A Technological and Organisational Explanation for the Size Distribution of Firms

Joshua S. Gans
John Quiggin

ABSTRACT. This paper combines insights from the literature on the economics of organisation with traditional models of market structure to construct a theory of equilibrium firm size heterogeneity under the assumption of a homogeneous product industry. It is possible that configurations consisting entirely of small firms (run by entrepreneurs with limited attention) and with larger firms (using managerial techniques to substitute away these limits to allow increasing returns technologies to become profitable) can arise in equilibrium. However, there also exist equilibrium configurations with the co-existence of large and small firms. The efficiency properties of these respective equilibria are discussed. Finally, the implications of an expanding market size are considered.

It is a commonly observed characteristic of virtually every industry that firms of different sizes co-exist for long periods of time (see Simon and Bonini, 1958; Audretsch, 1995; and Sutton, 1997, 1998). This empirical phenomenon contrasts with traditional industrial organisational theory that predicts that firms with access to the same technologies will, in a competitive environment, operate at a similar scale of operation. Such considerations form the basis of competitive market theory (Viner, 1931; Stigler, 1958) and of the more recent theory of contestable markets (Baumol, Panzar and Willig, 1982).

The competitive model does not provide an explanation of firm size heterogeneity because it is based on an implicit assumption that all technologies and organisational arrangements are infinitely reproducible. Under such an assumption, all firms use the technology generating the lowest average costs. However, when one technology or organisational arrangement uses an essential entrepreneurial input (or a factor supplied by an increasing cost industry), reproduction cannot be undertaken without rising costs. Thus, average costs involved in using that technology will rise as output expands. Only a few firms might be able to enter using that technology, leaving the market to other firms with technologies of different minimum efficient scales. In this paper, we show that allowing for limited reproducibility of this kind can generate equilibrium firm size heterogeneity in a competitive/free entry setting.

By using a competitive framework, our approach stands in contrast to previous theoretical work on the size distribution of firms. Ijiri and Simon (1977) tackled the regularity of industry configurations with a few large firms and many small firms by using stochastic methods in association with Gibrat’s Law, that is, the observation that growth rates of firms were independent of their size. They showed that a purely stochastic mechanism could account for many of the stylised facts regarding the observed size distribution of firms in many industries. Sutton (1998) has built on the stochastic approach. He views firms as clusters of investment opportunities such as plants, outlets and products that are at least as likely to be integrated in large as in small firms and incumbent firms as opposed to entrants. Under such conditions, he shows that the distribution of firm sizes will be skewed with industry assets concen-
trated in the hands of a few firms. Sutton then
demonstrates that the empirically observed
skewness is consistent with this viewpoint.

Not all of the features of the observed distrib-
ution of firms appear to be consistent with these
stochastic models. They are based on the assump-
tion that not only the mean but also the distribu-
tion of growth rates is unrelated to firm size. As
a result, the distribution of firm sizes they generate
is unimodal (typically lognormal). In many
industries, such as retailing and restaurants, we
observe a seemingly bimodal pattern with many
one and two-establishment enterprises and a small
number of large enterprises operating hundreds of
establishments. In the service industries, large
enterprises take the form of one or more chains
of small establishments, operated either on a fran-
chise basis or fully owned by the parent company.
The stochastic approach would appear to predict
larger numbers of middle-sized enterprises, say,
those operating between two and one hundred
establishments, than is in fact the case for those
industries.

A similar objection can be raised with regard
to the competitive models of Lucas (1978), and
Calvo and Wellisz (1980). They postulated an
underlying distribution of entrepreneurial or
managerial talent in the economy and then showed
that purposeful selection by agents into worker and
managerial roles, combined with an assump-
tion of Gibrat’s Law, could provide a competitive
foundation for heterogeneous firm sizes in equi-
librium. This heterogeneity was, however, fully
accounted for by the distribution of entrepreneurial
ability. In the absence of detailed information
about distributions of the ability of entrepreneurs
attracted to different industries, their model had
limited predictive power.3

In contrast to either the stochastic process or
static equilibrium approaches, Jovanovic (1982)
examines the firm size issue from a dynamic
perspective. He argues that when firms begin
production they discover potentially useful infor-
mation regarding their costs; in particular, the
costs associated with expansion. Those firms
receiving favourable cost signals grow and
become large firms. Those receiving unfavourable
signals exit. At any point in time, therefore, large
established firms and small firms who may only
be there temporarily may populate an industry.

This theoretical line can explain a potentially
bimodal firm size distribution.4

In this paper, we combine insights regarding the
technology of firms and the economics of organi-
sation to build a simple but, we believe, very
general theory of why one should expect hetero-
genility in firm size and, in particular a bimodal
size distribution, in many industries. In so doing,
we assume that free entry and exit exist in a strong
form, in that firms can be established using any
available technology or organisational form. There
are no barriers to production based on differential
resources or exclusive access to key technologies –
established firms and potential entrants are
symmetric. In addition, our model is of an industry
with a homogenous product, ruling out explana-
tions for firm size differences based on niche
exploitation or quality search considerations (see
Salop and Stiglitz, 1977; and Bagwell and Ramey,
1994). Given our focus on the choice of low cost
technology and organisational form by firms,
assuming no start-up barriers and a homogenous
product puts the maximum pressure on costly
production methods to be excluded from produc-
tion.

As a result, and in contrast to Jovanovic (1982),
our theory is a static equilibrium approach. It
can, therefore, explain the co-existence of large
and small firms in an industry without requiring
differences in survival or learning processes
associated with entry. Our theory’s value is in its
simplicity and its potential to rationalise the per-
sistence of firm size distributions across almost all
industries; including those that may be more
mature and where small firms may persist but stay
small.

Section I outlines a basic model with these
elements. There we postulate two organisational
modes. First, there are entrepreneurial firms.
These are managed by a single entrepreneur.
Because of this restriction, any technology
managed by such firms has a limited efficient
scale.5 This limits their size and also means that
they will operate as price takers, entering produc-
tion so long as revenue outweighs input costs and
compensates the entrepreneur for the opportunity
cost of their effort. We do assume, however, that
entrepreneurial opportunities for a given industry
are limited. That is, entrepreneurs can start pro-
duction in an industry by seizing an opportunity.