and was not employed in the last two months. For the time period from 1994 to 2007 when both series are available, the extended unemployment rates U5 calculated using either the marginally attached or the OLF without recent employment are closely aligned. Following Polivka and Miller (1998), the number of those working PTfER is scaled by a factor of 0.806 prior to the 1994 CPS redesign.

U3 is "understated" by about 0.3 percentage points for the extended BLS U6 unemployment rate, whereas it is "overstated" for the NEIs by up to one percentage point in the case of NEI2 for the sample 1976–2014.

4. CONCLUSION

All the measures of resource utilization in the labor market that we review in this article suggest that as of 2014 nonemployment has declined since the peak in 2010. In particular, even though the standard unemployment rate is still above its 2007 level, it has declined significantly. The decline in the standard unemployment rate is occasionally discounted because extended measures of unemployment that include those working part time for economic reasons seem to suggest that, following the Great Recession, the standard unemployment rate has understated "true" unemployment. In our view broader measures of nonemployment need to account for the heterogeneity in workforce attachment of the nonemployed. Extended measures of unemployment rates provided by the BLS do not. We have constructed such alternative measures of nonemployment and find that for most of the years following the Great Recession the standard unemployment rate actually overstated "true" unemployment and that as of 2014 the standard unemployment rate provides a reasonably accurate measure of "true" unemployment.

APPENDIX

Data for the BLS unemployment rates have been downloaded from Haver. The time series for the CBO estimate of the natural rate of unemployment has been downloaded from FRED. Data for the population shares and employment transition rates for nonemployment by reason and LFS history are from Kudlyak and Lange (2014).

REFERENCES

Appelbaum, Binyamin. 2014. "Still Needed: Millions of Jobs." *The New York Times*, April 4.

Bregger, John E., and Steven E. Haugen. 1995. "BLS Introduces New Range of Alternative Unemployment Measures." *Monthly Labor Review* 118 (October): 19–26.

Diamond, Peter. 2013. "Cyclical Unemployment, Structural Unemployment." Cambridge, Mass.: National Bureau of Economic Research Working Paper 18761 (February).

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin. 2013. "On the Importance of the Participation Margin for Labor Market Fluctuations." Federal Reserve Bank of San Francisco Working Paper Series 2013-05 (February).

Eriksson, Stefan, and Dan-Olof Rooth. 2014. "Do Employers Use Unemployment as a Sorting Criterion when Hiring? Evidence from a Field Experiment." *The American Economic Review* 104 (March): 1,014–39.

Flinn, Christopher J., and James J. Heckman. 1983. "Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States?" *Journal of Labor Economics* 1 (January): 28–42.

Fujita, Shigeru. 2014. "On the Causes of Declines in the Labor Force Participation Rate." Federal Reserve Bank of Philadelphia Research Rap Special Report (February).

Hall, Robert E., and Sam Schulhofer-Wohl. 2013. "Measuring Matching Efficiency with Heterogeneous Jobseekers." Mimeo.

Hornstein, Andreas. 2012. "Accounting for Unemployment: The Long and Short of It." Federal Reserve Bank of Richmond Working Paper 12-07 (November).

Hornstein, Andreas, Marianna Kudlyak, and Fabian Lange. 2014. "A New Measure of Resource Utilization in the Labor Market." Federal Reserve Bank of Richmond, Mimeo (April). Available at www.richmondfed.org/research/economists/bios/pdfs/hornstein_new_measure_resource_utilization.pdf.

Hornstein, Andreas, Marianna Kudlyak, Fabian Lange, and Tim Sablik. 2014. "Does the Unemployment Rate Really Overstate Labor Market Recovery?" Federal Reserve Bank of Richmond *Economic Brief* (June).

Jones, Jerry, Michael Joyce, and Jonathan Thomas. 2003. "Non-Employment and Labour Availability." *Bank of England Quarterly Bulletin* Autumn: 291–303.

Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *Quarterly Journal of Economics* 107 (February): 35–78.

Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo. 2013. "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment." *Quarterly Journal of Economics* 128 (3): 1,123–67.

Kroft, Kory, Fabian Lange, Matthew J. Notowidigdo, and Lawrence F. Katz. 2013. "Long-term Unemployment and the Great Recession: The Role of Composition, Duration Dependence, and Non-participation." Available at http://scholar.harvard.edu/files/lkatz/files/klnk_ltu_and_great_recession_resubmit.pdf.

Krueger, Alan B., Judd Cramer, and David Cho. 2014. "Are the Long-Term Unemployed on the Margins of the Labor Market?" *Brookings Papers on Economic Activity* (Spring).

Kudlyak, Marianna, and Fabian Lange. 2014. "Measuring Heterogeneity in Job Finding Rates among the NonEmployed Using Labor Force Status Histories." Federal Reserve Bank of Richmond Working Paper 14-18.

Madrian, Brigitte C., and Lars John Lefgren. 1999. "A Note on Longitudinally Matching Current Population Survey (CPS) Respondents." Cambridge, Mass.: National Bureau of Economic Research Working Paper t0247 (November).

Polivka, Anne E., and Stephen M. Miller. 1998. "The CPS After the Redesign: Refocusing the Economic Lens." In *Labor Statistics Measurement Issues; Studies in Income and Wealth,* Volume 60, edited by John Haltiwanger, Marilyn E. Manser, and Robert Topel. Cambridge, Mass.: National Bureau of Economic Research, 249–89.

Schweitzer, Mark E. 2003. "Ready, Willing, and Able? Measuring Labour Availability in the UK." Federal Reserve Bank of Cleveland Working Paper 0303.

Shimer, Robert. 2012. "Reassessing the Ins and Outs of Unemployment." *Review of Economic Dynamics* 15 (April): 127–48.

Veracierto, Marcelo. 2011. "Worker Flows and Matching Efficiency." Federal Reserve Bank of Chicago *Economic Perspectives* 35 (Fourth Quarter): 147–69.

Yellen, Janet L. 2014. "Labor Market Dynamics and Monetary Policy." Speech held at the Federal Reserve Bank of Kansas City Economic Symposium, Jackson Hole, Wyo., August 22.

Economic Quarterly—Volume 100, Number 1—First Quarter 2014—Pages 23–50

The Financial Crisis, the Collapse of Bank Entry, and Changes in the Size Distribution of Banks

Roisin McCord and Edward Simpson Prescott

The recent financial crisis has had an enormous impact on the banking industry. There were numerous bank failures, bank bailouts, and bank mergers. One of the more striking effects was the decline in the number of banks. At the end of 2007, as the recent financial crisis was developing, there were 6,153 commercial banks in the United States. At the end of 2013, as the direct effects of the crisis were wearing off, the number of banks had dropped 14 percent, reaching 5,317.[1]

The purpose of this article is to document the size and scope of these recent changes to the size distribution of banks, particularly among the smaller banks, and explain the sources of these recent changes. In doing so, we also update the work of Janicki and Prescott (2006), who studied the size distribution in the banking industry from 1960–2005.

Our most significant finding is that the recent decline in the number of banks is *not* due to exit from banking. Despite the financial crisis, the exit rate—the percentage of active banks that disappeared due to

We would like to thank Huberto Ennis, Joe Johnson, Thomas Lubik, and John Weinberg for helpful comments. The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: Edward.Prescott@rich.frb.org.

[1] There were even larger percentage declines in the number of savings institutions (savings and loans and savings banks) and credit unions. In 2007, there were 1,250 savings institutions and 8,268 credit unions. In 2013, there were 936 savings institutions and 6,687 credit unions, drops of 25 percent and 19 percent, respectively. (Sources: Savings institution numbers are from the Federal Deposit Insurance Corporation [FDIC] State Banking Performance Summaries. Credit union numbers are from NCUA Quarterly Call Reports.)

failure or merger with another bank—over the period 2008–2013 is not that different from 2002–2007. There are significant differences in how banks exited—in the earlier period virtually all of the exit was due to acquisitions and mergers, while in the later period there were also many failures—but mechanically it is the number of exits, not the reason for them, that matters for calculating the total number of banks.

Instead, nearly two-thirds of the recent decline is due to the collapse of entry into commercial banking. Very few new banks have started since 2008 and most of these are thrifts or credit unions changing their charter or, in a smaller number of cases, banks that were spun out of multi-bank holding companies. Entry by newly created banks, commonly called *de novo* banks, has been minimal and was actually zero in 2012. This is unprecedented over the last 50 years. Even during the previous banking crisis of the late 1980s and early 1990s when large numbers of banks failed or merged, there was still substantial entry.

The recent lack of entry has large implications for the number of banks and bank size distribution. Most new banks start small, so without that flow into banking, the number of small banks will decline. Indeed, we find that the biggest drop is in the smallest size class, those with less than $100 million in assets, and that two-thirds of this decline can be attributed to the lack of entry. This drop is of potential concern because small banks are considered to have a comparative advantage in small business, relationship-type lending (Berger and Udell 2002). For better or worse, a drastic change in the bank size distribution could have an impact on the allocation of credit to different sectors in the economy.[2]

To demonstrate the importance of entry for the future number of banks, we provide forecasts of the number of banks based on different assumptions about entry rates and show how these depend on the degree to which entry recovers to historical rates. Finally, we discuss various reasons for why entry has been so low.

1. DATA

Historically, in the United States there have been many legal and regulatory limits on bank size. For example, in the 1960s banks could not branch across state lines, and in some states banks were required to be unit banks, that is, they could not even have a branch. These limits were removed gradually starting in the 1970s, more rapidly in the 1980s, and mostly eliminated in the 1990s with the Riegle-Neal

[2] Bank size distribution should have an effect on bank productivity as well, but it is difficult to measure bank productivity.

Interstate Banking and Branching Efficiency Act of 1994. This law allowed bank holding companies to acquire banks in different states and allowed interstate bank mergers.[3]

A bank holding company is a company that directly or indirectly owns at least 25 percent of a bank's stock, controls the election of a majority of a bank's directors, or is deemed to exert controlling influence over a bank's policy by the Federal Reserve (Spong 2000). Often, a bank holding company will have multiple banks—or even another bank holding company—under its control. Historically, this structure was used to avoid some of the restrictions on bank branching (Mengle 1990) while still allowing the bank holding company to jointly manage many activities. For this reason, we follow Berger, Kashyap, and Scalise (1995) and treat all banks and bank holding companies under a bank holding company as a single banking entity. For convenience, we will call one of these entities a bank.

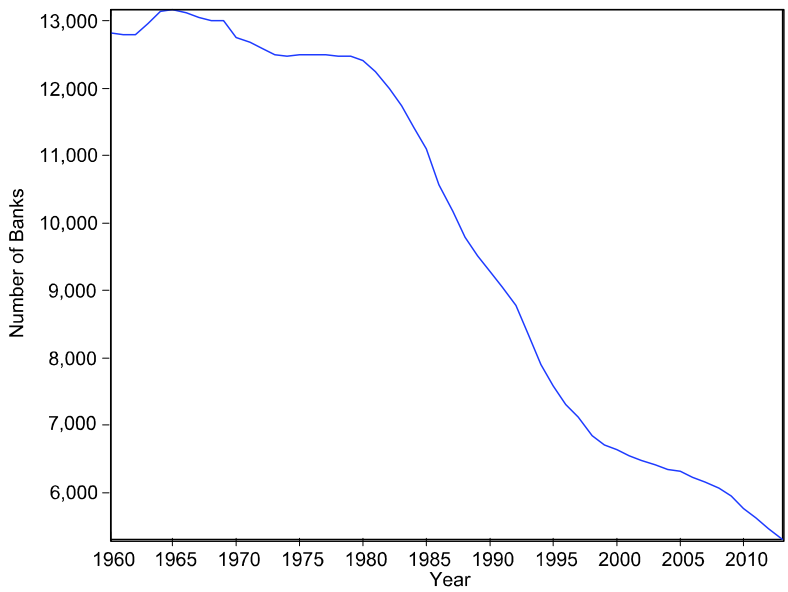
Bank structure and bank size data are measured at the end of each year from 1960–2013. Data on bank structure are taken from the Federal Reserve's National Information Center bank structure database. We only include commercial banks and exclude savings and loans, savings banks, and credit card banks.

Bank size data comes from the Reports on Condition and Income (the "Call Report"), which is collected by federal bank regulators. Bank size is measured by assets, though in a few places we use additional size measures. For the analysis, assets are also adjusted by off-balance sheet items starting in 1983. Starting in that period, banks, and larger banks in particular, began to undertake numerous activities like providing lines of credit, supporting securitizations, and issuing derivatives that expose a bank to risk but are not reported on a traditional balance sheet. These adjustments significantly increase the size of the largest banks. The Appendix contains more information on these adjustments.

To facilitate comparison of bank size across years, we report size measures relative to 2010 dollars. Data in other years are scaled by the change in total bank assets between those years and 2010. The resulting number is essentially a market share number, but scaled by the size of the commercial banking industry in 2010. For example, total bank assets in 2000 were 50.5 percent of total bank assets in

[3] See Jayaratne and Strahan (1997) for a history of bank branching restrictions and Kane (1996) for a description of the Riegle-Neal Act.

Figure 1 Total Number of Independent Banks



Notes: All banks and bank holding companies that are under a higher-level holding company are treated as a single independent bank. A more precise definition of an independent bank is given in Section 1.

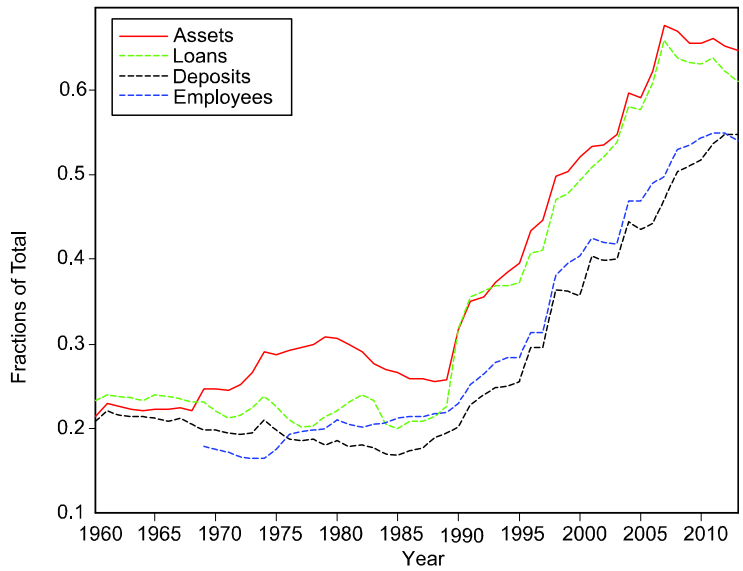
2010. Consequently, we roughly double the size of a bank in 2000 to make it comparable to a bank in 2010.[4]

2. CHANGES IN BANK SIZE DISTRIBUTION

Figure 1 shows the number of banks from 1960 through 2013. Several distinct periods are apparent in the graph. From 1960 to 1980, the number of banks is relatively stable. There is a drop in the early 1970s, which overlaps with the sharp recession of 1973–1975, but compared with future changes this drop is proportionally small. The most dramatic changes start in 1980 and last through the late 1990s. This

[4] An alternative way for scaling the data would be to use a price index like the consumer price index. We do not use this measure because that price index was designed to measure changes in the price of goods and we are interested in changes to the size of a bank's balance sheet, not what it charges to provide bank services. Furthermore, there have been much larger changes in total assets in the banking industry than in price levels.

Figure 2 Market Share of 10 Largest Commercial Banks



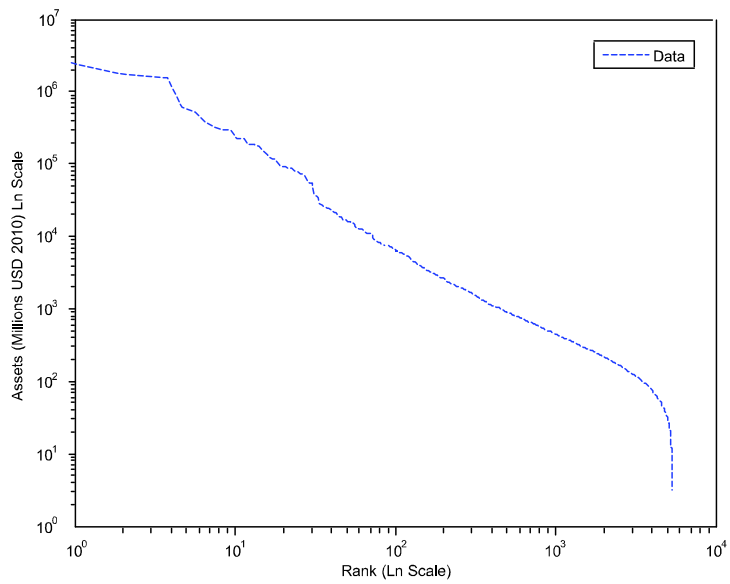
Notes: Market share of the 10 largest commercial banks for four different measures. The number of employees is only reported starting in 1969 because the Call Report did not collect that information until then.

is the era when many regulatory restrictions were removed from bank branching and interstate banking, and there was a commercial banking crisis in the 1980s and early 1990s when many banks failed. These factors led to a large amount of consolidation through both merger and failure. Starting in the late 1990s, however, the decline continues, but the rate of decline slows down. This trend lasts until about 2005, before the crisis, and then the numbers begin to rapidly decline again.

A second phenomenon associated with the latter period of bank consolidation is an increase in concentration, particularly for the largest banks. Figure 2 shows the market share of the 10 largest banks for four different measures of firm size. Interestingly, the big increase in concentration starts around 1990 and continues until the financial crisis, at which point it levels off.

Like many industries, the size distribution of banks consists of a large number of small firms and a small number of large ones. One class of distributions that is often used to fit the size distribution of

Figure 3 Zipf Plot for Bank Size Distribution in 2013



firms is one that is based on a power law, that is, it satisfies the relation

$$f(x) = cx^{-\alpha},$$

where $c > 0$. Power laws also describe a large number of other empirical phenomena in economics as well as in the natural sciences.[5]

In this article, we will look at the data with a Zipf plot, or rank-frequency plot. In our context, this means we rank banks by size and then plot the log of the rank versus the log of the size of the bank. If this relationship is linear, then it satisfies a power law because

$$y_r = cr^{-\alpha},$$

where r is the rank of a bank measured by size and y_r is the size of the rth largest bank. Furthermore, when $\alpha = 1$ (or is close to it), the data is said to satisfy Zipf's Law, that is, size is inversely proportional to rank. In other words, the largest bank would be twice the size of the second-largest bank, three times the size of the third-largest bank, etc.

[5] For a description of the use of power laws in economics, see Gabaix (2009). For a discussion of their use to applications as diverse as word frequency, population of cities, and earthquake strength, see Newman (2005). For examples of their application to firm size, see Axtell (2001) and Luttmer (2007).

Table 1 Ten Largest Banks

Bank	2007 (billions)	Bank	2013 (billions)
JP Morgan Chase	2,503	JP Morgan Chase	2,518
Bank of America	2,096	Bank of America	1,756
Citigroup	1,824	Citigroup	1,614
Wachovia	904	Wells Fargo	1,519
Bank of New York Mellon	823	Bank of New York Mellon	600
State Street	708	State Street	523
Wells Fargo	580	U.S. Bancorp	386
U.S. Bancorp	290	PNC	323
HSBC Holdings	277	Capital One	298
Northern Trust	258	Goldman Sachs	292

Notes: Size of the 10 largest banks measured by assets, expressed in 2010 dollars. The asset measure includes off-balance sheet conversions and only includes activities under the banks' charters.

Janicki and Prescott (2006) found that Zipf's Law did an excellent job of fitting the size distribution of banks in 1960 and 1970, but starting in 1980 it underpredicts the size of the largest banks.

Figure 3 shows the Zipf plot for 2013. The graph suggests that Zipf's Law still underpredicts the size of the largest banks and, furthermore, there are different ranges of the size distribution where bank size is proportional to rank, but these proportions differ along different segments of the size distribution. Furthermore, it is obvious that the size distribution of the smallest banks is poorly described by a power law and therefore needs to be described by some other distribution.[6]

Interestingly, despite the severity of the financial crisis, the Zipf plot for 2007 (not shown) looks virtually identical to Figure 3. One reason is that changes among the distribution of smaller banks are hard to see in the curve and, as we will see, there were significant changes there. However, the other reason is that there were not significant changes in concentration among the largest banks. This is apparent in Figure 2, which shows that the market share of the 10 largest banks levels off after the crisis.

[6] It is common in applications that the bottom part of the distribution is not well described by a power law distribution, so scientists typically leave this part out of their analysis. For example, when looking at bank size distribution, Janicki and Prescott (2006) only consider the largest 3,000 banks when they assess how Zipf's Law fits the size distribution of banks. Recent work by Goddard et al. (2014) develops a more general formulation by fitting a distribution in which there is a power law for the largest banks, a lognormal distribution for small banks, and an endogenous cutoff between the two classes of banks. See also Goddard, Liu, and Wilson (2014) for an analysis of bank growth rates.

Table 2 Drop in Number of Banks by Size Class

Size Class (millions)	2007	2013	Change	% Change
< 100	2,538	1,771	−767	−30.2
100–500	2,706	2,634	−72	−2.7
500–1,000	455	453	−2	−0.0
1,000–5,000	338	333	−5	−1.5
5,000–10,000	48	50	2	4.2
10,000–50,000	39	44	5	12.8
> 50,000	29	32	3	10.3
Total	6,153	5,317	−836	−13.6

While the size distribution among the largest banks did not change much, there were significant changes among the relative size of the largest banks. Table 1 lists the size of the largest 10 banks in 2007 and 2013. The top three largest banks did not change, but Wachovia ceased to exist after being acquired by Wells Fargo. Northern Trust and HSBC exited the top 10 list, while PNC, Capital One, and Goldman Sachs entered it.

There are two features of these numbers worth noting. First, off-balance sheet activities have a large effect on the size of some of these firms. For example, Wachovia is listed as having about $900 billion in assets in 2007. Nearly a third of that number ($269 billion) came from the off-balance sheet adjustments.[7] See Appendix A for figures showing how big this adjustment is for the banking sector as a whole. Second, by using Call Report data we are only measuring assets (and off-balance sheet assets) that are held under a bank holding company's commercial bank charters.[8] For some financial institutions, this matters. For example, most of Goldman Sachs' activities are done outside its bank charter. In 2013, its balance sheet was about $912 billion (FR Y-9C), which is much larger than the $292 billion reported in Table 2. For others it is less important. A traditional commercial bank like Wells Fargo has most of its assets under its commercial bank charters.

The largest changes in the bank size distribution have occurred among smaller banks, which is something that the Zipf plot does not show that well. Consequently, we break banks into size classes and look at the number of banks in each class. Table 2 reports these

[7] The four largest off-balance sheet equivalents for Wachovia were unused loan commitments with an original maturity exceeding one year ($74 billion), securities lent ($59 billion), derivatives ($50 billion), and financial standby letters of credit ($40 billion).

[8] For an analysis of how the activities of large bank holding companies have changed over the crisis, see Ennis and Debbaut (2014).

Table 3 Drop in Number of Small Banks by Size Class

Size Class (millions)	2007	2013	Change	% Change
< 50	1,230	725	−505	−41.1
50–100	1,308	1,046	−262	−20.0
100–200	1,407	1,357	−50	−3.6
200–300	687	678	−9	−1.3
300–400	359	372	13	3.6
400–500	253	227	−26	−10.3
500–750	290	296	6	2.1
750–1,000	165	157	−8	−4.8

numbers. Not surprisingly, the biggest drop in the number of banks is in the smallest class of banks because the majority of banks are small. More interesting, however, is the percentage change. The biggest such change is in banks that hold less than $100 million in assets. The drop in this size class is about 30 percent in just five years. This is an extraordinarily large decline. In the next three size classes, the number of banks does not change that much, while there are increases in the three largest categories.[9]

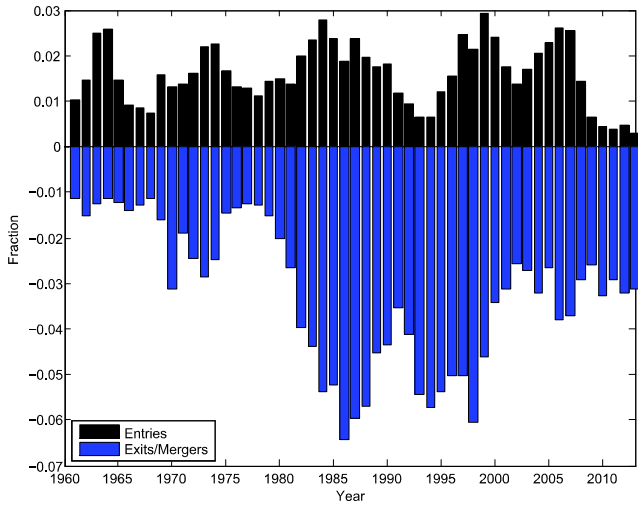
A closer look at banks that hold less than $1 billion further illustrates that the smallest banks are disappearing. Table 3 breaks down the size classes even further. There is an enormous drop of about 40 percent in the number of banks that hold less than $50 million. In the $50–$100 million range, there is a smaller, but still large, percentage drop of 20 percent. Above $100 million, the change is more mixed. In some categories, the number of banks increases and in others it decreases.

3. ENTRY AND EXIT

The recent decline in the number of banks shown in Figure 1 appears to be a continuation of a trend that started around 1980 and, when measured solely by the number of banks, that view would be correct. However, there is a significant difference from any previous period. Figure 4 reports the number of entries and exits into commercial banking expressed as a fraction of the banking population.

[9] To check the robustness of this result we also performed this analysis on other measures of bank size including on-balance sheet assets, deposits, and loans, both scaled and unscaled (nominal). Qualitatively, the results were similar for all these measures except for scaled loans.

Figure 4 Banks that Enter and Exit by Fraction of Total Banks



Notes: Entry and exit of banks expressed as a percentage of banks existing in each year.

The most striking observation from Figure 4 is the unprecedented collapse of bank entry since 2009. Entry rates are on the order of 0.05 percent, which is much smaller than the long-term average of 1.5 percent. Furthermore, as we will see in the next section, entry is actually weaker than these numbers indicate. The only period that is at all close to this is 1993 and 1994, which followed the previous banking crisis and the recession of the early 1990s.

The other striking observation from Figure 4 is that despite large numbers of exits in different periods, like the mid-1980s and the mid-1990s, entry was usually strong. For example, in 1984, when more than 5 percent of banks exited because of failure or merger, there were so many entrants that they equaled 3 percent of the banks that operated at the beginning of that year. The late 1990s were similar. During the merger wave of that period, there was a lot of entry.

It is also apparent from Figure 4 that despite the financial crisis, exit rates during the crisis are very similar to those from the 2002–2007 pre-crisis period. The one significant difference between these periods

Table 4 Commercial Bank Exit by Reason since 2002

Year	Total Exits	Failures	Acquisition/Mergers
2002	169	7	162
2003	176	1	175
2004	206	3	203
2005	169	0	169
2006	240	0	240
2007	232	1	231
2008	180	17	163
2009	158	98	60
2010	195	126	69
2011	168	80	88
2012	181	37	144
2013	171	18	153

Notes: Failed banks were obtained from the FDIC's Historical Statistics on Banking and then compared with our calculated list of exits. Banks that did not fail were treated as an acquisition/merger. Because we are measuring a bank at the holding company level and multiple failed banks can be part of the same holding company, we report fewer failures than the FDIC.

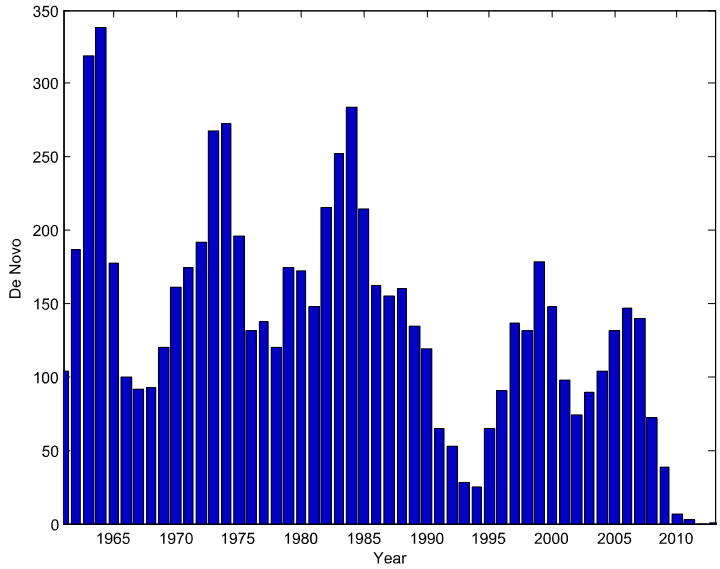
is the reason for exit. Table 4 lists bank exits by reason from 2002–2013. Before the crisis, almost all exit was due to an acquisition or merger while, during 2009–2010, failure was the most common reason for exit. Starting in 2011, failure accounts for about half of all exits, after which the rate of failure quickly declines.

The entry and exit rates demonstrate that the normal dynamics of the banking industry are *not* such that there is a fixed stock of banks from which banks exit over time. Instead, it is of a dynamic industry with lots of entry and exit in both good and bad economic times. By these perspectives, the collapse of entry is what is so striking about the last few years.

4. A DEEPER LOOK INTO ENTRY (OR THE LACK THEREOF)

A deeper look into the source of entry implies that entry in recent years is actually weaker than the numbers suggest. In our data, we can identify three distinct types of entry. First, there is a charter conversion, that is, a savings and loan, a savings bank, or a credit union that changes its charter to a commercial banking charter. Second, there is a spinoff, which is a bank that was formerly part of a holding company but has become independent. Third, there is a *de novo* entrant, which is a newly formed bank.

Figure 5 Number of De Novo Entrants by Year



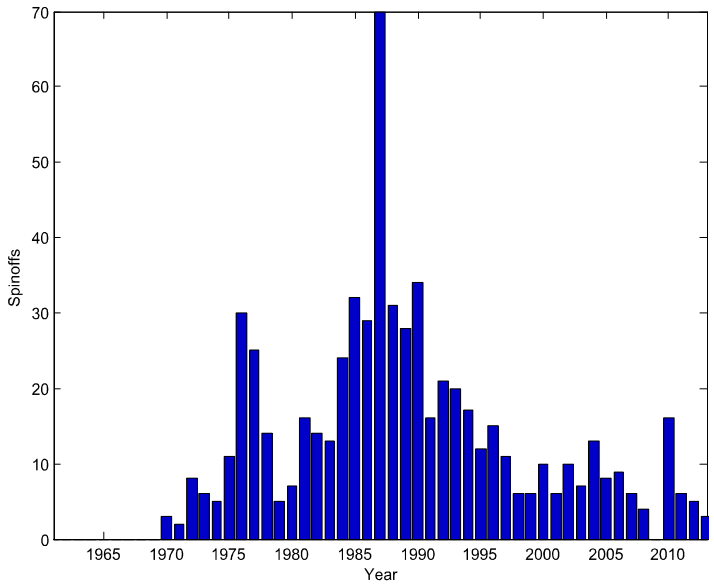
Notes: A *de novo* bank is a newly formed bank.

A *de novo* bank is a good measure of interest in entering banking because it represents new capital, new management, and a new organization. A charter conversion to a degree is just a relabeling of an existing institution since there is overlap between the activities of a commercial bank and other depository institution charters. Similarly, a spinoff is just another way of legally organizing bank assets and managers that are already in the banking sector.

Figure 5 lists the number of *de novo* entries for each year since 1960. The only two periods in which there is a sharp decline in the number of *de novo* banks are the early 1990s and the last few years. The former period coincides with the recession of the early 1990s and the end of a commercial banking crisis, but *de novo* entry numbers quickly rebound. In contrast, the *de novo* entry numbers in the recent period are truly abysmal. In 2011, there were three *de novo* banks;[10] in 2012, there were zero, and in 2013, there was only one. This last

[10] These three banks were Alostar, Cadence, and Certusbank, which were all formed to acquire failed banks.

Figure 6 Number of Spinoffs by Year



Notes: A spinoff is a newly independent bank that used to be part of a bank holding company. We identify a spinoff by taking the bank ID of each new entrant and seeing if that ID was a bank that was in a holding company in the previous year.

one was Bank of Bird-in-Hand, which was formed in Lancaster County, Pa., to serve the Amish community.

Spinoffs are unusual and to our knowledge have not been studied in the banking literature. There are several reasons for why a bank holding company might undertake one. One reason is that a bank holding company might sell one of its healthy bank charters to outside investors because the holding company is in financial trouble. For example, in 2012 the bank holding company Capital Bancorp sold several of its banks to local investors while it filed for Chapter 11 bankruptcy (Stewart 2012). A second reason is that management thinks the bank will be better managed separately rather than jointly. For example, in 2005 Midwest Bank Holdings sold one of its subsidiaries, Midwest Bank of Western Illinois, to local managers and investors because the bank's agricultural lending focus did not fit well with the holding company's Chicago growth strategy (Jackson 2005).

Table 5 Commercial Bank Entry by Type since 2002

Year	De Novo	Spinoff	Conversion
2002	74	10	6
2003	90	7	10
2004	104	13	12
2005	132	8	3
2006	147	9	5
2007	140	6	8
2008	72	4	10
2009	38	0	0
2010	7	16	3
2011	3	6	12
2012	0	5	20
2013	1	3	11

Figure 6 reports the number of spinoffs by year for our data set. In general, spinoffs are unusual, though there was a spike in the mid-1980s and there were 16 in 2009.

The final type of entry that we can identify is a charter conversion. A depository institution may want to switch charters because it wants to expand certain types of lending (e.g., savings and loans and credit unions face limits on the type of lending that they do). Table 5 shows entry by type since 2002, and this makes clear that most entries since 2011 came from charter conversions.

5. DECOMPOSING THE DROP IN THE NUMBER OF BANKS

The two trends we have identified—the decline in the number of small banks and the collapse of entry—are related. As we emphasized earlier, the dynamics of bank growth matter for the size distribution. In particular, the pool of small banks changes over time. Some grow to a new size class and some exit. These factors alone would reduce the number of small banks, so the flow into this pool matters a lot. For the smallest class of banks, *de novo* banks are a critical part of the flow in. Many of these banks start small, so they replenish the stock of small banks, even as other ones are leaving that class.

We can get a sense of just how much the recent decline in small banks is due to the drop in bank entry by running a simple counterfactual. We break banks into the seven size classes of Table 2, calculate the fraction of banks in each size class that move to another size class

Table 6 Annual Transition Probabilities Between Size Classes for 2008–2013

Size Class (millions)	**Exit**	**< 100**	**100–500**	**500–1,000**	**1,000–5,000**	**5,000–10,000**	**10,000–50,000**	**> 50,000**
< 100	0.03	**0.91**	0.06	0.00	0.00	0	0	0
100–500	0.03	0.02	**0.93**	0.02	0.00	0	0	0
500–1,000	0.04	0.00	0.06	**0.84**	0.05	0	0	0
1,000–5,000	0.04	0.00	0.01	0.04	**0.90**	0.01	0	0
5,000–10,000	0.04	0	0	0	0.03	**0.86**	0.06	0.00
10,000–50,000	0.03	0	0	0	0.01	0.04	**0.92**	0.01
> 50,000	0.01	0	0	0	0	0	0	**0.99**

Notes: Entries in bold reflect the fraction that stay in the same size class. Rows may not add to 1 due to rounding.

in each year, and then take the average over the 2008–2013 period.[11] We then use these transition probabilities along with some counterfactual assumptions on entry to see what would have happened to the number of banks under more typical entry conditions.

Table 6 shows the average annual transition probabilities for the 2008–2013 period. Each row takes all the banks in a given size class and reports the fraction of them that are in each size class in the succeeding year. For example, of banks that are less than $100 million, 3 percent exited, 91 percent stayed in the same size class, and 6 percent moved up to the next highest size category.[12]

For our counterfactual experiment, we take the number of banks in each size category in 2007 (column 2 in Table 2) and multiply this by the transition probabilities. For entry, we take the average entry rate over the 2008–2013 period and—for the counterfactual part—we add enough additional entrants so that the number of entrants equals 129, which was the average number of new entrants over the 2002–2007 period. We put these entrants into size categories in the same proportion as new entrants during the 2002–2007 period.[13]

Table 7 reports the number of banks in each size category for 2013 and the number that would have existed under the counterfactual assumption on entry. It also lists the difference, expressed in absolute and percentage terms. With the counterfactual entry, there would have been 567, or 10.7 percent, more banks. There would still be fewer banks than in 2007, when there were 6,153, but a lot more than the actual 5,317 in 2013. The actual number of banks dropped in this period by 836, while under the counterfactual the number would have only dropped by 269. This means that the weaker entry accounts for the rest of the drop, which is about 68 percent of the total.

Among size classes, the biggest difference among banks is in the less than $100 million size class. In the counterfactual, there are 22 percent more banks. Much of this difference is directly accounted for by the lack of entry. Under the counterfactual entry assumptions, 129 banks enter per year, and most of them enter the smallest size class. Furthermore, in each year, 91 percent of those new entrants stay in this class, so over time new entry adds a lot of banks to this size class.

---

[11] See Adelman (1958), Simon and Bonini (1958), and Janicki and Prescott (2006) for more information about transition probabilities and how they can be used to assess the dynamics of an industry.

[12] Appendix B contains some more analysis of the transition matrix.

[13] In the 2002–2007 period, 81 percent of new entrants started in the under $100 million size category, 15 percent started in the $100–$500 million size category, 2 percent started in the $500–$1,000 million size category, and 2 percent started in the $1,000–$5,000 million size category.

Table 7 Number of Banks by Size Class with Counterfactual Entry

Size Class (millions)	Data 2013	Counterfactual 2013	Difference	% Difference
< 100	1,771	2,276	505	28.5%
100–500	2,634	2,711	77	2.9%
500–1,000	453	448	−5	−1.1%
1,000–5,000	333	329	−4	−1.2%
5,000–10,000	50	48	−2	−4.0%
10,000–50,000	44	40	−4	−9.1%
> 50,000	32	32	0	0.0%
Total	5,317	5,884	567	10.7%

Notes: Number of banks in each size class in 2013 compared with numbers under the counterfactual assumption that the number of entrants in 2008–2013 is the same as in 2002–2007.

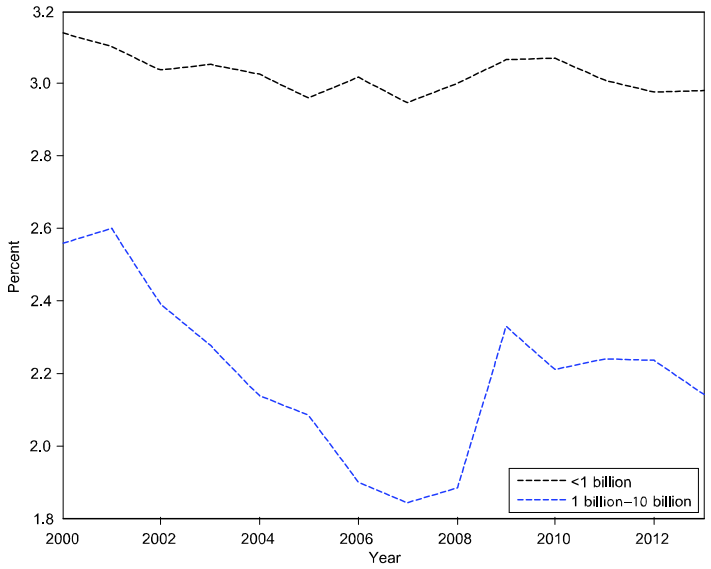
6. WHAT ACCOUNTS FOR THE LACK OF ENTRY?

The literature on bank entry has identified three main factors that are positively correlated with bank entry. The first is that entry is more likely in fast-growing, profitable, and concentrated markets (Dunham 1989; Moore and Skelton 1998), presumably because potential profits are higher in this type of market. The second is that entry is more likely after recent mergers (Dunham 1989; Keeton 2000; and Berger et al. 2004). Starting a bank requires experienced bankers and there are more people available after a merger since mergers often involve layoffs.[14] The third factor is that entry is more likely when regulatory restrictions on entry are relaxed (Ladenson and Bombara 1984; Lindley et al. 1992), presumably because any decrease in entry cost will make it more profitable for a potential bank to enter.

Analysis and discussion of the recent lack of entry have focused on the poor economic conditions and the increase in regulatory compliance costs. The recent economic recovery has been very weak, which has certainly reduced the potential return from entering. Adams and Gramlich (2014) examine entry at the county level with an ordered probit model estimated on U.S. data from 1976 to 2013. Based on this model, they conclude that 75 percent to 80 percent of the decline in bank entry over the last few years is due to low interest rates and a lack of demand for banking services. They point out that community bank profits are heavily dependent on the net interest margin, that is, the

[14] At a longer time horizon, an industry with frequent mergers may create an incentive to start a bank with the goal of selling it in the future.

Figure 7 Non-Interest Expense as a Percentage of Assets for Banks with Less $1 Billion in Assets and with $1 Billion to $10 Billion in Assets

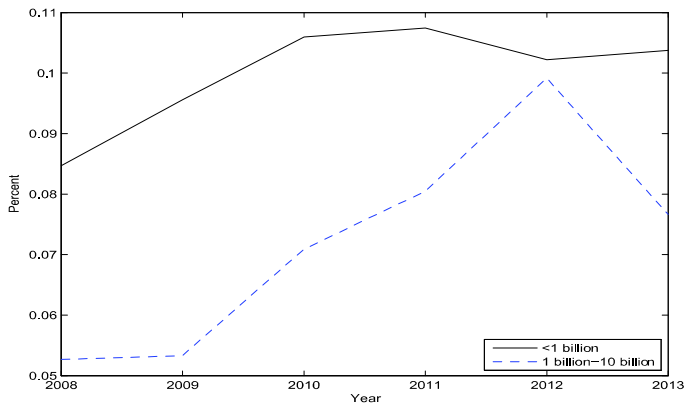


Notes: Nominal value of expenses and assets are used.

spread between deposit rates and lending rates, and with present Federal Reserve monetary policy pushing lending rates down, this margin is relatively small.

While these results are suggestive, they are far from definitive. There are plenty of periods where net interest margins declined, yet entry did not collapse. Morris and Regehr (2014) study the historical pattern of net interest income in community banks after recessions since the mid-1970s. They observe significant drops in this revenue source during all recessions and argue that the recovery in net interest income after the recent recession is not that different from the 2001–2002 recession and is actually higher than in the 1981–1982 recession. Furthermore, as we showed in Figure 4, entry rates were much higher after *every* earlier recession. Indeed, the Adams and Gramlich (2014) model includes a dummy variable for the post-crisis period (2010 and after) that is also important for explaining the recent lack of entry. Their model also predicts that, even if the net interest margin and the economy recovered to 2006 levels, there would still be almost no entry.

Figure 8 Legal Fees, Accounting, Auditing, Consulting, and Advisory Expenses to Asset Ratio



Notes: Combined legal fees and expenses, accounting and auditing expenses, and consulting and advisory expenses measured as a percentage of assets for banks with less than $1 billion in assets and with $1 billion to $10 billion in assets. Nominal values of expenses and assets are used.

It seems then that while the net interest margin is important, there may be other factors at work.

The other line of analysis is that regulatory costs are discouraging entry. There are two distinct, but often mixed together, arguments used here. The first argument is that the general increase in regulations resulting from the implementation of the Dodd-Frank Act of 2010 have made banking significantly more costly by requiring more resources to be used for complying with regulations and that, furthermore, there are economies of scale in complying with these regulations.

Peirce, Robinson, and Stratmann (2014) surveyed community bankers about compliance costs. The bankers responded that their median number of compliance staff increased from one to two.[15] Other than for the smallest banks, this is not a big increase in number of employees, but there are other sources of compliance costs that could be reflected in the non-interest expense category of the Call Report income statement.

[15] For an analysis of potential increases in costs to community banks of hiring additional compliance staff, see Feldman, Schmidt, and Heineche (2013).

Figure 7 shows non-interest expense as a percentage of assets for banks with less than $1 billion in assets and for those with $1 billion to $10 billion in assets. For the smaller class, this ratio did not change much between 2007 and 2013, and while it is higher for the larger class, it is still lower than it was in 2000. If compliance costs are really increasing, then they are being swamped by changes in other expenses.

The non-interest expense number does not break out expenses between compliance and non-compliance costs, but starting in 2008 the Call Report added some subcategories of expenses, including costs related to legal fees, auditing, consulting, and advisory expenses. Presumably, some of these costs are related to the costs of complying with regulations. Figure 8 shows these costs measured as a percentage of assets for banks with less than $1 billion in assets.

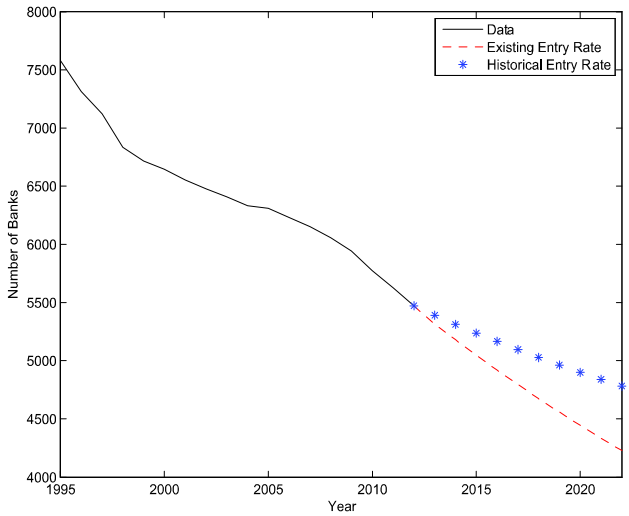
There is an increase in these expenses from 2008 to 2011, but the increase is relatively small and, more importantly, the size of these expenses is just too small to have a big effect on bank profitability. For example, entirely eliminating these expenses would only increase the return on assets by 10 basis points.

Based on this data, if regulatory costs are significantly impacting bank expenses and profitability, it is because other costs are declining to offset the increase or regulatory costs are affecting the operations of banks in such a way that less revenue is being generated. For example, many community bankers say that their leaders spend a lot of their time reading, interpreting, and reacting to the rules, and that for small banks, in particular, this pulls them away from things like making loans and managing their staff.[16] This kind of cost is not something we can measure in the Call Report data.

The second argument related to regulatory change is that the costs of entry have increased due to regulations. To start a bank in the United States, organizers are required to get a banking charter from either a state or the federal government and to obtain deposit insurance from the FDIC. Once the organizers pass these hurdles, the *de novo* bank is under heightened supervision for a period of time. One way in which these costs have gone up is that the intensity of supervision of newly chartered banks has increased. In 2009, the FDIC raised the period from three to seven years under which FDIC-supervised, newly insured depository institutions are subject to higher capital requirements and more frequent examinations. Furthermore, FDIC approval is now required for changes in business plans during this seven-year period (Federal Deposit Insurance Corporation 2009).

[16] Personal conversations with bankers by the second author.

Figure 9 Trends



Notes: Forecasts of the future number of banks assuming existing transition rates between size classes and with existing entry rates and historical entry rates.

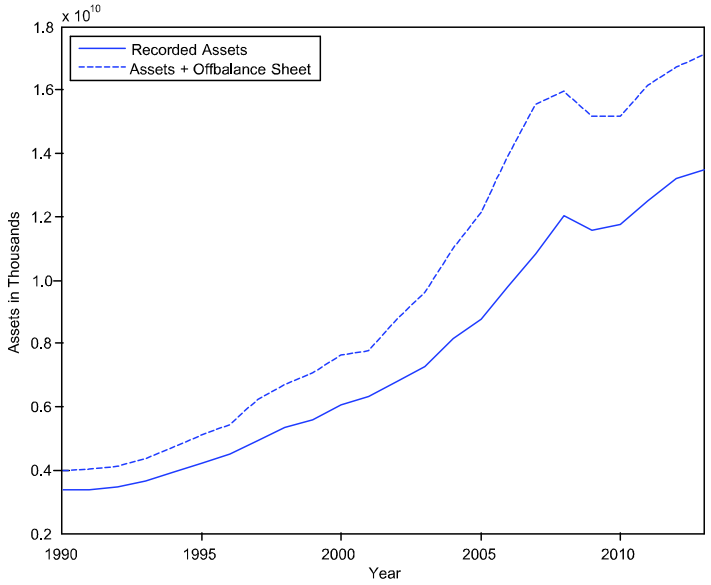
A second way in which these entry costs may have gone up is that the application process has lengthened, become more rigorous, and gotten more expensive. There have been so few *de novo* banks the last few years that there is not much direct data on this cost. However, organizers of the one *de novo* bank in 2013 claim that the application process was significantly longer and more intensive than in the past (Peters 2013).

7. LOOKING AHEAD

The future number of banks will depend on the conditions under which bank entry rates recover. If the main reason for the lack of entry is the low net interest margin, then entry numbers should recover when the economy improves and the Federal Reserve raises interest rates. If regulatory costs are the main reason for the lack of entry, then it will depend on how these change over time.

Regardless of the reason for the lack of entry, until entry recovers (and assuming that exit does not decrease) the number of banks will

Figure 10 Total Banking Assets With and Without Off-Balance Sheet Adjustment



continue to decline. To illustrate how this drop could be affected by changes in entry rates, we ran two experiments similar to the counterfactual that we ran earlier. In both, we divided the banking industry into the same seven size categories we used earlier. Like in the earlier counterfactual experiment, we took the annual transition probabilities between size categories, exit rates from each category, and the entry rate for the 2008–2013 period. We then took the size distribution of banks in 2013 and calculated what the number of banks would be in 10 years if these transition rates did not change. We then took the same transition probabilities and only raised the entry rate to match the historical average of 1.5 percent and then calculated what the number of banks would be in 10 years under this more typical entry rate.[17]

Figure 9 shows the number of banks through 2013 and then the two different forecasts. While both forecasts predict a continued decline in the number of banks, there is a substantial quantitative difference

[17] There are obvious limitations to this exercise. In particular, entry and exit decisions are determined simultaneously in a market. Nevertheless, we think this simple exercise is useful because exit rates did not change that much from before the crisis to after it, so this assumption is plausible.

Table 8 Off-Balance Sheet Items and Credit Equivalents as of 2013

Item	Conversion Factor
Financial Standby Letters of Credit	1.00
Performance and Standby Letters of Credit	1.00
Commercial Standby Letters of Credit	0.20
Risk Participations in Bankers' Acceptances	1.00
Securities Lent	1.00
Retained Recourse on Small Business Obligations	1.00
Recourse and Direct Credit Substitutes	1.00
Other Financial Assets Sold with Recourse	1.00
Other Off-Balance Sheet Liabilities	1.00
Unused Loan Commitments (maturity >1 year)	0.50
Derivatives	—

Notes: Conversion factors used by regulators for determining credit equivalents of off-balance sheet items in 2013. The source is FFIEC 041 Schedule RC-R www.ffiec.gov/forms041.htm. Credit equivalents for derivatives do not have a direct conversion factor but instead are based on the current and future possible credit exposure.

between them. The number of banks under the existing trend drops another 1,000 banks over 10 years, while it only drops by about 500 banks under historic entry rates.

8. CONCLUSION

Since the financial crisis began, the biggest change to the size distribution of banks has been the decline in the number of small banks. We document that much of this decline is due to the lack of entry. We discussed several reasons for why there might be less entry, including macroeconomic conditions, regulatory costs, and regulatory barriers to entry. Regardless of the reasons for the decline, however, it is clear that to a large degree the number of banks as well as the size distribution of banks in the future will depend on whether entry recovers.

APPENDIX A: OFF-BALANCE SHEET ITEMS

Banks can make commitments that are not directly measured by a traditional balance sheet. For example, a loan commitment is a promise to make a loan under certain conditions. Traditionally, this kind of

promise was not measured as an asset on a balance sheet. As documented by Boyd and Gertler (1994), providing this and other off-balance sheet items has become an important service provided by banks, which means that traditional balance sheet numbers do not accurately report some of the implicit assets and liabilities of a bank.

We account for loan commitment and other off-balance sheet items like derivatives by converting them into *credit equivalents* and then adding them to on-balance sheet assets. We use the weights used by regulators to determine credit equivalents for capital purposes. Some of these adjustments are made starting in 1983, but many are added in 1990. The weights as of 2013 are reported in Table 8. Figure 10 demonstrates the importance of the adjustment starting in 1990 by plotting aggregate assets and loans with and without the adjustment.

APPENDIX B: TRANSITION MATRIX

One interesting thing that can be done with the transition matrix is to calculate the steady-state distribution of bank size. If the size distribution at time t is vector s_t and P is the transition matrix (also commonly called a Markov matrix), then the size distribution at $t+n$ is

$$s_{t+n} = P^n s_t.$$

If the transition matrix has the property that a bank starting in any category has a positive probability of moving to any other size category in a finite number of steps, then several theorems can be proven. In particular, there exists a unique stationary size distribution, that is, there exists s, such that $s = Ps$. Furthermore, regardless of the initial distribution, the size distribution will converge to this unique distribution.

For the transition matrix in Table 6, Table 9 shows the stationary distribution. There is a large fraction of banks in the over $50 billion size category. The reason for this concentration is that in the transition probabilities over the 2008–2013 period, 99 percent of banks in the largest size category stayed in it each year. Consequently, if a bank enters this category, it is very unlikely to leave, so banks accumulate there. In the recent period, this reflects the lack of merger activity among the largest banks and that the largest banks were prevented from failing by the federal government during the crisis. In past periods, transition probabilities for the largest size class were very different. For

Table 9 Stationary Distribution based on Transition Probabilities between Size Classes for 2008–2013

Size Class (millions)	Stationary Distribution	Distribution in 2013 Data
< 100	0.23	0.33
100–500	0.45	0.50
500–1,000	0.10	0.09
1,000–5,000	0.10	0.06
5,000–10,000	0.02	0.01
10,000–50,000	0.03	0.01
> 50,000	0.06	0.01

Notes: Columns do not add to 1 due to rounding.

example, over the 2000–2005 period, only 91 percent of banks in the largest size class stayed there.

The stationary distribution is useful for illustrating what direction the transition probabilities are taking the size distribution. As a long-term forecast, however, it is less valuable. It can take many iterations for a distribution to converge to its stationary distribution (over 200 in this case) and, as Janicki and Prescott (2006) show, properties of transition matrices for the banking industry have changed several times over the last 50 years.

REFERENCES

Adams, Robert M., and Jacob Gramlich. 2014. "Where Are All the New Banks? The Role of Regulatory Burden in New Charter Creation." FEDS Working Paper 2014-113 (December).

Adelman, Irma G. 1958. "A Stochastic Analysis of the Size Distribution of Firms." *Journal of the American Statistical Association* 53 (December): 893–904.

Axtell, Robert L. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science* 293 (September): 1,818–20.

Berger, Allen N., and Gregory F. Udell. 2002. "Small Business Credit Availability and Relationship Lending: The Importance of Bank Organisational Structure." *Economic Journal* 112 (February): F32–F53.

Berger, Allen N., Seth D. Bonime, Lawrence G. Goldberg, and Lawrence J. White. 2004. "The Dynamics of Market Entry: The Effects of Mergers and Acquisitions on Entry in the Banking Industry." *Journal of Business* 77 (October): 797–834.

Berger, Allen N., Anil K. Kashyap, and Joseph M. Scalise. 1995. "The Transformation of the U.S. Banking Industry: What a Long, Strange Trip It's Been." *Brookings Papers on Economic Activity* 2: 55–201.

Boyd, John H., and Mark Gertler. 1994. "Are Banks Dead? Or Are the Reports Greatly Exaggerated?" Federal Reserve Bank of Minneapolis *Quarterly Review* 18 (Summer): 2–23.

Dunham, Constance R. 1989. "New Banks in New England." Federal Reserve Bank of Boston *New England Economic Review* January: 30–41.

Ennis, Huberto M., and Peter S. Debbaut. 2014. "Large U.S. Bank Holding Companies During the 2007–2009 Financial Crisis: An Overview of the Data." Manuscript.

Federal Deposit Insurance Corporation. 2009. "Enhanced Supervisory Procedures for Newly Insured FDIC-Supervised Depository Institutions." FIL-50-2009, August 28.

Feldman, Ron J., Jason Schmidt, and Ken Heineche. 2013. "Quantifying the Costs of Additional Regulation on Community Banks." Federal Reserve Bank of Minneapolis Economic Policy Paper 13-3 (May).

Gabaix, Xavier. 2009. "Power Laws in Economics and Finance." *Annual Review of Economics* 1 (May): 255–94.

Goddard, John A., Hong Liu, Donal G. McKillop, and John O.S. Wilson. 2014. "The Size Distribution of US Banks and Credit Unions." *International Journal of the Economics of Business* 21 (February): 139–56.

Goddard, John, Hong Liu, and John O.S. Wilson. 2014. "Entry, Exit and Growth of US Commercial Banks." Available at http://ssrn.com/abstract=2398888 (February 20).

Jackson, Ben. 2005. "Urban Parent's Refocus Freeing a Farm Lender." *American Banker, 3* June, 4.

Janicki, Hubert P., and Edward S. Prescott. 2006. "Changes in the Size Distribution of U.S. Banks: 1960–2005." Federal Reserve Bank of Richmond *Economic Quarterly* 92 (Fall): 291–316.

Jayaratne, Jith, Philip E. Strahan. 1997. "The Benefits of Branching Deregulation." Federal Reserve Bank of New York *Economic Policy Review* 3 (December): 13–29.

Kane, Edward J. 1996. "De Jure Interstate Banking: Why Only Now?" *Journal of Money, Credit and Banking* 28 (May): 141–61.

Keeton, William R. 2000. "Are Mergers Responsible for the Surge in New Bank Charters." Federal Reserve Bank of Kansas City *Banking Studies* 85 (1): 21–41.

Ladenson, Mark L., and Kenneth J. Bombara. 1984. "Entry in Commercial Banking, 1962–78." *Journal of Money, Credit and Banking* 16 (May): 165–74.

Lindley, James T., James A. Verbrugge, James E. McNulty, and Benton E. Gup. 1992. "Investment Policy, Financing Policy, and Performance Characteristics of De Novo Savings and Loan Associations." *Journal of Banking and Finance* 16 (April): 313–30.

Luttmer, Erzo G. J. 2007. "Selection, Growth, and the Size Distribution of Firms." *Quarterly Journal of Economics* 122 (3): 1,103–44.

Mengle, David L. 1990. "The Case for Interstate Branch Banking." Federal Reserve Bank of Richmond *Economic Review* 76 (November/December): 3–17.

Moore, Robert R., and Edward C. Skelton. 1998. "New Banks: Why Enter When Others Exit?" Federal Reserve Bank of Dallas *Financial Industry Issues* Q1: 1–7.

Morris, Charles S., and Kristen Regehr. 2014. "What Explains Low Net Interest Income at Community Banks?" Federal Reserve Bank of Kansas City *Economic Review* Q2: 5–33.

Newman, M.E.J. 2005. "Power Laws, Pareto Distributions and Zipf's Law." *Contemporary Physics* 46 (5): 323–51.

Peirce, Hester, Ian Robinson, and Thomas Stratmann. 2014. "How Are Small Banks Faring under Dodd-Frank?" Mercatus Center Working Paper 14-05.

Peters, Andy. 2013. "Amish Bank Charter to Set Standard for Future Applications." *American Banker*, 4 December.

Simon, Herbert, A., and Charles P. Bonini. 1958. "The Size Distribution of Business Firms." *American Economic Review* 48 (September): 607–17.

Spong, Kenneth. 2000. *Banking Regulation*. Kansas City, Mo.: Federal Reserve Bank of Kansas City.

Stewart, Jackie. 2012. "North Carolina Investors Buy Bank from Capitol." *American Banker*, 12 August.

Economic Quarterly—Volume 100, Number 1—First Quarter 2014—Pages 51–85

A Business Cycle Analysis of Debt and Equity Financing

Marios Karabarbounis, Patrick Macnamara, and Roisin McCord

The recent turmoil in financial markets has highlighted the need to better understand the link between the real and the financial sectors. For example, a widespread view holds that real shocks can propagate themselves by adversely affecting credit markets (financial accelerator). An informative way to establish such linkages is to look at the co-movement between financial flows and macroeconomic conditions. The magnitude and direction of this relationship can guide our thinking regarding how strong these linkages are and the particular way in which they manifest themselves.

This article takes a modest step in this direction. In particular, we provide an introductory, yet comprehensive, business cycle analysis of firm financing. We first document empirically the cyclical properties of debt and equity issuance. We then build a simple two-period model to analyze the optimal capital structure as well as the response of firm financing to exogenous shocks such as a productivity shock. Finally, we examine how well a fully dynamic, reasonably calibrated, heterogeneous-firm model replicates the business cycle properties of debt and equity issuance.

We document empirical patterns of firm financing based on Compustat for the period 1980–2013. We find that firms issue more debt during expansions. In contrast, the cyclical properties of equity issuance depend on the exact definition of equity. If we define equity issuance using the sale of stock net of equity repurchases (following Jermann and Quadrini [2012]), we find a countercyclical equity issuance (or a

For very helpful suggestions we thank Kartik Athreya, Arantxa Jarque, Marisa Reed, and Alex Wolman. Any opinions expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. E-mail: Marios.Karabarbounis@rich.frb.org; Patrick.Macnamara@manchester.ac.uk.

procyclical equity payout). If we follow Covas and Den Haan (2011) and define equity issuance based on the change in the book value of equity, we find equity issuance to be weakly procyclical. Equity financing through mergers explains much of the discrepancy between the two measures. Stock compensation also explains the discrepancy but to a smaller degree. Moreover, regardless of the measure used, the countercyclicality of net equity issuance is driven by a strongly procyclical dividend payout and not countercyclical gross equity issuance. The data also reveal a substantial degree of heterogeneity in firms' financial decisions. Compared to large firms, the debt issuance of small firms tends to be less procyclical while equity issuance tends to be more procyclical.

To build intuition, we analyze the firm's optimal capital structure within a simple two-period model. Each period, firms receive an idiosyncratic productivity shock. The firm chooses how much to invest and how it will finance this decision. Financing can take the form of a one-period bond (debt) and external equity. The firm chooses debt issuance to balance the tax benefits of debt with the expected bankruptcy costs of default. External equity is also assumed to be costly. We show how the policy functions for investment, debt, and equity vary with internal equity, the costs of issuing equity, and idiosyncratic productivity.

Our fully dynamic model incorporates many of the elements outlined in the two-period model. Firms experience both aggregate and idiosyncratic productivity shocks. Nevertheless, we keep the analysis simple and assume a partial equilibrium framework. The model is calibrated to match several cross-sectional moments as calculated from Compustat. We then examine how well our model can explain the cyclical properties of debt and equity issuance. As in the data, firms issue more debt in response to a positive productivity shock. Higher productivity implies that firms desire to invest more, which makes default more costly and, hence, borrowing easier. Moreover, equity issuance is countercyclical. This is driven by large firms issuing more dividends during expansions. The model also captures the firm-size relationship in firm financing. Specifically, the model is able to match the empirical observation that net equity issuance of small firms is procyclical, while debt issuance is less procyclical than for larger firms.

This article contributes to the literature on firm financing in two ways. First, we highlight how equity financing through mergers and stock compensation can account for the different measures of net equity issuance used in the literature. In particular, we show that if one excludes mergers and stock compensation, the measures used by Covas and Den Haan (2011) and Jermann and Quadrini (2012) (change in book value of equity and net sale of stock, respectively) lead to the

same conclusion. Moreover, we show that a countercyclical net equity issuance in the data is driven by dividend payouts falling during recessions, not gross equity issuance increasing during recessions. Although such a distinction is crucial for understanding how firm financing varies over the cycle, it is not stressed in the literature. Second, we test these predictions within a quantitative model of firm financing with heterogeneous firms. Although this is certainly not the first quantitative article of firm financing, our article makes several novel contributions. For example, we build intuition regarding the determinants of firm financing using a simple two-period model. Moreover, using our heterogeneous-firm model we can test if the model captures the empirical firm-size relationship and especially the decomposition of equity financing into gross equity issuance and payout components.

1. RELATED LITERATURE

Our analysis borrows many elements from the work of Covas and Den Haan (2011), who look at disaggregated data from Compustat and document the cyclical properties of firm finance for different firm sizes. Their finding is that debt and (net) equity issuance is procyclical as long as the very large firms are excluded. Hence, Covas and Den Haan (2011) stress the importance of incorporating heterogeneity in quantitative models of firm financing.[1] Jermann and Quadrini (2012) document the cyclical properties of financial flows using aggregate data from the flow of funds accounts. The authors find a procyclical debt issuance but a countercyclical net equity issuance. Their article also examines the macroeconomic effects of financial shocks by constructing a shock series for the financial shock and then feeding the shock into a real business cycle model. Beganau and Salomao (2014) also document financial flows from Compustat. Following the equity definition of Jermann and Quadrini (2012), Beganau and Salomao (2014) also find net equity issuance is countercyclical.

Although the focus on the cyclicality of financial flows has been relatively new, there is ample work on the cross-sectional determinants of capital structure and firm dynamics. Rajan and Zingales (1995) investigate the relationship between leverage and firms' characteristics for a set of countries. They report that most of the empirical regularities found in the United States (such as the positive relationship between firm size and leverage) are also true for other countries. Cooley and

[1] In a related article, Covas and Den Haan (2012) build a quantitative model of debt and equity finance. Our model in Section 4 uses many of their modeling assumptions.

Quadrini (2001) introduce financial frictions in a model of industry dynamics and study the relationship between firm financing and firm size and age. Hennessey and Whited (2007) build a structural model of firm financing and estimate the magnitude of external financing costs. Recently, Katagiri (2014) builds a general equilibrium model of firm financing to study the distribution of leverage.

2. EMPIRICAL ANALYSIS

In this section, we describe several empirical patterns regarding firm financing. We first explain how we construct the variables used in the analysis. We next present aggregate statistics both in the cross-section of firms and along the business cycle. The main findings emerging from the analysis are the following. First, debt issuance is strongly procyclical. Second, the cyclicality of equity issuance depends on the specific measure used. However, smaller firms seem to issue more equity in expansions relative to larger firms, independent of the measure. Third, there is widespread heterogeneity in firm financing decisions.

Data Construction

To construct our variables we use annual data from Compustat. Compustat contains financial information on publicly held companies. Following the literature on firm financing, we focus on the period between 1980 and 2013. Jermann and Quadrini (2012) document that during this period there was a break in macroeconomic volatility as well as significant changes in U.S. financial markets. We exclude financial firms and utilities as these industries are more heavily regulated.[2] One important concern is whether we include firms affected by a merger or an acquisition. For this purpose, we separately report results for two cases. In the first case, we follow Covas and Den Haan (2011) and drop all firm-year observations that are affected by a "major" merger or acquisition. By "major" we mean that the merger or acquisition causes the firm's sales to increase by more than 50 percent. In the second case, we drop all observations affected by any kind of merger. After imposing these restrictions and dropping all observations affected by a major merger, we are left with an unbalanced panel of 19,101 firms and a total of 168,295 firm-year observations. When we also drop observations affected by any merger, we are left with 18,486 firms and 141,379 observations.

[2] For more details on the construction of our data, see Appendix A.

Variable Definitions

The literature uses two different methods to measure equity issuance. Fama and French (2005) and Covas and Den Haan (2011) use changes in the book value of equity (reported on the firm's balance sheet) to measure equity issuance. Jermann and Quadrini (2012) use the "net sale of stock" (from the statement of cash flows) in the construction of equity issuance. To clarify the difference between these two measures, it is useful to define a company's accounting identity:

$$A_{i,t} = SE_{i,t} + RE_{i,t} + L_{i,t}.$$

For firm i at date t, assets $A_{i,t}$ must equal equity plus liabilities $L_{i,t}$ (all variables are book values). Total equity includes retained earnings $RE_{i,t}$, which is the portion of the company's net income it has retained rather than distributed to shareholders as dividends. Therefore, $SE_{i,t}$ is the company's total equity net of retained earnings. This part of the firm's balance sheet reflects equity that the firm has obtained from "external" sources such as sale of common stock.

Under the first definition, equity issuance is the annual change in $SE_{i,t}$ minus cash dividends distributed to shareholders $d_{i,t}$. We subtract cash dividends from our definition because, effectively, they represent one of two ways firms can distribute funds to shareholders: They can buy back stock, which would decrease $SE_{i,t}$, or they can issue dividends, which would decrease $RE_{i,t}$ instead. Therefore, in our first definition, the equity issuance of firm i at date t is

$$\Delta E_{i,t}(1) \equiv \Delta SE_{i,t} - d_{i,t}, \tag{1}$$

where $\Delta SE_{i,t} \equiv SE_{i,t} - SE_{i,t-1}$ is the annual change in $SE_{i,t}$. This corresponds to one of the primary definitions of equity issuance in Covas and Den Haan (2011). Our second definition of equity issuance is defined as follows:

$$\Delta E_{i,t}(2) \equiv \Delta SS_{i,t} - d_{i,t}; \tag{2}$$

$\Delta SS_{i,t}$ is the net sale of stock, which is defined as the gross revenue from the sale of stocks minus stock repurchases. This corresponds to the definition of equity issuance utilized by Jermann and Quadrini (2012).

Ideally these two measures would be equivalent, as the net sale of stock $\Delta SS_{i,t}$ affects $SE_{i,t}$. Nevertheless, the two definitions lead to different conclusions about the cyclicality of equity issuance. This discrepancy has to do with the way firms choose to issue equity. Apart from equity offerings to the public, equity issuance can take place through mergers, warrants, employee options, grants, and benefit plans among others. Hence, as Fama and French (2005) note, the net sale of stock measure captures only a few of the ways in which firms can raise outside equity. Take, for example, a merger or an acquisition. Suppose

a firm acquires another firm by issuing equity to the shareholders of the target firm. This transaction will change the book value of equity. However, it will not alter the sale of stock measure because no actual revenue is raised by the transaction. Moreover, suppose a firm were to compensate its employees with a stock. Again, if equity is measured using the book value of equity, equity issuance will increase. This is because employee compensation will decrease retained earnings and thus increase $SE_{i,t}$, the company's equity net of retained earnings. Meanwhile, as before, the sale of stock measure will not record the equity issuance because no actual revenue is raised.

In the data, a situation in which no firms issue equity (on net) will look the same as a situation in which some firms issue equity while others reduce equity. To uncover such heterogeneity, we break up our first definition of equity issuance into a "gross equity issuance" and "gross equity payouts" component.[3] In particular, we define gross equity issuance $E^I_{i,t}$ to be

$$E^I_{i,t} \equiv \begin{cases} \Delta SE_{i,t} & \text{if } \Delta SE_{i,t} > 0 \\ 0 & \text{if } \Delta SE_{i,t} \leq 0 \end{cases} . \tag{3}$$

Similarly, we define gross equity payouts $E^P_{i,t}$ to be

$$E^P_{i,t} \equiv \begin{cases} d_{i,t} & \text{if } \Delta SE_{i,t} > 0 \\ -\Delta SE_{i,t} + d_{i,t} & \text{if } \Delta SE_{i,t} \leq 0 \end{cases} . \tag{4}$$

Note that $\Delta E_{i,t}(1) = E^I_{i,t} - E^P_{i,t}$ by construction. By looking at gross flows, we can separately identify firms that raise equity and firms that reduce equity.

Moreover, we also consider several other variables of interest. In particular, w_S will denote employee stock compensation; $\Delta RE_{i,t} \equiv RE_{i,t} - RE_{i,t-1}$ is the change in retained earnings. A firm's net debt issuance $\Delta D_{i,t} \equiv D_{i,t} - D_{i,t-1}$ is defined to be the change in the firm's book value of debt between period $t-1$ and t. A firm's net change in sales $\Delta S_{i,t} \equiv S_{i,t} - S_{i,t-1}$ is defined to be the change in the firm's nominal sales between t and $t-1$. Finally, $I_{i,t}$ is the firm's investment while $K_{i,t}$ is the firm's capital stock.

Construction of Group Aggregates

To uncover any underlying heterogeneity in the financing decisions of firms, we sort firms by size. At each date t, we sort firms into four possible groups based on their size (more on the construction of these

[3] We can similarly break up the second definition of equity. However, as discussed earlier, the second definition of equity tends to understate equity issuance.

groups later). Then, for every date t, we aggregate each firm-level variable across all the firms in each bin. To be precise, let $X_{i,t}$ be a variable of interest for firm i at date t. For example, this might be $\Delta D_{i,t}$, the net debt issuance of a particular firm. Let $G_{j,t}$ denote the set of firms in group j at date t. Then, we can construct the group aggregate $X_{j,t}$ as follows:

$$X_{j,t} = \frac{\sum_{i \in G_{j,t}} X_{i,t}}{\sum_{i \in G_{j,t}} K_{i,t}}. \tag{5}$$

The numerator is the sum of $X_{i,t}$ across all firms in group j at date t. Therefore, if $X_{i,t}$ is $\Delta D_{i,t}$, then the numerator of (5) is the net amount of debt issued by all firms in group j at date t. Meanwhile, the denominator of (5) is the total amount of capital in group j at date t. The denominator is used to normalize the resulting aggregate variable and capital is chosen because it is acyclical. Following this procedure, we obtain a time series for the aggregate variable X for each group. Note, however, that the composition of firms in each group varies over time. Not only may a firm transition between groups over time, but the groups may include newly listed firms.

To construct the firm groups, we sort firms based on the previous period's book value of their assets. At each date, we sort firms into four groups. The first group consists of firms with assets below the median ([0, 50]). The second group consists of firms between the 50th and 75th percentile ([50, 75]), and the third group consists of firms between the 75th and 99th percentile ([75, 99]). And finally, the last group consists of firms in the top 1 percent ([99, 100]). As the book value of assets tends to grow over time, we have to be careful in how we determine the asset boundaries for these size groups. Define $A_{50,t}$, $A_{75,t}$, and $A_{99,t}$ to be the asset boundaries between the four size bins. In other words, a firm with assets $A_{i,t} < A_{50,t}$ will be in the [0, 50] group at date $t+1$. Following Covas and Den Haan (2011), we construct $A_{50,t}$, $A_{75,t}$, and $A_{99,t}$ by fitting a (log) linear trend through the asset values that correspond to the 50th, 75th, and 99th percentiles at each time t.

Cross-Sectional Analysis

We begin our analysis by looking at group aggregates for the whole period between 1980 and 2013 (Table 1). Each variable is expressed as a percentage of the group capital stock. In the top panel we exclude major mergers from the sample while in the lower panel we exclude all mergers. Looking at the top panel, we see that relative to their size, small firms tend to issue more debt and equity than large firms. Debt issuance decreases monotonically from 14.1 percent of the group's

Table 1 Summary Statistics

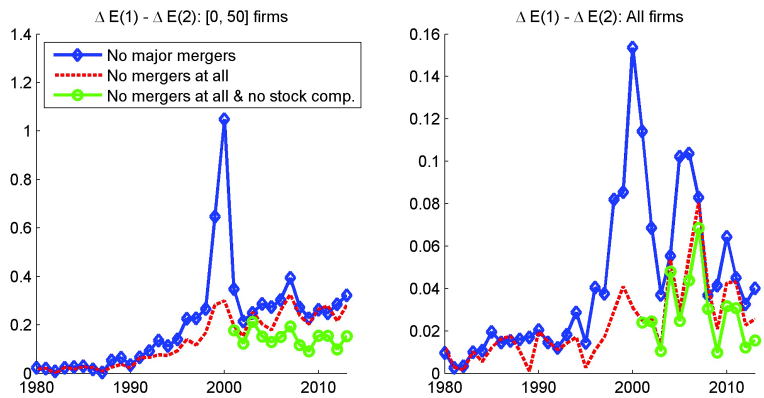
	Size Class (Percent)				
No Major Mergers	**[0, 50]**	**[50, 75]**	**[75, 99]**	**[99, 100]**	**[0, 100]**
ΔD	10.1	8.6	5.4	2.7	4.7
$\Delta E(1)$	69.0	8.6	−3.0	−5.3	−3.0
$\Delta E(2)$	40.6	−1.5	−9.9	−10.0	−9.4
$\Delta E(1) - \Delta E(2)$	28.4	10.0	6.9	4.7	6.3
w_S	7.9	2.6	0.9	0.3	0.8
E^I	79.8	18.0	6.9	4.7	6.9
E^P	10.8	9.4	10.0	10.0	10.0
ΔRE	−35.7	3.6	4.5	3.4	4.0
I	4.6	1.9	0.7	1.1	0.9
ΔS	42.1	28.3	14.5	14.3	15.0
	Size Class (Percent)				
No Mergers At All	**[0, 50]**	**[50, 75]**	**[75, 99]**	**[99, 100]**	**[0, 100]**
ΔD	4.0	0.7	1.7	0.9	1.5
$\Delta E(1)$	76.8	3.6	−5.8	−7.4	−5.6
$\Delta E(2)$	53.3	−1.7	−9.5	−10.7	−9.3
$\Delta E(1) - \Delta E(2)$	23.4	5.3	3.7	3.3	3.7
w_S	8.7	2.6	0.8	0.4	0.8
E^I	87.8	14.1	3.6	2.4	4.0
E^P	11.0	10.5	9.4	9.9	9.6
ΔRE	−49.9	4.1	4.5	6.1	4.8
I	4.8	1.5	0.3	0.5	0.5
ΔS	30.7	15.2	9.4	10.6	9.9

Notes: This table reports the average of various group aggregates between 2001 and 2013. Each variable is expressed as a percentage of the group capital stock. ΔD is net debt issuance. $\Delta E(1)$ is the first measure of net equity issuance and is defined in (1). Similarly, $\Delta E(2)$ is the second measure of equity issuance and is defined in (2). w_S is stock compensation. E^I is gross equity issuance and is defined in (3). E^P is gross equity payouts and is defined in (4). ΔRE is the net change in retained earnings. I is investment. ΔS is the net change in sales. Note that we only have data on stock compensation between 2001 and 2013.

capital stock in the [0, 50] bin to 2.9 percent for firms in the top 1 percent. Equity issuance $\Delta E(1)$ decreases from 64.9 percent of capital for firms in the [0, 50] bin to −3.9 percent for firms in the top 1 percent. For our second measure, $\Delta E(2)$, these numbers are 44.4 percent and −6.4 percent, respectively. Nevertheless, the two measures of equity do differ in a significant way. $\Delta E(2)$, which is based on the net sale of stock, underestimates the amount of equity that firms raise. While the measures differ across all size groups, they are significantly different for smaller firms.

As noted earlier, mergers financed through the issuance of stock may also explain part of the difference between the two equity

Figure 1 Effect of Mergers and Stock Compensation on Equity Issuance



Notes: The graph plots the difference between our two measures of equity issuance $\Delta E(1)$ and $\Delta E(2)$ for the period 1980–2013. We plot the difference if (i) no major mergers are included in the sample, (ii) no mergers at all are included in the sample, and (iii) no mergers at all are included in the sample and stock compensation is subtracted from the difference. The left panel shows the differences for firms in the [0, 50] size class. The right panel shows the differences for all firms. In each case, the differences are plotted as a percentage of the group capital stock. Information on stock compensation is available only after 2003.

measures. To investigate how much mergers and acquisitions explain the difference, we repeat our earlier analysis, but we exclude all mergers from the sample. The bottom panel of Table 1 reports the results when all mergers are excluded. While the same results hold as before, firms on average issue less debt than before (1.8 percent versus 5.0 percent). Firms also issue less equity under both definitions (−3.9 percent and −6.2 percent versus −1.7 percent and −6.0 percent, respectively). Overall, the difference between the two equity measures falls by almost half. Moreover, stock compensation (which is not reflected in $\Delta E(2)$) does explain some of the remaining discrepancy between the two measures.[4] In fact, for small firms it is a major explanation for the discrepancy between the two measures. Still, after accounting for mergers and stock compensation, significant differences remain.

[4] However, note that our data for stock compensation only begins in 2001.

Figure 1 shows how our equity measures, $\Delta E(1)$ and $\Delta E(2)$, differ between 1980 and 2013. Similar to Table 1, we plot the difference $\Delta E(1) - \Delta E(2)$ for three different cases: (i) if no major mergers are included, (ii) if no mergers at all are included, and (iii) if no mergers at all are included and we subtract from the difference equity issuance related to stock compensation. The left panel of Figure 1 shows the differences for firms in the [0, 50] size group, while the right panel shows the differences for all firms. This figure highlights how the differences between these two measures have grown since the late 1990s. Moreover, it also demonstrates the importance that mergers and acquisitions have had on equity financing, especially in the late 1990s. $\Delta E(1)$ can capture these effects while $\Delta E(2)$ cannot. However, in the period after 2007, mergers seem to account for only a small part of the discrepancy. Nevertheless, during that period, stock compensation seems to account for a larger fraction of the difference. As seen in Figure 1, this is especially true for firms in the [0, 50] size group.

Finally, from Table 1 (both top and bottom panels) it is readily apparent that small firms grow faster (in terms of sales growth) and invest at a higher rate. Moreover, excluding the top 1 percent, smaller firms have lower growth in retained earnings and ΔRE is even negative for firms in the [0, 50] size group. These results are consistent with the findings of Covas and Den Haan (2011).

Business Cycle Analysis

We next turn to the business cycle analysis of debt and equity issuance. In Table 2, we report the correlation of various group aggregates with real corporate gross domestic product (GDP). To compute these correlations, both GDP and the group aggregates are de-trended with an H-P filter.[5] First consider the top panel of Table 2, which includes results for the case when only major mergers are excluded from the sample. Consistent with Covas and Den Haan (2011) and Jermann and Quadrini (2012), debt issuance is strongly procyclical. The cyclicality is stronger for larger firms. The correlation between debt issuance and corporate GDP increases from 0.536 for the [0, 50] size group to 0.755 for the [75, 99] size group. The correlation falls to 0.547 for firms in the top 1 percent. However, note that there is a relatively small number of firms in this group.[6]

[5] Throughout this article, we use a smoothing parameter of 100 to de-trend annual data.

[6] There are, on average, 31 firms in the top 1 percent every year.

Table 2 Business Cycle Correlations of Debt and Equity Issuance

	Size Class (Percent)				
No Major Mergers	**[0, 50]**	**[50, 75]**	**[75, 99]**	**[99, 100]**	**[0, 100]**
ΔD	**0.536**	**0.611**	**0.755**	**0.547**	**0.785**
	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
$\Delta E(1)$	**0.345**	0.191	0.016	0.044	0.096
	(0.046)	(0.280)	(0.927)	(0.804)	(0.589)
$\Delta E(2)$	0.243	−0.250	**−0.617**	−0.312	**−0.509**
	(0.166)	(0.155)	(0.000)	(0.072)	(0.002)
w_S	0.010	−0.050	0.066	0.022	0.341
	(0.973)	(0.859)	(0.816)	(0.937)	(0.214)
E^I	**0.353**	0.268	0.306	0.250	**0.363**
	(0.041)	(0.125)	(0.079)	(0.153)	(0.035)
E^P	0.069	0.279	**0.654**	0.314	**0.588**
	(0.697)	(0.110)	(0.000)	(0.071)	(0.000)
	Size Class (Percent)				
No Mergers At All	**[0, 50]**	**[50, 75]**	**[75, 99]**	**[99, 100]**	**[0, 100]**
ΔD	**0.646**	**0.590**	**0.661**	**0.418**	**0.661**
	(0.000)	(0.000)	(0.000)	(0.014)	(0.000)
$\Delta E(1)$	0.264	−0.168	**−0.419**	−0.303	−0.322
	(0.132)	(0.343)	(0.014)	(0.082)	(0.064)
$\Delta E(2)$	0.230	−0.312	**−0.687**	−0.193	**−0.506**
	(0.191)	(0.072)	(0.000)	(0.274)	(0.002)
w_S	0.035	0.027	0.009	−0.005	0.225
	(0.901)	(0.924)	(0.975)	(0.985)	(0.421)
E^I	0.283	−0.040	0.033	−0.113	0.100
	(0.105)	(0.824)	(0.852)	(0.524)	(0.573)
E^P	0.240	0.273	**0.627**	0.314	**0.586**
	(0.172)	(0.119)	(0.000)	(0.071)	(0.000)

Notes: This table reports correlations of group aggregates with real corporate GDP. All variables are de-trended with an H-P filter. The p-values for each correlation are shown in parentheses. Coefficients that are significant at the 5 percent level are shown in bold. ΔD is net debt issuance. $\Delta E(1)$ is the first measure of net equity issuance and is defined in (1). Similarly, $\Delta E(2)$ is the second measure of equity issuance and is defined in (2). w_S is stock compensation. E^I is gross equity issuance and is defined in (3). E^P is gross equity payouts and is defined in (4). Note that we only have data on stock compensation between 2001 and 2013.

Overall, equity issuance, as measured by $\Delta E(1)$, is acyclical. However, according to this measure, equity issuance tends to be procyclical and statistically significant for firms in the [0, 50] size group. The cyclicality of equity issuance monotonically decreases across size groups and becomes essentially uncorrelated with output for the top 1 percent. In particular, the correlation decreases from 0.345 for firms in the [0, 50] size group to 0.044 for firms in the top 1 percent. These results are

consistent with Covas and Den Haan (2011). In contrast, if we measure equity issuance using the net sale of stock ($\Delta E(2)$), then equity issuance becomes strongly countercyclical. This is consistent with Jermann and Quadrini (2012). However, even according to this measure, equity issuance of the smallest firms tends to be procyclical (although statistically insignificant) with a correlation of 0.243. Meanwhile, $\Delta E(2)$ is significantly countercyclical for firms in the [75, 99] size group with correlation equal to -0.617. This pattern of financing across firm size is consistent with the net sale of stock measure reported in Covas and Den Haan (2011).

In Table 2, we also report how the cyclicality of $\Delta E(1)$ breaks into a gross equity issuance and gross equity payouts component, both defined in (3) and (4). For smaller firms, gross equity issuance is driving the (pro)cyclicality of net equity issuance. But for all other firm sizes, procyclical gross equity issuance is associated with a more procyclical gross equity payout. Both statistics may explain the weak cyclicality of net equity issuance $\Delta E(1)$. This decomposition can also shed some light on the discrepancy between our net equity measures $\Delta E(1)$ and $\Delta E(2)$. Since $\Delta E(2)$ underestimates gross equity issuance, it is mostly affected by a countercyclical gross equity payout.

Similar to our cross-sectional analysis, we trace the discrepancy in the cyclical behavior of $\Delta E(1)$ and $\Delta E(2)$ to mergers and stock compensation. In the bottom panel of Table 2, we report the business cycle correlations when we exclude all mergers from the sample. In this case, debt issuance is slightly less correlated with GDP but is still strongly procyclical (0.661 versus 0.785). However, according to $\Delta E(1)$, equity issuance for all firms now becomes significantly countercyclical (at the 10 percent level) and significantly countercyclical for the top 25 percent. For example, the correlation for firms in the [75, 99] size group is -0.419 when we exclude all mergers versus 0.016 when we do not. Nevertheless, for the smallest firms, equity issuance according to $\Delta E(1)$ is still procyclical, but it is not significant. Moreover, gross equity issuance is now statistically insignificant for all size groups. For all firms, gross equity payouts is still significantly procyclical. Therefore, the procyclical nature of merger activity[7] appears to play an important role in explaining the differences in the cyclicality between $\Delta E(1)$ and $\Delta E(2)$.

Another candidate to explain the discrepancy in the cyclicality of the two measures is stock compensation. Table 2 includes information on the cyclicality of stock compensation by firms. As mentioned

[7] Eisfeldt and Rampini (2006) document that capital reallocation due to acquisitions is procyclical.

earlier, this type of equity issuance is captured by $\Delta E(1)$ but not by $\Delta E(2)$. Therefore, it could help explain the discrepancy between the two measures. However, as we see from Table 2, stock compensation is itself acyclical. Therefore, while it does explain some of the difference in levels between the two measures (especially for small firms), it does not help explain the different cyclicalities.

3. OPTIMAL CAPITAL STRUCTURE: A TWO-PERIOD MODEL

In this section we outline a simple two-period model to explain how the firm chooses its capital structure. Firms are perfectly competitive and produce a single homogeneous good. Capital is the only input in the firm's production function, $zf(k)$, and z is the firm's productivity. Productivity follows an AR(1) process. We denote by $F(z'|z)$ and $f(z'|z)$ the cumulative distribution and probability density functions for next period's productivity z', conditional on the current-period productivity z.

Budget Constraint

The firm enters the first period with an initial level of capital, k, and a required debt payment, b. Given k and b in the first period, the firm (i) produces $zf(k)$, (ii) chooses investment $i = k' - (1-\delta)k$, (iii) issues dividends d (or raises external equity if $d < 0$), and (iv) issues new debt, $q(z, k', b')b'$. The firm borrows using a defaultable one-period noncontingent bond. It promises to pay b' tomorrow and in return the firm receives $q(z, k', b')b'$ today, where q is the price of the bond. Later in this section we discuss how this price is determined. To facilitate the analysis, we follow Gourio (2013) by assuming that the firm receives a tax subsidy from the government proportional to the amount borrowed. In other words, for every dollar the firm raises in the bond market, the government gives the firm a subsidy of τ.[8]

The firm chooses dividends d, tomorrow's capital k', and debt b' subject to the following budget constraint:

$$d + k' = e(z, k, b) + (1+\tau)q(z, k', b')b', \tag{6}$$

where $e(z, k, b) \equiv zf(k) + (1-\delta)k - b$ is defined to be internal equity. Therefore, when choosing tomorrow's capital stock, the firm has access

[8] We are assuming that the tax subsidy takes place at issuance. However, in reality, the implicit tax subsidy takes place when the firm's earnings are taxed, as interest payments can be deducted from corporate taxable income.

to three sources of funding: (i) internal equity e, (ii) debt qb', which is supplemented by the tax subsidy, and (iii) external equity (when $d < 0$).

As discussed in Fazzari, Hubbard, and Petersen (1988), there are many reasons why external equity is costly, including taxes and flotation costs. Thus, we assume that issuing equity is costly and specify the cost $\Lambda(d)$ as follows:

$$\Lambda(d) = \begin{cases} -\lambda_0 d & \text{if } d < 0 \\ 0 & \text{if } d \geq 0 \end{cases} . \tag{7}$$

When $d < 0$, the firm is issuing external equity and the cost is assumed to be proportional to the amount of funds raised. Moreover, note that $\Lambda(d)$ does not appear in (6). Therefore, when $d < 0$, $-d$ is the amount of funds actually received by the firm. However, shareholders actually pay $-d + \Lambda(d) = -(1 + \lambda_0)d$, of which only $-d$ goes to the firm.

Default Decision

We also allow firms to default on their debt obligations. In particular, in period 2, the firm chooses whether it will pay b' or declare bankruptcy. If the firm does not default, it receives

$$V^{ND}(z', k', b') = z' f(k') + (1 - \delta)k' - b'. \tag{8}$$

In this case, the firm's shareholders receive output and the undepreciated capital minus the debt payment. However, if the firm defaults, we assume that the firm can hide and keep a fraction θ of its assets. Therefore, in this case, the firm receives

$$V^{D}(z', k') = \theta \left[z' f(k') + (1 - \delta)k' \right]. \tag{9}$$

Due to bankruptcy costs, lenders will only recover a fraction $1-\psi$ of the total remaining assets in the case of default. In other words, the lender recovers $(1 - \psi)(1 - \theta) \left[z' f(k') + (1 - \delta)k' \right]$ when the firm defaults.

Given (8) and (9), the firm will default tomorrow when $V^{ND}(z', k', b'^{D}(z', k')$. This implicitly defines a productivity threshold $z^*(k', b')$ such that the firm will default if and only if $z'^*(k', b')$. This threshold is defined to be the value of productivity, z^*, such that the firm is indifferent between defaulting and not defaulting: $V^{ND}(z^*, k', b'^{D}(z^*, k')$. Using (8) and (9), we can then obtain the following functional form for $z^*(k', b')$:

$$z^*(k', b') = \begin{cases} \frac{b'/(1-\theta)-(1-\delta)k'}{f(k')} & \text{if } b' \geq (1 - \theta)(1 - \delta)k' \\ 0 & \text{if } b' < (1 - \theta)(1 - \delta)k' \end{cases} . \tag{10}$$

Consequently, default is only possible when $b' > (1-\theta)(1-\delta)k'$. Moreover, when b' is above this threshold, z^* depends negatively on k' and

positively on b'. The more firms invest, the more output and capital the firm will have next period. This will make default more costly. Consequently, the default threshold decreases (i.e., $\partial z^*/\partial k' < 0$). In contrast, the more debt the firm issues, the more attractive default will be next period. In this case, the default threshold will increase (i.e., $\partial z^*/\partial b' > 0$).

Bond Price

We assume there are households willing to lend their savings to firms. The price that lenders charge, $q(z, k', b')$, takes into account the probability that a firm will default, which depends on the firm's choices for k' and b'. Specifically, it is assumed that q is set to guarantee the lender an expected return equal to the risk-free rate r. Hence, q will be given by

$$q(z, k', b') = \frac{1}{1+r}\left[1 - F(z^*(k', b')|z) + \frac{R(z, k', b')}{b'}\right], \tag{11}$$

where

$$R(z, k', b') \equiv (1-\psi)(1-\theta)\int_0^{z^*(k',b')} \left[z' f(k') + (1-\delta)k'\right] f(z'|z)dz'$$

is the unconditional expected recovery value of the bond in the case of default. Therefore, the price of debt is composed of two terms. With probability $1 - F(z^*|z)$, the firm will not default and the lender receives b'. However, when the firm does default, the lender receives a fraction $(1-\psi)(1-\theta)$ of total assets.

Firm's Problem

We can now write the firm's problem as a dynamic programming problem. Define $V(z, k, b)$ as the value of a firm with productivity z, capital k, and debt b. This value function is given by

$$V(z, k, b) = \max_{d, k', b'} \left\{ \begin{array}{c} d - \Lambda(d) + \\ \frac{1}{1+r}\int_0^\infty \max\left\{V^{ND}(z', k', b'), V^D(z', k')\right\} f(z'|z)dz' \end{array} \right\} \tag{12}$$

subject to the budget constraint in (6), which is repeated here:

$$d + k' = e(z, k, b) + (1+\tau)q(z, k', b')b'.$$

The firm's objective is to choose next period's capital stock k', debt b', and dividends d in order to maximize its lifetime valuation.

Characterizing the Solution

In this subsection we explain what determines the firm's optimal capital structure. To do so, it is useful to first re-write the firm's value function defined in (12). Specifically, using the bond price function defined in (11), the firm's value function can be re-written as[9]

$$V(z,k,b) = \max_{d,k',b'} \Big\{ e(z,k,b) - k' - \Lambda(d) + \tau q(z,k',b')b' - B(z,k',b') + \frac{1}{1+r}\left[E[z'f(k')|z] + (1-\delta)k'\right] \Big\}, \tag{13}$$

subject to the budget constraint in (6). Recall that $e(z,k,b) \equiv zf(k) + (1-\delta)k - b$ is defined to be internal equity. Let $T(z,k',b') = \tau q(z,k',b')b'$ denote the total value of the tax subsidy. This term reflects the tax benefit of debt issuance. Similarly, $B(z,k',b')$ is defined to be the expected bankruptcy costs and is given by

$$B(z,k',b') = \frac{\psi(1-\theta)}{1+r} \int_0^{z^*(k',b')} \left[z'f(k') + (1-\delta)k'\right] f(z'|z)dz'.$$

As before, firms will choose k', b', and d to maximize the firm's lifetime valuation. As is clear from (13), the effect of marginal changes in k' and b' on $T(z,k',b')$ and $B(z,k',b')$ will play a key role in determining the firm's optimal capital structure. To ease the exposition of the firm's problem, we will first consider the case where issuing equity is costless (i.e., $\lambda_0 = 0$) and describe how the optimal policies for k', b', and d are determined. We then allow for costly equity ($\lambda_0 > 0$) and analyze how the firm's optimal choices change.

Costless Equity Issuance

We first assume that $\lambda_0 = 0$, which implies that $\Lambda(d) = 0$ for all d. In this case, the first order conditions for k' and b' become

$$\tau \frac{\partial q}{\partial k'} - \frac{\partial B}{\partial k'} + \frac{E[z'f'(k')|z] + 1 - \delta}{1+r} = 1 \tag{14}$$

$$\tau \left[q + \frac{\partial q}{\partial b'} b' \right] = \frac{\partial B}{\partial b'}. \tag{15}$$

When $\tau = \psi = 0$, the first-order condition for k' in (14) reduces to the familiar expression that the expected marginal product of capital equals interest plus depreciation (i.e., $E[z'f'(k')|z] = r + \delta$). Therefore, the firm invests the first-best amount of k'. Moreover, when $\tau = \psi = 0$,

[9] The readers can find the exact derivation of this expression in Appendix B.

both sides of (15) are always zero. Therefore, the Modigliani-Miller theorem[10] applies and the optimal capital structure is indeterminate. In this case, there is no benefit or cost from issuing debt.

However, when $\tau > 0$ and $\psi > 0$, the Modigliani-Miller theorem no longer applies. As seen in (14), the tax subsidy and bankruptcy costs now affect the firm's investment decision. By affecting the net tax benefit, $\tau q b' - B$, a marginal change in k' now has an additional benefit or cost. Consequently, whether the optimal k' is above or below the first-best level of k' depends on how a marginal change in k' affects the net tax benefit. Under our benchmark parameterization, $\tau \frac{\partial(qb')}{\partial k'} > \frac{\partial B}{\partial k'}$, we imply that k' can be higher than the first-best level of k'. Moreover, when $\tau > 0$ and $\psi > 0$, debt is beneficial to the firm because it increases the tax subsidy it receives. At the same time, more debt makes default more likely and increases the expected costs of bankruptcy. Consequently, as seen in (15), firms choose b' to equate the marginal tax benefits of debt with marginal bankruptcy costs.

The left panel of Figure 2 provides a visual characterization of the optimal capital structure. Since external equity is costless, internal and external equity are perfect substitutes. Hence, internal equity does not have any effect on the optimal value for k' and b', which are both horizontal lines. In what follows, we denote by k^* and b^* the firm's optimal choice of k' and b' when $\lambda_0 = 0$. Given that $k' = k^*$ and $b' = b^*$ for any value of e, it follows from the firm's budget constraint in (6) that the optimal dividend policy is then just a straight line (with a slope of 1). Firms with low (or even negative) internal equity are able to choose $k' = k^*$ because they can issue equity costlessly. Firms with large amounts of internal equity choose $k' = k^*$ and also choose to issue a positive dividend.

Costly Equity Issuance

Now we assume that external equity is costly (i.e., $\lambda_0 > 0$). In this case, the first-order conditions for b' become

$$(\tau + I_{d<0}\lambda_0(1+\tau))\left[q + \frac{\partial q}{\partial b'}b'\right] = \frac{\partial B}{\partial b'}. \tag{16}$$

This condition will only hold when $d \neq 0$. In the case of costly external equity, the marginal cost of an additional unit of debt is the same. Nevertheless, there is potentially an additional benefit of debt. In particular, an additional unit of debt allows the firm to substitute away from costly external equity. As seen in (16), a marginal increase in b'

[10] See Modigliani and Miller (1958).

means that the firm is able to raise $(1+\tau)\left[q+\frac{\partial q}{\partial b'}b'\right]$ in extra funds through the debt market (and through an additional tax subsidy). For each unit of extra funds raised, the firm is able to save on the external equity cost λ_0.[11]

Similarly, the first-order condition for k' is now given by

$$\tau\frac{\partial q}{\partial k'}-\frac{\partial B}{\partial k'}+\frac{E[z'f'(k')|z]+1-\delta}{1+r}=1+I_{d<0}\lambda_0\left[1-(1+\tau)\frac{\partial q}{\partial k'}b'\right]. \tag{17}$$

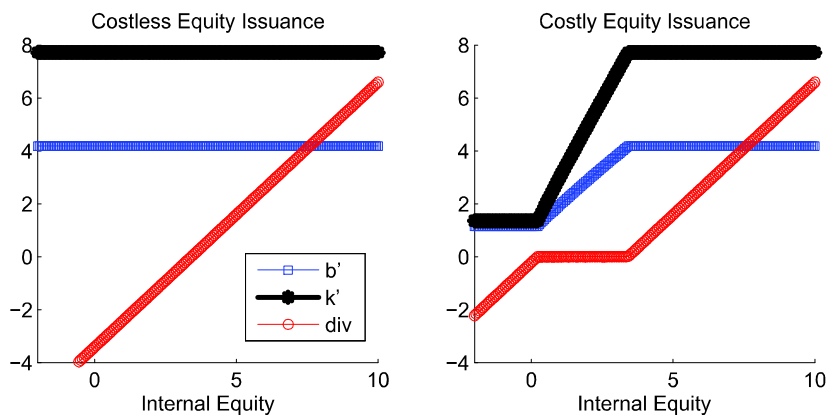
This condition only holds with equality when $d \neq 0$. In the case of costly external equity, the marginal benefit of additional investment is the same. However, there is now potentially an additional cost associated with increasing k'. When the firm is already relying on external equity ($d < 0$), the additional unit of k' must be financed with expensive external equity. Since a higher k' tends to lower the price on existing debt, the firm only needs to raise $1-(1+\tau)\frac{\partial q}{\partial k'}b'$ of external equity. For every unit of additional external equity the firm raises, it must pay the cost λ_0.

The right panel of Figure 2 plots the policy functions for k', b', and d as a function of internal equity when external equity is costly. Examination of Figure 2 reveals that firms now behave differently depending on how much internal equity they have (their initial size). There are three regions of interest: (1) firms with low levels of internal equity, (2) firms with medium levels of internal equity, and (3) firms with high levels of internal equity.

First consider firms with low (but not necessarily negative) levels of internal equity. From Figure 2, it can be seen that $k' < k^*$, $b' < b^*$, and $d < 0$. Because these firms start out with low levels of internal equity, they need to issue equity to reach even low levels of k'. Consequently, it is still beneficial to issue even a small amount of external equity to increase their investment. However, because of the cost, they do not issue as much as they would when $\lambda_0 = 0$. Nevertheless, even though they choose $b' < b^*$, it is the case that $b'/k' > b^*/k^*$. Because of the high cost of external equity, they still do substitute toward more debt relative to a lower level of k'. As internal equity increases they substitute external with internal equity while maintaining the same amount of investment and debt issuance.

[11] We should note that in the infinite-horizon version of this model, issuing debt will be associated with one more cost. In particular, the firm might want to issue less debt in case it ends up receiving a bad draw tomorrow and issuing costly equity to avoid default. This is a precautionary savings mechanism for the firm. In our two-period version there are only positive payments to shareholders in the second period.

Figure 2 Policy Functions



Notes: Both panels plot k', b', and d as a function of internal equity. Left panel: External equity is costless ($\lambda_0 = 0$). Right panel: External equity is costly ($\lambda_0 > 0$). The policy functions at the left panel coincide with the policy functions at the right panel for high values of internal equity.

Now consider firms with medium levels of internal equity. These firms choose $k' < k^*$ and $b' < b^*$, but also $d = 0$. Intuitively, the first-order conditions for b' and k' in (16) and (17) do not hold with equality. Because they have more internal equity, they avoid issuing costly external equity. Instead, they rely only on internal funds and debt to finance investment. However, firms in this region will use any additional internal equity to increase their investment (while maintaining $d = 0$). As a result, both k' and b' are increasing with e. Moreover, as firms obtain more internal equity, b'/k' is decreasing toward b^*/k^*.

Finally, consider firms with high levels of internal equity. These firms have so much internal equity that they are able to choose $k' = k^*$ and $b' = b^*$ without having to raise external equity. When external equity was costless, they chose $d > 0$. Costly external equity has no effect on them because they were not raising external equity anyway. Hence, their behavior coincides with the case of costless external equity where investment and debt issuance are constant and the firms are issuing positive dividends.

Cyclicality of Debt and Equity Issuance

Here we use our stylized framework to analyze the effects of productivity changes (z) on investment, debt, and equity issuance (k', b', and

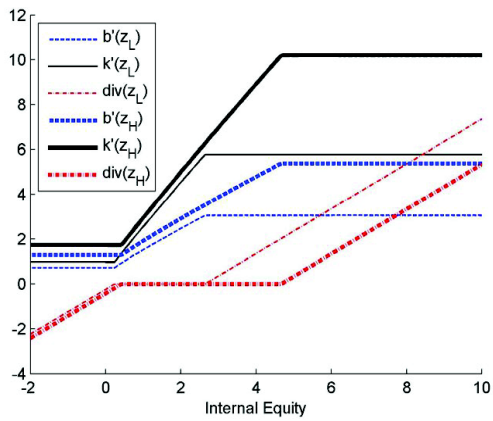
d, respectively). Figure 3 plots the policy functions for k', b', and d as a function of internal equity when external equity is costly. We plot the policy functions when productivity is low ($z = z_L$) and when productivity is high ($z = z_H$). A higher value of productivity will affect the firm's capital structure in two ways. First, internal equity $e(z, k, b) \equiv zf(k) + (1 - \delta)k - b$ will increase. Second, if shocks are autocorrelated (which is true in our simple example), a higher z in the first period will imply a higher expected z' in the next period. Using Figure 3, we can distinguish between the two since we plot how the policy functions change for a given amount of internal equity.

Looking at Figure 3, we see that higher productivity shifts k' upward since the marginal benefit of investing increases (see Equation [14]). This means that a fraction of previously unconstrained firms will find themselves constrained since the same amount of e will not be enough to sustain the larger amount of investment. Debt issuance b' will also increase. As firms invest more, the default threshold decreases for any given $b' > (1 - \theta)(1 - \delta)k'$ (i.e., $\partial z^*/\partial k' < 0$). This increases the borrowing capacity of the firm and lowers the marginal bankruptcy costs for each individual b'. Since the tax benefit of debt is $\tau q b'$, the higher borrowing capacity also increases the marginal benefit of issuing debt. Both effects cause b' to increase for a given level of internal equity. The increase in debt issuance is not uniform across firm sizes though. Smaller firms issue less debt than larger firms.

External equity issuance will increase (or dividend payout will decrease) in response to an increase in productivity. Firms with low amounts of internal equity e will increase their equity issuance to sustain a larger amount of investment. Since equity issuance is costly, they will change their issuance by only a small amount. Firms with a medium level of e will not issue equity or distribute any dividends. However, the set of (constrained) firms that do not distribute any dividends will increase. Similarly, firms with a high level of e will decrease the amount of dividends that they pay out.

Hence, for a given amount of internal equity our simple model predicts a procyclical debt and equity issuance. Of course, as stated before, e will also increase if z increases. A larger internal equity will represent a movement along the policy functions. This can potentially increase debt issuance but decrease external equity issuance (or increase dividend payout). So while debt issuance is definitely procyclical, equity issuance might be procyclical or countercyclical depending on how strong the opposing effects are. Based on Figure 3 it seems that for smaller firms the equity issuance is more likely to be countercyclical but for larger firms it is more likely to be procyclical.

Figure 3 Policy Functions for Different Levels of Productivity



Notes: We plot k', b', and d as a function of internal equity for a low value of productivity z_L and a high value of productivity z_H. Thin lines are used for z_L and thick lines are used for z_H. External equity is costly in both cases.

4. FULL MODEL

Utilizing the basic ingredients of our stylized two-period model in Section 3, we now build a fully dynamic model with heterogeneous firms and aggregate productivity shocks. Nevertheless, to keep the analysis simple, we assume a partial equilibrium framework.

Entrepreneurs and Firms

The economy is populated by a continuum of entrepreneurs. Each entrepreneur operates a firm. Entrepreneurs, and thus the firms they operate, differ with respect to their idiosyncratic productivity z. Firms are perfectly competitive and produce a single homogeneous good. Capital k and labor l are inputs into the firm's production function, $y = Az(k^{\alpha}l^{1-\alpha})^{\gamma}$, where A is aggregate productivity. We assume that $\gamma \in (0, 1)$, implying that there are decreasing returns to scale at the firm level. With the assumption of perfect competition, diminishing returns to scale enable heterogeneity to exist in equilibrium. Assuming a competitive labor market, the firm's profits can be denoted by

$$\pi(A, z, k) = \max_{l}\{Az(k^{\alpha}l^{1-\alpha})^{\gamma} - wl\}, \tag{18}$$

where w is the real wage. Since this is a partial equilibrium analysis, the wage w is normalized to 1.

We assume that both $\ln z$ and $\ln A$ follow an AR(1) process:

$$\ln z' = \rho_z \ln z + \varepsilon^z$$
$$\ln A' = \rho_a \ln A + \varepsilon^A,$$

where $\varepsilon^z \sim N(0, \sigma_\varepsilon^z)$ and $\varepsilon^A \sim N(0, \sigma_\varepsilon^A)$. Since z is an idiosyncratic shock, ε^z is assumed to be independent of ε^A. We denote by $F(z'|z)$ and $f(z'|z)$ the cumulative distribution and probability density functions for next period's productivity z', conditional on the current productivity z. Similarly, let $p(A'|A)$ denote the probability density function for A', conditional on current aggregate productivity A.

Every period firms choose how much capital to invest for next period k'. Investment is subject to a capital adjustment cost $g(k, k')$. We will assume that this function takes the form $g(k, k') = \phi\frac{(k'^2}{k}$. This will guarantee a gradual transition of firms toward their optimal size. Firms issue bonds b', which are priced at $q(A, z, k', b')$. This price will be determined endogenously based on the investment and debt issuance decisions of the firm as well as the idiosyncratic and aggregate shocks. As in Section 3, firms receive a tax subsidy from the government, $\tau q(A, z, k', b')b'$. Firms also have the option of distributing dividends ($d > 0$) or issuing equity ($d < 0$). As in Section 3, we assume that external equity is costly. However, now we specify the cost $\Lambda(A, d)$ as follows:

$$\Lambda(A, d) = \begin{cases} A^{-\lambda_1}\frac{\lambda_0}{2}d^2 & \text{if } d < 0 \\ 0 & \text{if } d \geq 0 \end{cases}.$$

Following Covas and Den Haan (2012), we assume that equity issuance costs are lower during expansions. This assumption will be critical to match the procyclicality of equity issuance in the data.

After the firm chooses k', b', and d, it may exit next period. We assume there are two reasons a firm may exit. First, a constant fraction η will exogenously be forced to exit. In this case, it is assumed that the entire firm value is destroyed. This implies that the firm will default and both the entrepreneur and lender will recover nothing. Second, depending on tomorrow's realization of A' and z', some entrepreneurs will endogenously default on their debt obligations. In this case, we assume that the firm is liquidated but that the entrepreneur lives on to found a new firm (a start-up). We discuss this default decision in the next subsection in more detail.

Default Decision

In deciding whether or not to default, the entrepreneur compares the value of "not defaulting" to the value of "defaulting." We define $V^{ND}(A, z, k, b)$ to be the value of not defaulting for a firm with state (A, z, k, b). Similarly, we define $V^D(A, z, k)$ to be the firm's value of default. These value functions will be defined below. Given these value functions, the firm's total value $V(A, z, k, b)$ is defined to be

$$V(A, z, k, b) = \max\left\{V^{ND}(A, z, k, b), V^D(A, z, k)\right\}. \tag{19}$$

If $V^{ND}(A, z, k, b) \geq V^D(A, z, k)$, the firm pays back its debt b and continues its operations. Otherwise, the firm chooses not to pay back its debt b and defaults.

The value of not defaulting, $V^{ND}(A, z, k, b)$, is then defined to be

$$V^{ND}(A, z, k, b) = \max_{d,k',b'} \left\{d - \Lambda(A, d) + \frac{1-\eta}{1+r} E\left[V(A', z', k', b')|A, z\right]\right\} \tag{20}$$

$$\text{s.t.} \quad d = \pi(A, z, k) + (1-\delta)k - b + (1+\tau)q(A, z, k', b')b' - k' - g(k, k'). \tag{21}$$

If the firm does not default, it chooses how much to invest (k'), how much debt it will issue (b'), and if it will distribute dividends ($d > 0$) or issue equity ($d < 0$). It makes these decisions subject to the budget constraint in (21). As noted earlier, the firm must also pay an equity issuance cost ($\Lambda(A, d) > 0$) if it issues equity ($d < 0$). Next period, with probability η, the entrepreneur receives the exogenous exit shock and receives nothing. With probability $1 - \eta$, however, the firm does not exogenously exit. In this case, depending on tomorrow's realization of A' and z', the firm can decide tomorrow whether to default or continue operating.

If the firm defaults, it shuts down its operations and is liquidated. Nevertheless, the entrepreneur can hide a fraction θ of the firm's undepreciated capital. Moreover, the entrepreneur can start a new firm next period. Hence, the owner can transfer his idiosyncratic productivity to a different project while eliminating his debt obligations. Given these assumptions, the value of defaulting, $V^D(A, z, k)$ is assumed to be

$$V^D(A, z, k) = \left\{\theta(1-\delta)k + \frac{1}{1+r} E\left[V^s(A', z')|A, z\right]\right\}, \tag{22}$$

where $V^s(A', z')$ is the value of a start-up tomorrow with aggregate productivity A' and idiosyncratic productivity z'. This value function will be defined later.

In general, we can define a threshold $z^*(A, k, b)$ such that firms with capital k, debt b, and idiosyncratic productivity lower than $z^*(A, k, b)$

will default. This threshold is defined to be the value of idiosyncratic productivity z^* such that the firm is just indifferent between defaulting and not defaulting:

$$V^{ND}(A, z^*, k, b) = V^D(A, z^*, k). \tag{23}$$

Consequently, this default threshold will depend on the aggregate level of productivity (A) as well as the firm's individual levels of capital (k) and debt (b). The default threshold z^* increases if debt b is large and decreases if capital k is large or if the economy is booming (A is high).

Bond Price

The firm issues bonds that are purchased by risk-neutral households. Households lend $q(A, z, k', b')b'$ to firms today, and in return the firm promises to pay b' next period. Given that the default is possible, the price $q(A, z, k', b')$ is set to guarantee the lender an expected return equal to the risk-free rate r. Consequently, the bond price will be given by

$$q(A, z, k', b') = \frac{1-\eta}{1+r}\left[1 - F(z^*(A', k', b')|z) + \frac{R(A, z, k', b')}{b'}\right], \tag{24}$$

where

$$R(A, z, k', b') = (1-\psi)(1-\theta)\int_0^\infty \int_0^{z^*(A',k',b')} (1-\delta)k' f(z'|z)p(A'|A)dz'dA'$$

is the unconditional recovery value of the bond. With probability η, the firm receives an exogenous exit shock and the lender receives nothing. However, with probability $(1-\eta)$, the firm does not receive an exit shock. In this case, the firm does not default with probability $1 - F(z^*|z)$ and the lender receives b'. However, if the firm defaults, then the lender receives fraction $(1-\psi)(1-\theta)$ of its undepreciated capital. The parameter θ controls how much of the capital stock the entrepreneur can hide while ψ reflects the bankruptcy costs.

Entry

As noted earlier, there are two reasons firms exit in this model. First, a fraction η of firms will exogenously exit. The entrepreneurs of these firms are assumed to be replaced by "new" entrants. Therefore, while a constant fraction of entrepreneurs exit each period, a constant mass of entrepreneurs are born each period. These new entrepreneurs are assumed to draw their initial idiosyncratic productivity from the invariant distribution for z. Second, some of the remaining firms will

endogenously choose to default. The entrepreneurs of these firms, however, are able to continue. In particular, these entrepreneurs can start a new firm (start-up) in the next period.

Therefore, in every period, firms will be destroyed and created at the same time. Because firms are assumed to be born with no capital, a start-up will have zero profits in the first period. Then, a start-up firm will choose how much to invest (k'). This investment can be financed by raising equity ($d < 0$) or by issuing debt (b'). Let $V^s(A, z)$ denote the value of a start-up with aggregate productivity A and idiosyncratic productivity z. This value is defined to be

$$V^s(A, z) = \max_{d,k',b'} \left\{ d - \Lambda(A, d) + \frac{1-\eta}{1+r} E\left[V(A', z', k', b')|A, z\right] \right\} \quad (25)$$

$$\text{s.t.} \qquad d = (1+\tau)q(A, z, k', b')b' - k'.$$

Therefore, the problem of a start-up is very similar to the problem of a continuing firm. However, a start-up begins its life with no debt and no assets. Because the start-up has no initial capital, it is assumed that it does not pay any capital adjustment costs.

Timing

The timing of the economy can be described as follows.

1. All entrepreneurs/firms receive productivity draws A and z.

2. A fraction η of firms are exogenously destroyed.

3. Surviving firms with state $\{z, A, k, b\}$ decide to default if $z < z^*(A, k, b)$. Firms that default exit.

4. Firms that did not default, as well as new start-ups, make investment and firm financing (debt and equity) decisions.

5. QUANTITATIVE ANALYSIS

In this section we quantitatively characterize our model of firm financing. We calibrate our model either using parameters commonly used in the literature or targeting specific moments computed in the data. We compare the model's predictions for the same set of statistics computed from Compustat in Section 2.

Table 3 Parameter Values

Parameter	Description	Value	Target
r	Real interest rate	0.04	Standard
δ	Depreciation rate	0.10	Standard
α	Capital share	0.36	Standard
γ	Returns to scale	0.65	Gomes and Schmid (2010)
η	Exit rate	0.04	Cooley and Quadrini (2001)
ψ	Bankruptcy cost	0.25	Arellano, Bai, and Zhang (2012)
λ_0	Equity issuance cost	0.75	Covas and Den Hann (2012)
λ_1	Equity issuance cost	20	Covas and Den Hann (2012)
ρ_z	Persistence of z	0.55	Clementi and Palazzo (2014)
ρ_a	Persistence of A	0.68	Clementi and Palazzo (2014)
σ_ε^z	Standard deviation of ε^z	0.18	S.D. of sales growth
σ_ε^A	Standard deviation of ε^A	0.016	Clementi and Palazzo (2014)
τ	Tax credit	0.07	Mean leverage
θ	Hidden fraction	0.93	Mean default
ϕ	Capital adjustment cost	0.10	Mean of sales growth

Notes: This table reports the parameter values used in the quantitative model. Each parameter is calibrated either based on the literature or targeting a specific moment.

Calibration

All parameter values are reported in Table 3. The model is computed at an annual frequency. We normalize the wage rate to 1 and set an annual risk-free rate of 4 percent. The depreciation rate is set at 10 percent, a value commonly employed in the literature. The capital share equals $\alpha = 0.36$ and, based on Gomes and Schmid (2010), the decreasing returns to scale parameter is $\gamma = 0.65$. The firms' exit rate η is set to 0.04 based on Cooley and Quadrini (2001). Bankruptcy cost equals $\psi = 0.25$ based on Arellano, Bai, and Zhang (2012). Following Covas and Den Haan (2012), we assume that equity issuance costs are lower during expansions and set $\lambda_0 = 0.75$ and $\lambda_1 = 20$.

The persistence of idiosyncratic productivity $\rho_z = 0.55$ is based on Clementi and Palazzo (2014). Although the authors provide an estimate for σ_ε^z, we choose to use this parameter to match a specific moment (see below). We also borrow their estimates to calibrate the persistence and standard deviation of the aggregate productivity process. In particular, ρ_a is set to 0.68 and σ_ε^A is chosen to be 0.016.

The remaining parameters, $\{\tau, \theta, \phi, \sigma_\varepsilon^z\}$, are chosen to match specific model moments. In particular, a higher tax benefit τ will encourage firms to issue more debt and increase their leverage ratio. Therefore, to match the mean leverage ratio observed in Compustat, τ is set to 0.02. Conditional on the value of ψ, a larger value of θ induces more

firms to default since they can hide and keep a larger fraction of their assets. Thus, to match the mean default rate in the economy, θ is set to 0.93. The adjustment cost parameter ϕ affects how fast firms grow. Hence, to match the average cross-sectional growth rate of sales, ϕ is set to 0.10. Finally, a larger dispersion in idiosyncratic productivity will lead to a larger dispersion in the growth of sales. With a value of $\sigma_\varepsilon^z = 0.18$, the model matches the cross-sectional standard deviation of sales growth.

Steady-State Results

We start by characterizing the steady state of the economy. In the steady state, aggregate productivity is constant in every period ($A = 1$). Based on our policy functions, we simulate a panel of firms and track their behavior over time. We use the stationary distribution to construct several statistics and compare them to the ones computed from Compustat. Table 4 gives a summary of the results.

In Compustat the distribution of leverage across firms is found to be highly skewed to the right. Excluding firms at the top 1 percent of the distribution, the average leverage ratio is 27 percent. Our model economy is able to match this statistic by targeting the tax credit τ. In contrast, leverage ratios are more dispersed in the data than our model. The standard deviation of leverage in Compustat is 0.37, much higher than the model's result of 0.15. A reason for this failure is the relatively low value for the persistence of idiosyncratic productivity ρ_z. If idiosyncratic shocks are not very persistent then even unproductive firms can easily get access to credit. Indeed, we have experimented with higher values of ρ_z and found that the standard deviation of leverage increases. Moreover, the model can perform well with respect to sales growth. The mean of sales growth in the model is 0.12, very close to the value computed in the data (0.11). This statistic was targeted using the adjustment cost parameter ϕ. The model can also capture the dispersion in sales growth rates (0.45 in the model versus 0.51 in the data). To match this moment, we used the dispersion of idiosyncratic productivity shocks σ_ε^z.

Table 4 Steady-State Results

	Data-Compustat			Model		
Statistic	**Small Firms**	**Large Firms**	**All**	**Small Firms**	**Large Firms**	**All**
Mean (Leverage)	0.27	0.28	0.27	0.30	0.23	0.27
S.D. (Leverage)	–	–	0.34	–	–	0.15
Mean (Sales Growth)	0.12	0.10	0.11	0.21	0.04	0.12
S.D. (Sales Growth)	–	–	0.51	–	–	0.45

Notes: This table shows the steady-state results for the mean and standard deviation of leverage and sales growth, respectively. Statistics on leverage and sales growth are calculated from the data (Compustat).

We next compare the behavior of small versus large firms. Using data from Compustat and consistent with Rajan and Zingales (1995) and Cooley and Quadrini (2001), we find a positive relationship between leverage and total assets. However, the differences seem to be minor as firms with assets smaller than the median have a leverage equal to 0.27 while firms with assets larger than the median have a leverage equal to 0.28. In our model, these numbers are 0.30 and 0.23, respectively. In Section 3, we saw that as firms obtain more internal equity, the ratio $\frac{b'}{k'}$ decreases. Smaller firms (with lower internal equity) substitute more toward debt to avoid using costly external equity. Moreover, due to decreasing returns to scale, the model replicates qualitatively the empirical observation that smaller firms grow faster. In the model, sales growth is 0.21 for small firms and 0.04 for large firms. In Compustat, these numbers are 0.12 and 0.10, respectively. In general the model captures the basic features of the data with some success.

Business Cycle Results

We now allow the economy to experience aggregate productivity shocks. To avoid further computational complexity we assume that the prices do not adjust in response to productivity changes. If we allowed for a general equilibrium framework, we would have to keep track of the distribution of firms over debt, capital, and equity, which would greatly increase the state space.

Table 5 reports the correlation between debt and equity issuance with aggregate output. To facilitate the comparison with the data, we include information from the top panel of Table 2 that excludes only major mergers from the sample. In Section 2, we showed that mergers are an important way that firms raise equity. The model replicates the positive correlation between debt issuance and aggregate output (0.868 in the model versus 0.785 in the data). As explained in Figure 3, a higher productivity increases k', allowing the firm to issue more debt. Table 5 also reports how the cyclicality differs among small and large firms. In Section 2, we documented that the cyclicality is stronger for larger firms (excluding the top 1 percent). Our model replicates this pattern and can match very closely the cyclicality of firms in the [75, 99] bin (0.737 in the model versus 0.755 in the data). In response to an increase in productivity, a small firm may disproportionately increase b' by disproportionately decreasing external equity issuance. In contrast, large firms that issue a small amount of external equity will

Table 5 Business Cycle Results

	Size Class (Percent)				
Data	**[0, 50]**	**[50, 75]**	**[75, 99]**	**[99, 100]**	**[0, 100]**
ΔD	**0.536**	**0.611**	**0.755**	**0.547**	**0.785**
$\Delta E(1)$	**0.345**	0.191	0.016	0.044	0.096
$\Delta E(2)$	0.243	−0.250	**−0.617**	−0.312	**−0.509**
E^I	**0.353**	0.268	0.306	0.250	**0.363**
E^P	0.069	0.279	**0.654**	0.314	**0.588**
	Size Class (Percent)				
Model	**[0, 50]**	**[50, 75]**	**[75, 99]**	**[99, 100]**	**[0, 100]**
ΔD	0.260	0.244	0.737	0.277	0.868
ΔE	0.447	−0.326	−0.714	−0.942	−0.764
E^I	0.528	0.445	0.356	—	0.386
E^P	0.287	0.335	0.715	0.942	0.759

Notes: This table reports the model-generated business cycle properties of debt and equity issuance. This table also reports empirical statistics as calculated in Section 2. For simplicity we report empirical measures that exclude only major mergers from our sample. For the empirical section, we show coefficients that are significant at the 5 percent level in bold.

increase b' in a relatively proportional manner.[12] As a result, we find the correlation between debt issuance and output (productivity) to be much higher in the case of large firms. A similar nonlinearity occurs for the largest firms when they start distributing dividends, which explains why the correlation decreases for that group.

The model also generates a countercyclical equity issuance. In Section 2, we documented that equity issuance can be weakly procyclical or countercyclical depending on the way we measure equity. Moreover, we have shown that much of the procyclicality is due to raising equity through mergers and that the cyclicality becomes negative if we just consider net sale of stock. In the model, net equity issuance ΔE is strongly countercyclical. Similar to the data, we break net equity issuance ΔE into a gross equity issuance E^I and a gross equity payout E^P component, with $\Delta E = E^I - E^P$. Our decomposition reveals that the strong countercyclicality of net equity issuance is driven by a strongly procyclical gross dividend payout. In the model, smaller firms prefer to raise more gross equity than paying out gross dividends during expansions. This leads to a procyclical equity finance for firms in the [0, 50] bin, similar to what we observe in the data. Nevertheless, the

[12] To understand these properties better we refer the reader to Figure 3. Although in our fully dynamic model we include quadratic capital adjustment costs and quadratic equity issuance costs, the basic properties of the policy functions remain intact.

procyclicality of equity issuance for small firms relies on our assumption (among others) of countercyclical equity issuance costs. Overall, the model is consistent with the empirical patterns we see in equity financing.

6. CONCLUSION

This article provides an introductory, yet comprehensive, business cycle analysis of firm financing. We first document several empirical patterns of debt and equity issuance based on data from Compustat. While we find that debt issuance is strongly procyclical, the cyclicality of net equity issuance depends on the exact definition used. If we define equity using the net sale of stock (following Jermann and Quadrini [2012]), we find net equity issuance to be countercyclical. Alternatively, if we define equity issuance using the change in the book value of equity (following Covas and Den Haan [2011]), we find net equity issuance to be weakly procyclical. Nevertheless, we find that equity financing through mergers and, to a lesser extent, stock compensation can explain much of the discrepancy between the two measures. Moreover, regardless of the measure used, the countercyclicality of net equity issuance is driven by a strongly procyclical gross payout to equity and not countercyclical gross equity issuance. Overall, these empirical findings should be useful in evaluating theoretical models, which stress the role of the financial sector in propagating aggregate fluctuations. Of particular interest, perhaps, is the heterogeneous behavior of firm financing and the role of mergers and acquisitions.

To help build intuition, we analyze the firm's optimal capital structure within a simple two-period model. Then, to determine how well our framework can match the cyclical properties of firm financing, we build a fully dynamic quantitative model. The model features heterogeneous firms that endogenously choose their capital structure by balancing the tax benefits against the bankruptcy costs of debt issuance and the expenses associated with equity issuance. The model generates a procyclical debt and countercyclical net equity issuance. Moreover, the model can match the firm-size relationship regarding debt and especially equity issuance. Overall, the model is useful for illustrating the important mechanisms involved. While firms issue more debt to finance more investment, the model highlights that equity issuance provides conflicting motives for the firm. On the one hand, firms would like to issue more equity (which may be costly) to finance more investment. On the other hand, firms would like to pay out more dividends in good times. For most firms the second effect dominates in our model.

However, to generate procyclical net equity issuance for small firms, we assume that equity issuance costs are lower during expansions.

APPENDIX A: DATA SOURCES

We obtain annual data from Compustat between 1980 and 2013. We exclude financial firms (SIC 6000–6999) and utilities (SIC 4900–4999). We drop any firm-year observations if we do not have any information on assets, capital stock, debt, or both equity measures. We drop observations that violate the accounting identity by more than 10 percent. We drop firms affected by 1988 accounting change (GM, GE, Ford, Chrysler).[13] We only include firms reporting in USD. One important concern is whether we include firms affected by a merger or an acquisition. For this purpose, we separately report our results for two cases. In the first case, we follow Covas and Den Haan (2011) and drop all firm-year observations that are affected by a "major" merger or acquisition. By "major" we mean that the merger or acquisition causes the resulting firm's sales to increase by more than 50 percent. In the second case, we drop all observations affected by any kind of merger. To identify whether a firm was involved in a merger, we use the footnote code on sales. Compustat assigns the footnote code AB if the data reflects a major merger or acquisition. Meanwhile, footnote code AA reflects other acquisitions.

SE is defined as the book value of stockholder's equity (data item #216) minus retained earnings (data item #36). $\Delta E(1)$ is defined to be the annual change in SE minus cash dividends (data item #127). The net sale of stock is defined to be the funds received from the issuance of common and preferred stocks (data item #108) minus equity repurchases (data item #115). $\Delta E(2)$ is defined to be the net sale of stock minus cash dividends. RE is the balance sheet item for retained earnings (data item #36). w_S is stock compensation (data item #398). Sales is given by data item #12, which represents gross sales (i.e., the amount of actual billings to the customers). Total assets is the book value of assets (data item #6). We define debt as the sum of debt in current liabilities (data item #34) and long-term debt (data item #9). The capital stock K is (net) property, plant, and equipment (data item

[13] See Bernake, Campbell, and Whited (1990) for details.

#8). Investment I equals capital expenditures on property, plant, and equipment (data item #30).

And finally, we obtain real corporate GDP from the Bureau of Economic Analysis's National Income and Product Accounts. Particularly, we use Table 1.14, which reports the gross value added of domestic nonfinancial corporate business, in billions of chained (2009) dollars.

APPENDIX B: SIMPLIFIED VALUE FUNCTION

In (12), the firm's problem was given by

$$V(z,k,b) = \max_{d,k',b'} \left\{ \begin{array}{c} d - \Lambda(d) + \\ \frac{1}{1+r}\int_0^\infty \max\left\{V^{ND}(z',k',b'), V^D(z',k')\right\} f(z'|z)dz' \end{array} \right\},$$

subject to the budget constraint, which is

$$d + k' = e(z,k,b) + (1+\tau)q(z,k',b')b'.$$

Using the definitions of $V^{ND}(z',k',b')$ and $V^D(z',k')$ in (8) and (9), we can re-write the firm's value function as follows:

$$V(z,k,b) = \max_{d,k',b'} \left\{ \begin{array}{c} d - \Lambda(d) + \frac{1}{1+r}\int_0^\infty [z'f(k') + (1-\delta)k']f(z'|z)dz' \\ -\frac{(1-F(z^*(k',b')|z))b'}{1+r} \\ -\frac{1-\theta}{1+r}\int_0^{z^*(k',b')}[z'f(k') + (1-\delta)k']f(z'|z)dz'] \end{array} \right\}.$$

When we substitute for d using the firm's budget constraint, this becomes

$$V(z,k,b) = \max_{d,k',b'} \left\{ \begin{array}{c} e - k' - \Lambda(d) + (1+\tau)q(z,k',b')b' \\ +\frac{1}{1+r}\left[E[z'f(k')|z] + (1-\delta)k'\right] \\ -\frac{(1-F(z^*(k',b')|z))b'}{1+r} - \\ \frac{1-\theta}{1+r}\int_0^{z^*(k',b')}[z'f(k') + (1-\delta)k']f(z'|z)dz'] \end{array} \right\}.$$

Moreover, in (11), the bond price was defined to be

$$q(z,k',b') = \frac{1}{1+r}\left[1 - F(z^*(k',b')|z) + \frac{R(z,k',b')}{b'}\right],$$

where

$$R(z,k',b') \equiv (1-\psi)(1-\theta)\int_0^{z^*(k',b')} \left[z'f(k') + (1-\delta)k'\right] f(z'|z)dz'.$$

Therefore, using that $qb'(1+r) = [1 - F(z^*)]\, b' + R$, we arrive at (13):

$$V(z,k,b) = \max_{d,k',b'} \left\{ \begin{array}{l} e - k' - \Lambda(d) + \tau q(z,k',b')b' + \\ \frac{1}{1+r}\left[E[z'f(k')|z] + (1-\delta)k'\right] + \\ \quad - \dfrac{\psi(1-\theta)}{1+r} \displaystyle\int_0^{z^*(k',b')} [z'f(k') + (1-\delta)k']f(z'|z)dz' \end{array} \right\}.$$

REFERENCES

Arellano, Cristina, Yan Bai, and Jing Zhang. 2012. "Firm Dynamics and Financial Development." *Journal of Monetary Economics* 59 (6): 533–49.

Begenau, Juliane, and Juliana Salomao. 2014. "Firm Financing over the Business Cycle." Working Paper, Stanford University.

Bernanke, Ben S., John Y. Campbell, and Toni M. Whited. 1990. "U.S. Corporate Leverage: Developments in 1987 and 1988." *Brookings Papers on Economic Activity* 21 (1): 255–86.

Clementi, Gian Luca, and Berardino Palazzo. 2014. "Entry, Exit, Firm Dynamics and Aggregate Fluctuations." Working Paper, New York University.

Cooley, Thomas F., and Vincenzo Quadrini. 2001. "Financial Markets and Firm Dynamics." *American Economic Review* 91 (December): 1,286–310.

Covas, Francisco, and Wouter J. Den Haan. 2011. "The Cyclical Behavior of Debt and Equity Finance." *American Economic Review* 101 (April): 877–99.

Covas, Francisco, and Wouter J. Den Haan. 2012. "The Role of Debt and Equity Finance over the Business Cycle." *Economic Journal* 122 (December): 1,262–86.

Eisfeldt, Andrea L., and Adriano A. Rampini. 2006. "Capital Reallocation and Liquidity." *Journal of Monetary Economics* 53 (April): 369–99.

Fama, Eugene F., and Kenneth R. French. 2005. "Financing Decisions: Who Issues Stock?" *Journal of Financial Economics* 76 (June): 549–82.

Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Petersen. 1988. "Financing Constraints and Corporate Investment." *Brookings Papers on Economic Activity* 1988 (1): 141–195.

Gomes, Joao F., and Lukas Schmid. 2010. "Levered Returns." *Journal of Finance* 65 (April): 467–94.

Gourio, Francois. 2013. "Credit Risk and Disaster Risk." *American Economic Journal: Macroeconomics* 5 (July): 1–34.

Hennessey, Christopher A., and Toni M. Whited. 2007. "How Costly is External Financing? Evidence from a Structural Estimation." *Journal of Finance* 62 (August): 1,705–45.

Jermann, Urban, and Vincenzo Quadrini. 2012. "Macroeconomic Effects of Financial Shocks." *American Economic Review* 102 (February): 238–71.

Katagiri, Mitsuru. 2014. "A Macroeconomic Approach to Corporate Capital Structure." *Journal of Monetary Economics* 66: 79–94.

Modigliani, Franco, and Meron H. Miller. 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment." *American Economic Review* 48 (June): 261–97.

Rajan, Raghuram G., and Luigi Zingales. 1995. "What Do We Know about Capital Structure? Some Evidence from International Data." *Journal of Finance* 50 (December): 1,421–60.