NEW DEVELOPMENTS IN FIRM DYNAMICS IN UNDERSTANDING BUSINESS DYNAMISM

Declining Business Dynamism: What We Know and the Way Forward

By Ryan A. Decker, John Haltiwanger, Ron S. Jarmin, and Javier Miranda

Creative destruction has long been considered an important element of market economies. Continual creation and destruction of jobs, firms, and industries is closely linked to growth in employment, productivity, and living standards, though it can be costly for individuals.

Measures of creative destruction include data on gross job or worker flows and entry and exit rates of new firms or establishments. These data reveal rich dynamics underlying the net changes in employment or output that dominate popular discussion. Gross flows of jobs and workers are large in comparison to net employment growth (Davis, Haltiwanger, and Schuh 1996; Davis, Faberman, and Haltiwanger 2012). New firms account for a significant share of activity in a given year, and young firms make a long-lasting contribution to aggregate employment despite high failure rates (Haltiwanger, Jarmin, and Miranda 2013).

The pace of creative destruction has declined in recent decades in the United States. What do we know about the decline in business dynamism? What are its potential implications?

I. Recent Developments

Since the early 1980s, various measures of entrepreneurial activity in the United States have seen a secular decline. Decker et al. (2014) show that this decline can be seen in falling young firm activity as a share of firms, job creation, and employment (where “young” firms have age five or less). Pugsley and Şahin (2015) provide evidence that the dearth of young firms has rendered the United States less responsive to business downturns, highlighting a possible cause of “jobless” recession recoveries. But the economy-wide declines mask important variation across sectors. The Retail Trade, Services, and Manufacturing sectors display clear downward trends in the employment share of young firms, while the Information sector had rising young firm activity until around 2000, after which even that sector saw a decline (Decker et al. forthcoming).

Given the heterogeneity in entrepreneurial types highlighted by Schoar (2010) and Hurst and Pugsley (2012), a key question is whether the decline in entrepreneurship reflects fewer “subsistence” and “lifestyle” entrepreneurs or fewer “transformational” entrepreneurs with high growth potential. Decker et al. (forthcoming) show that, since 2000, high-growth entrepreneurship has joined the broader decline in entry, with a large decline in the ninetieth percentile of the young firm growth rate distribution.
Another key dynamism concept is labor market churn, typically measured as gross job or worker flows. Gross job flows and within-firm employment volatility have been declining since the early 1980s (Decker et al. 2014). This decline is manifest in both creation and destruction, and it occurs in a “stair-step” pattern that appears to be catalyzed by recessions. Changes in the composition of the US economy cannot explain the decline. The US industrial shift from Manufacturing to Services works to increase flows, since Services tends to exhibit higher flows than Manufacturing. The aging of the firm distribution brought on by declining entrepreneurship has some explanatory power, since young firms are typically more dynamic (though declining entry cannot fully explain the gross flows decline as all age classes have seen lower flow rates). However, controlling for composition effects along firm size, firm age, and industry still accounts for less than one-third of the decline in job reallocation (Decker et al. 2014).

Gross job flows can be thought of as a second moment (dispersion) of the employment growth rate distribution, and indeed, measures of the interdecile range of firm growth rates closely track measures of job reallocation. Prior to 2000, the decline in the interdecile range reflected similar declines in the gap between the tenth and fiftieth percentiles and the gap between the ninetieth and fiftieth percentiles, suggesting that declining dispersion was not broadly accompanied by changes in the third moment, skewness (Decker et al. forthcoming). However, the character of the change in dispersion changed around 2000, after which the 90/50 gap fell much more than the 50/10 gap. That is, the decline in dispersion began to reflect a decline in skewness, indicating a reduction in high-growth firm activity. This change in the character of the dispersion decline is particularly pronounced in the High-Tech and Information sectors, which actually saw increasing dispersion prior to 2000 but have seen large skewness-driven declines in dispersion since then. Services likewise shows large skewness declines particularly after 2000, while the dispersion decline in Retail Trade has always involved only very little change in skewness.

Worker flows similarly show a decline in recent decades, even relative to job flows (Davis, Faberman, and Haltiwanger 2012). Hyatt and Spletzer (2013) show that the decline in job flows explains at most one-third of the decline in worker flows, and demographic composition effects account for less than one quarter. There has been a related decline in short-duration jobs, which raises questions about potential rising returns to tenure. Evidence suggests that the trend in worker flows is also driving a decline in interstate migration (Molloy, Smith, and Wozniak 2014), highlighting the pervasive consequences of labor market dynamics.

It is important to note that the trends discussed above may in some measure be appearing in other countries as well. Most developed countries saw declines in young firm activity between 2001 and 2011 (Criscuolo, Gal, and Menon 2014), and several countries saw declines in job reallocation between 2002 and 2009 (Davis and Haltiwanger 2014).

II. Interpreting the Data through Models

Canonical models of firm dynamics provide clues as to the causes of declining dynamism. Hopenhayn (1992) describes an industry equilibrium model that is at the core of most modern firm dynamics treatments. The model consists of a continuum of incumbent firms along with endogenous entrants and exiters. Firms receive time-varying idiosyncratic productivity shocks, and these shocks completely determine firm-level labor demand.

A firm’s optimal scale is determined by its productivity, so a firm that receives a productivity increase (decrease) creates (destroys) jobs. In the absence of adjustment costs these responses are instant, with firms immediately reaching their optimal scale; so the pace of gross job flows is directly related to the dispersion and persistence of productivity shocks. Entry is determined endogenously by a free entry condition. A firm exits when its continuation value falls below the fixed costs of continued operation. Both entry and exit are increasing in the persistence of the productivity shock, since each new productivity draw changes a (potential) firm’s profitability relative to entry costs and fixed costs. This model gives rise to rich empirical predictions, including entry, exit, and a high pace of job flows, without the use of aggregate shocks.

A key implication of the Hopenhayn model is that gross job flows as well as entry and exit rates depend on the persistence and dispersion of productivity shocks. In this model, then, a decline in gross flows and entry would primarily
be the result of declining dispersion or persistence of productivity or profitability shocks, but more realistic models generate an additional possible explanation. Hopenhayn and Rogerson (1993) model frictions to reallocation by adding a tax on job destruction to the Hopenhayn model and setting it in general equilibrium.

Adjustment costs reduce the firms’ desire to destroy jobs in the face of lower expected productivity, creating a form of specificity that also reduces firms’ desire to create jobs in the face of higher expected productivity. The result is ranges of inaction such that firms prefer maintaining nonoptimal scale to making small labor demand adjustments, dampening their responsiveness to their productivity draws and thereby reducing both job creation and job destruction. This qualitative result generalizes to other forms of specificity; for example, search-and-matching models of labor markets can reduce gross flows with vacancy costs, hiring costs, or higher job match surplus (or lower outside payoffs).

The model-based insights discussed above suggest empirical strategies for narrowing the set of possible causes of declining dynamism. Falling volatility indicates either falling dispersion and/or persistence of idiosyncratic business productivity, or falling responsiveness of businesses to their productivity. Both possibilities can be assessed empirically.

Recent evidence from US Manufacturing indicates that the dispersion of establishment-level total factor productivity has actually risen in the post-2000 period with relatively steady or mildly increasing dispersion in the 1980s and 1990s (Decker et al. 2016; for international evidence see Andrews, Criscuolo, and Gal 2015). The persistence of establishment-level productivity has been roughly flat during the same period, with AR(1) coefficients varying nonmonotonically between 0.6 and 0.7 in high tech and between 0.55 and 0.60 among other firms. If anything, the productivity distribution data suggest that one should expect rising gross job and worker flows, rather than falling flows as in the data.

Conversely, Decker et al. (2016) find evidence of declining productivity responsiveness. Historically, there has existed a robust relationship between establishment-level productivity and employment growth, with high productivity being associated with growth and low productivity with contraction or exit. They find these relationships have weakened over time, particularly since 2000. This can be seen in establishment-level regressions of employment growth and survival on estimated total factor productivity for the Manufacturing sector.

Results for young firms in High-Tech Manufacturing industries are particularly notable. These firms have historically displayed stronger productivity responsiveness than has the rest of the economy. The responsiveness of these young High-Tech establishments rose in the 1990s then fell dramatically in the 2000s, a pattern that mirrors the rise and fall of growth rate skewness and dispersion among young High-Tech firms over the same period shown by Decker et al. (forthcoming). That is, the rise then fall of high-growth activity among young High-Tech firms since 1980 has coincided with a strengthening then weakening of the productivity/growth relationship at the establishment level. Among businesses outside the High-Tech industry, productivity responsiveness has steadily declined since the 1980s, consistent with the coincident decline of growth rate dispersion and skewness seen in the broad economy outside high tech.

III. Searching for Explanations

The evidence is consistent with the presence of adjustment frictions or other factors that are changing the payoff to business scale adjustments. While this result narrows the field of explanations for changing dynamism, it still leaves many possible candidates. One candidate could be policy frictions that reduce incentives for labor adjustment, such as unlawful discharge regulations or occupational licensing; Davis and Haltiwanger (2014) review relevant studies and provide suggestive evidence. Another candidate may be financial frictions that might arise as a result of bank consolidation or perceived changes in financial risk or be exacerbated by increased need for external capital, as in Caballero and Hammour (2005).

In the High-Tech sector, another possible cause of declining productivity responsiveness during the post-2000 period is the transition from “general-purpose” to “special-purpose” equipment manufacturing in the United States.¹

¹We thank Christopher Foote for this insight.
Byrne (2015) shows that US production of general-purpose electronics declined markedly during the 2001 recession, with further declines since then. This post-2000 decline was accompanied by a strong increase in special-purpose electronics that trade in thinner markets and benefit less from scale economies. Businesses manufacturing these special-purpose products may be less responsive to productivity due to demand constraints or uncompetitive environments that reduce adjustment imperatives. This potential driver of declining responsiveness can be investigated through close examination of the affected industries as well as broader examination of industries outside of High-Tech Manufacturing and Manufacturing generally.

The evidence discussed above describes the response of labor demand to productivity shocks, but within-establishment employment growth is not firms’ only means of adjustment. Firms may increasingly adjust to shocks through investment in physical or intangible capital. Software may provide unprecedented ability to scale up without adding labor. Adjustment may occur in other establishments of affected firms. There may be an increasingly international dimension to productivity responsiveness; exposure to foreign trade may affect the responsiveness of domestic plants, or firms may facilitate adjustment by scaling international operations. Each of these hypotheses carries empirical predictions that can be taken to the data.

The changing nature of network externalities faced by certain types of young firms may also hold explanatory power. The platform issues that have long been associated with the High-Tech industry appear to have intensified in recent years with the rise of businesses like Facebook or more generally with the maturation of the industry and as larger firms become dominant. There may be a related shift in the objectives of young firms from desire to grow indefinitely to a desire to be acquired early. Firms facing these objectives may behave very differently from the firms that are typically conceptualized by economists, requiring more creativity from both theoretical and empirical researchers.

Regardless of cause, declining productivity responsiveness implies a decline in the contribution of reallocation to aggregate productivity growth. The reallocation contribution has been significant historically (Decker et al. 2016), so a decline in productivity responsiveness implies slower productivity growth holding other things equal. However, it may be that benign explanations for declining responsiveness yield more within-plant productivity growth or a contribution of reallocation through other margins (e.g., physical or intangible capital). The implications for the future of productivity growth in the United States highlight the importance of further research into declining business dynamism.

REFERENCES


Declining Business Dynamism


