Quantification of the Evolution of Firm Size Distributions Due to Mergers and Acquisitions

Sandro Claudio LERA
Singapore-ETH Centre and ETH Zurich

Didier SORNETTE
ETH Zurich and Swiss Finance Institute
Quantification of the evolution of firm size distributions due to mergers and acquisitions

Sandro Claudio Lera\textsuperscript{a,b}, Didier Sornette\textsuperscript{b,c}

\textsuperscript{a}ETH Zurich, Singapore-ETH Centre, 1 CREATE Way, #06-01 CREATE Tower, 138602 Singapore
\textsuperscript{b}ETH Zurich, Department of Management, Technology, and Economics, Scheuchzerstrasse 7, 8092 Zurich, Switzerland
\textsuperscript{c}Swiss Finance Institute, c/o University of Geneva, Geneva, Switzerland

Abstract

The distribution of firm sizes is known to be heavy tailed. In order to account for this stylized fact, previous studies have focused mainly on growth through investments in a company’s own operations (internal growth). Thereby, the impact of mergers and acquisitions (M&A) on the firm size (external growth) is often not taken into consideration, notwithstanding its potential large impact. In this article, we make a first step into accounting for M&A. Specifically, we describe the effect of mergers and acquisitions on the firm size distribution in terms of an integro-differential equation. This equation is subsequently solved both analytically and numerically for various initial conditions, which allows us to account for different observations of previous empirical studies. In particular, it rationalises shortcomings of past work by quantifying that, in order to observe a significant influence of mergers and acquisitions on the firm size distribution, more extensive datasets would have been required. Our approach is very flexible and can be extended to account for other sources of external growth, thus contributing towards a holistic understanding of the distribution of firm sizes.

Keywords: firm size distribution, mergers and acquisitions, coagulation equation, econophysics

JEL: C63, D21, G34, L11

1. Introduction

In any established economy, companies of all sizes can be observed, ranging from small local businesses up to multinational corporations. The size of a company can conveniently be measured for instance in terms of yearly revenue, asset value, number of employees, and so on. It is by now a well-established fact that the distributions of firm sizes are heavy tailed: See (De Wit, 2005; Zambrano et al., 2015) and references therein for a review of the different tail distributions found for firm sizes. Knowledge of the firm size distribution is important because it has direct implications for an economy, for instance for monetary policy (Gertler and Girchrist, 1994), employment (Davis et al., 1996) or innovation (Pagano and Schivardi, 2003). Generally, a heavy tailed firm size distribution implies that a large fraction of market impact is due to a few large firms that dominate the economy (Malevergne et al., 2009; Gabaix, 2011). It is then not surprising that the study of firm size distributions has received a great deal of attention both in the theoretical and empirical literature.

While theoretical models for the heavy tail nature of firm size distributions are manifold, at the heart of most explanations is the idea of proportional growth (Gibrat (1931)’s law). Gibrat’s law states that the
growth rate of a company is independent of its absolute size, or, equivalently, that the change in size is linearly proportional to its size. In absence of any other ingredient, the resulting firm size distribution is a log-normal function. It is noteworthy that, when the variance (in the log-variable) of the log-normal function is sufficiently large, it can appear close to a power law in an intermediate range of sizes (see (Montroll and Shlesinger, 1982) and the more recent analysis of section 4.1.3 and figures 4.2 and 4.3 of (Sornette, 2004) and (Malevergne et al., 2011)). Empirically however, the distribution of firms is fatter tailed than a log-normal function, and thus requires other ingredients that just proportional growth.

In addition to proportional growth, the various mathematical models have been enriched with other realistic ingredients such as firm heterogeneity (Luttmer, 2011; Axtell et al., 2008; Wyart and Bouchaud, 2003; Takayasu et al., 2014), minimum firm size (Gabaix, 1999), births and deaths of firms (Grossman and Helpman, 1993; Steindl, 1965; Luttmer, 2007), and sudden bankruptcy (Saichev et al., 2010; Malevergne et al., 2013). Zipf’s law is then usually obtained due to subtle balance conditions between the growth rate and available external resources (Gabaix, 1999; Cabral and Mata, 2003; Malevergne et al., 2013). Finally, deviations from Gibrat’s fundamental assumption of proportional growth have also been considered (Saichev et al., 2010; Gabaix et al., 2015). Sutton (1997) gives a chronological wrap-up of the study of firm size distributions and a detailed overview on recent empirical studies is provided by Segarra and Teruel (2012).

The papers referenced above focus mainly on growth through investments in a company’s own operations (internal growth). What is often ignored is the large impact of mergers and acquisitions (external growth). For instance, over the period of 1995-1999, Schenk (2005) observes that investments in acquisitions by North American and West European firms were approximately equal to sixty per cent of gross investments in machinery and equipment and they easily outpaced those in Research and Development (R&D). Investments in acquisitions were no less than about eight times higher than business enterprise expenditures on R&D. Although it must be mentioned that the period of 1995-1999 falls in the midst of a so-called merger wave (Harford, 2005; Kummer and Steger, 2008), these facts nevertheless make the point that mergers and acquisitions (M&A) are likely to be highly relevant for the study of firm size distributions. Taking into considerations M&A (as well as its counter-part, the spin-offs) constitutes an important component of a holistic understanding of the distribution of firm sizes.

In this article, we take a first step in this direction by formalizing the effect of M&A on the firm size distribution in terms of a non-linear partial integro-differential equation. As was pointed out by Saichev et al. (2010) (chapter 10), this equation is formally equivalent to the coagulation equation from physics. Here, we will show that the coagulation equation allows to account for various empirical observations. Furthermore, our approach is very flexible and can conveniently be extended to also take into consideration spin-offs, internal growth, firm birth, bankruptcy and other growth related phenomena.

The remainder of this paper is structured as follows. In section 2, we summarize the literature that has studied the effect of mergers and acquisitions on firm size distributions. In section 3 and 4, we suggest a quantitative explanation for these observations in terms of the coagulation equation. Section 5 establishes a connection between the empirical findings and our theoretical calculations that let us account for the empirical stylized facts. Section 6 confirms further our approach by comparison with numerical solutions. Section 7 concludes and discusses possible extensions of the model.

2. The effect of mergers and acquisitions on Zipf’s law

While there is a large amount of literature examining Zipf’s law in general, there are only a few studies that investigate its relation to mergers and acquisitions. In this section, we extract the main conclusions
from those studies formulated in terms of five key observations (O1)-(O5). In the remainder of this paper, we will then set up a model to account for these observations.

(O1) Examining a sample of large American firms, Ijiri and Simon (1974) notice that smaller firms have a higher chance of being absorbed. Singh (1975) concludes the same when studying firm sizes in the UK. Aaronovitch and Sawyer (1975) confirm this finding by observing that UK firm sizes and probability of acquisitions are inversely related. We will see in the next section how this property has a very simple mathematical expression in terms of the coagulation equation.

(O2) Ijiri and Simon (1971) examine the firm size distributions in 1956 and 1957 in a sample of large American firms. Their analysis supports the proposition that firm growth due to M&As follows Zipf’s law to the same extent as without M&As. They conclude that the distribution of the 500 largest firms remained relatively unchanged. Their analysis thus supports the proposition that firm growth due to M&As would follow Gibrat’s law of proportional growth to the same extent as internal growth.

(O3) In a later study, Ijiri and Simon (1974) examine the 831 largest US industrial firms in 1969 and conclude that the firm size distribution is clearly affected by M&A. Plotting the firm size distribution in a log-log plot after the M&A had taken place, they find a concave curve that clearly deviates from a straight line expected under Zipf’s law. In section 5, we will see how the difference between Ijiri and Simon’s (1971)- and (1974)-study finds a natural explanation in terms of the coagulation equation.

(O4) Cefis et al. (2009) examine the entire population of Dutch manufacturing firms. Unlike previous studies, they examine an entire population of firms and not only large firms. They conclude that M&As do not affect the size distribution when they consider the entire population of firms. Examining only firms that are at some point involved in an M&A, a shift of the firm size distribution towards larger sizes is noticed. This shift is not uniform but affects firms of different sizes in different ways. While the number of firms in the lower tail decreased, the number of firms in the central size classes increased.

(O5) Studying UK manufacturing firms in 1957, Hannah and Kay (1977) decompose the growth rate into internal and external contribution and conclude that the observed power law coefficient is attributed to external growth. Furthermore, had it been just for internal growth, smaller firms would have grown faster than larger firms. These findings were later challenged and debated (Hart, 1979; Hannah and Kay, 1981).

Observation (O5) addresses the interplay between internal and external growth. Albeit interesting and debated, accounting for (O5) is not part of the current work and will not be further mentioned. However, as will become clear below, our model can be extended to also address (O5). In the subsequent sections, we set up a model that allows us to account for (O1)-(O4).

3. Mergers and acquisitions as a coagulation process

Let us denote by \( p(m, t) \) the distribution of firm sizes \( m \) at time \( t \). We propose to model the effect of M&A on this distribution according to the following integro-differential equation

\[
\frac{dp(m, t)}{dt} = \frac{1}{2} \int_0^m dm' A(m', m-m') p(m', t) p(m-m', t) - p(m, t) \int_0^\infty dm' A(m, m') p(m', t). \tag{3.1}
\]
We call \( A(m, m') \) the merger kernel, describing the rate at which a firm of size \( m \) merges with a firm of size \( m' \) to produce a firm of size \( m + m' \) (we neglect here any dilution of size that may occur during the merger). The first term on the right-hand side of (3.1) accounts for all the firms of size \( m' \) that merge at time \( t \) with a company of size \( m - m' \) to form a company of size \( m \). Mergers of this type are weighted with the rate \( A(m', m - m') \) and then the ‘sum’ over all such \( m', m - m' \) pairs is taken. The factor \( 1/2 \) is to avoid double counting. The second term on the right hand side decreases the concentration of firms of size \( m \) by subtracting the sum of all firms of size \( m' \) that merge at time \( t \) with a firm of size \( m' \) to form a firm of size \( m + m' \). We choose here a continuous description for mathematical convenience. Alternatively, the integrals in (3.1) can be replaced by discrete sums. This does not qualitatively alter the presented results.

As was pointed already out by Saichev et al. (2010), equation (3.1) is well-known in physics as Smulowksi’s (1916; 1917) coagulation equation. This equation has been introduced to describe the evolution of the number density of particles with mass or size \( m \) as they consolidate (coagulation process). The coagulation equation has found many applications in astrophysics, cloud physics, polymer chemistry and formation of aerosols, see (Aldous, 1999; Leyvraz, 2003; Wattis, 2006) for detailed reviews.

Figure 1: Both the merger and acquisition of firms (left side) and the coagulation of physical particles (right side) are formally described by the same mathematical equation (3.1). The fact that the two processes are described by the same mathematical equation just reflects the generic process of growth by assembling.

Considering the huge amount literature in physics and mathematics that has been devoted to equation (3.1), it should be convenient to fall back on one of the many known solutions. It turns out, however, that our merger equation is different from what is usually considered in physical systems. This is for two reasons. First, as is discussed in the next section, our merger kernel \( A(m, m') \) is a decaying function in \( m \) and \( m' \). Most physical systems assume that the coagulation kernel \( A(m, m') \) be a (homogeneous and) increasing function in \( m \) and \( m' \). Second, at time \( t = 0 \), we start from a heavy tailed firm size distribution. In physical applications, one usually starts from a thin tailed or even point-like initial distribution. This changes significantly the solution methods discussed in the literature. Only recently, power law initial distributions in combination with decaying coagulation kernels have been considered (Fournier and Lauren, 2005; Mimouni and Wattis, 2009; Niethammer et al., 2013; Niethammer and Velázquez, 2013). However, merely approximate scaling results in the limits of small and large masses are known and several issues are not yet clearly resolved. The state of the art from a mathematical perspective is well summarized by Da Costa (2015). We conclude that
there are so far no known solutions to the merger equation that we develop and solve in the next section.

4. Coagulation equation for heavy tailed distributions: an approximate solution

The solution $p(m, t)$ of (3.1) describes the firm size distribution under the influence of M&A at time $t$. We assume that our observation starts at time $t = 0$ from a heavy tailed initial distribution. Here, time $t$ is an abstract parameter that will be related to real time only in the next section. In this section, we give an analytical first order solution to (3.1).

4.1. Specification of the merger kernel and initial distribution

The coagulation equation (3.1) is not yet fully determined. We must also specify a merger kernel $A(m, m')$ and an initial distribution $p(m, t = 0)$. The kernel $A(m, m')$ describes the rate at which a firm of size $m$ merges with a firm of size $m'$ to form a firm of size $m + m'$. According to observation (O1), this must be a decaying function in $m$ and $m'$. A natural choice is thus

$$A(m, m') = e^{-am} + e^{-am'}$$

(4.1)

for some $a > 0$. Clearly, other kernels are possible but we will see below that there is no strong dependence of our results on the kernel.

As stated in the introduction, the exact shape of the firm size distributions is debated. We could thus choose either a log-normal or a Pareto as initial distribution. Since the log-normal is considerably more difficult to treat analytically, we decide to work here with a power law. Numerical results for the log-normal are found in section 6. Assuming a power law initial distribution means

$$\mu \approx \mu_0$$

(4.2)

or, equivalently (Laskin, 2003),

$$f_\mu(m) = m^{\mu-1} E_{\mu, \mu}(-m^\mu),$$

(4.3)

with the Mittag-Leffler functions $E_\mu(z)$ and $E_{\mu, \nu}$ defined by

$$E_\mu(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\mu n + 1)} \quad \text{and} \quad E_{\mu, \nu}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\mu n + \nu)}.$$

(4.4)

Repin and Saichev (2000) show that $f_\mu$ is asymptotically a power law:

$$f_\mu(m) \approx \begin{cases} \frac{m^{\mu-1}}{\Gamma(\mu)}, & m \rightarrow 0; \\ \frac{e^{-am}}{\mu \Gamma(1 - \mu)}, & m \rightarrow \infty. \end{cases}$$

(5)
What makes the FED extremely useful compared to a strict Pareto initial distribution - in fact, this is its defining property - is its simple Laplace transform,

\[ \hat{f}_\mu(k) = \int_0^\infty dm \ e^{-mk} f_\mu(m) = \frac{1}{1 + k^\mu}. \]  

(4.6)

By working with \( f_\mu \), we restrict ourselves to exponents \( \mu \in (0, 1] \). Once again, numerical results in section 6 will confirm that our results are not strongly dependent on the exact choice of \( \mu \).

4.2. First order analytical solution

There are only a few known analytical solutions to equation (3.1). All of them involve a kernel \( A \) that is an increasing function in \( m \) and \( m_0 \) and most explicit solutions consider initial distributions with finite support. Here we consider an exponentially decaying kernel and initial distribution \( f_\mu \) with a power law tail.

We derive first order analytical results that account for empirical observations already very well. The reader that is only interested in the final result can skip directly to the next section. A detailed step by step solution is available from the authors upon request.

We start by noting that kernel (4.1) is of the form

\[ A(m, m_0) = F(m) + F(m_0) \]

for some function \( F \). Hendriks et al. (1983) have shown that kernels of this type allow for a convenient simplification of the coagulation equation. We introduce \( M_0(t) \) the zeroth order moment of \( p \), which is defined by

\[ M_0(t) = \int_0^\infty dm \ p(m, t). \]  

(4.7)

Note that \( M_0(t) \) is in general not constant since it is related to the total number of firms as a function of time (see expression (5.1) below). The total number of firms can vary as a function of the combined effects of creation, bankruptcies, merger and acquisitions as well as spin-offs. Reporting the substitutions

\[ q(m, \theta) = \frac{p(m, t)}{M_0(t)}, \quad d\theta = dt \ M_0(t), \quad \theta(t = 0) = 0 \]

(4.8)

into (3.1) yields

\[ \frac{dq(m, \theta)}{d\theta} = [Fq \ast q] (m, \theta) - F(m)q(m, \theta) \]  

(4.9)

where \( \ast \) denotes the convolution operator. We interpret \( q(m, \theta) \) as a rescaled version of \( q(m, t) \) with rescaled time parameter \( \theta \). Integro-differential equations such as (4.9) are best solved in Laplace space, where the convolution turns into a multiplication. We then take the Laplace transform of (4.9) with \( F(m) = \exp(-am) \) and obtain

\[ \frac{d\hat{q}(k, \theta)}{d\theta} = [\hat{q}(k, \theta) - 1] \ \hat{q}(k + a, \theta). \]  

(4.10)

Equation (4.10) is a mixed functional differential equation (Myshkis, 2005). Analytical solutions to nonlinear equations of this type are not common. Instead, we make use of the fact that \( a \) must be small. The coagulation kernel describes the rate at which a company of size \( m \) merges with a firm of size \( m' \). Since mergers are observed also between large firms (as measured for instance in terms of revenue or number of employees), kernel (4.1) must assign non-vanishing probability also to events with \( m, m_0 \). Consequently, \( a \ll 1 \) and we can approximate \( \hat{q}(k + a) \) with a first order Taylor expansion, \( \hat{q}(k) + a\hat{q}'(k) \).
Plugging this approximation into (4.10) results in
\[
\frac{\partial \hat{q}(k, \theta)}{\partial \theta} + \alpha(1 - \hat{q}) \frac{\partial \hat{q}}{\partial k} = \hat{q}(\hat{q} - 1).
\]
This is a quasi-linear first order partial differential equation and can be solved analytically with the method of characteristics. Its solution, together with initial condition (4.6), reads \( \hat{q}(k, \theta) = \left(1 + k_0^2(k, \theta)e^\theta \right)^{-1} \) where \( k_0(k, \theta) \) is the inversion of
\[
k(k_0, \theta) = \alpha \log \left( \frac{1 + k_0^2e^\theta}{1 + k_0^2} \right) + k_0.
\]
It is not difficult to see that for the entire range of realistic values of \( k, \alpha \) and \( \theta < 1 \), the approximation \( k \approx k_0 \) holds very well (see below for an explanation why \( \theta < 1 \)). We conclude
\[
\hat{q}(k, \theta) = \frac{1}{1 + k_0^2e^\theta}.
\]
Comparing with (4.6), the inverse Laplace transform of \( \hat{q}(k, \theta) \) is just a rescaled FED,
\[
q(m, \theta) = e^{-\theta}m^{\mu-1} \mathcal{E}_{m,\mu} \left( -e^{-\theta}m^\mu \right).
\]
Finally, we have to transform back to real time \( t \). Integrating the second equation of (4.8) gives \( \theta = \int_0^t d\tau \ M_0(\tau) \). The zeroth moment \( M_0 \) is nothing but \( \hat{p}(k = 0, t) \). Taking the Laplace transform of (4.1) with \( F(m) = \exp(-\alpha m) \) and setting \( k = 0 \) thus results in the following equation for \( M_0 \),
\[
\frac{dM_0}{dt} = -\hat{p}(\alpha, t)M_0 = -\hat{q}(\alpha, \theta(t))M_0^2 = -\frac{M_0^2}{1 + \alpha e^\theta}.
\]
This last equation involves \( \theta = \int_0^t d\tau \ M_0(\tau) \). To get rid of this dependency, we self-consistently set \( \theta(t) = \theta(0)t + O(t^2) \approx t \) in (4.15). Of course, this assumption has to be justified through a posteriori sanity checks. In fact, as will become visible below, our approximation only induces a slight overestimation of \( \theta(t) \), leading at most to an understatement of our main result. Equation (4.15) is now solved with separation of variables, giving
\[
M_0(t) = \left( 1 + t + \log \left( \frac{1 + \alpha e^\theta}{1 + \alpha e^\theta} \right) \right)^{-1} = 1 - \frac{t}{1 + \alpha e^\theta} + O(t^2),
\]
and then,
\[
\theta = \int_0^t d\tau \ M_0(\tau) = t - \frac{t^2}{2(1 + \alpha e^\theta)} + O(t^3).
\]
In conclusion, the first order solution to (3.1) is given by (4.14) with \( \theta(t) \) determined via (4.17). This solution is depicted in figure 2 for \( \mu = 0.5 \) at different times \( t \). We notice that, for small times \( t \), only mild deviations from the initial distribution are observed. This is in line with observations (O2)-(O4), as we show in the next section.
Figure 2: The first order solution of the coagulation equation (3.1) with exponential kernel (4.1) and initial distribution (4.3) is given by (4.14) with \( \theta(t) \) determined via (4.17). Here, we show the corresponding survival function (or complementary cumulative distribution function) \( \text{ccdf} \) for \( \mu = 0.5 \) and \( \alpha = 0.01 \). The dependence on \( \alpha \) is extremely weak, which is why we do not show different choices. The times \( t \) are chosen in correspondence to the dataset considered by Cefis et al. (2009).

5. Connecting the coagulation equation to empirical data

The first order solution to (3.1) was derived in the previous section as (4.14) with \( \theta(t) \) determined by (4.17). So far, we have treated time \( t \) as an abstract parameter. The connection to real time is established by noting that the coagulation kernel \( A(m,m') \) is actually a rate.

Denote the number of firms at time \( t \) by \( N(t) \). It holds that

\[
N(t) = N_0 \int_0^\infty dm \ p(m,t) = N_0 M_0(t)
\]

(5.1)

with \( N_0 \equiv N(t=0) \). Then, the total number of firms that have ‘vanished’ through M&A since time \( t = 0 \) is given by

\[
\Delta N(t) = N_0 - N(t) = N_0 (1 - M_0(t))
\]

(5.2)

A priori, \( \Delta N < 0 \) is also possible, meaning that newly born firms and spin offs outweigh the mergers and increase the total number of firms in an economy. The pure coagulation equation cannot account for such an observation. Here, we do not need to consider this case since (O2)-(O4) all report \( \Delta N > 0 \). However, it is, at least in principle, straightforward to extend equation (3.1) to the inclusion of spin-offs and firm births (cf. equation (10.5) in (Saichev et al., 2010)).
Knowing $M_0(t)$, equation (5.2) allows us to solve for time $t$ as a function of observed $\Delta N$. In our case, we deduce with (4.16) that

$$t = (1 + a^\omega) \frac{\Delta N(t)}{N_0} \approx \frac{\Delta N(t)}{N_0}. \quad (5.3)$$

In this last approximation, we have used that $a \ll 1$. This is justified for the empirical studies below, for which mergers between large companies are reported. We have thus systematically shown that there is negligible dependence on $a$, meaning that the kernel $A$ can be approximated to first order as constant for small times. This result can be made intuitive when noticing that, even for a constant kernel, the far right tail of the distribution is only affected for long times (see also numerical solution in section 6). For small times, it is mostly the lower part of the distribution that is affected by coagulation. For small $m$, $\exp(-am) \approx 1$ is a good approximation.

Let us now turn to observations (O2)-(O4). The most thorough study was conducted by Cefis et al. (2009). They study a comprehensive data set of roughly 60’000 Dutch manufacturing firms including firm entries, exits, spin-offs, mergers and acquisitions. Measuring firm size in number of employees and accounting exactly for all relevant events, Cefis et al. estimate both the firm size distribution in the beginning and in the end of the year 1997. The starting distribution is found to be best fitted by a log-normal. When plotting the starting and final distribution in one figure, it is seen that the starting and final distribution largely overlap, thus demonstrating that the firm size distribution seems to be unaffected by M&A (cf. figure 2 there). Since 1997 was a year of high M&A activity of Dutch firms, this observation seems puzzling. While Cefis et al. suggest a number of reasons, this observation finds a natural explanation within our framework. During 1997, 3’899 firms were involved in activities related to M&A and their spin-offs and divestitures. At the end of the year, 2’564 remained, which suggests that M&As, spin-offs and divestitures have decreased the numbers of firms active in the manufacturing sector by $\Delta N = 1’335$. With $N_0 = 57’329$, we predict from (5.3) a temporal evolution of merely $t \approx 0.02$. As is visible in figure 2, the $p(m, t = 0.02)$ curve is hardly distinguishable from the $p(m, t = 0)$ curve. The parameter $t$ is related to real time ($\approx 365$ days) via a rate, namely the rate of M&A events. The contribution of our analysis is thus to quantify that the M&A events over the full year of 1997 amount in aggregate to a tiny perturbation to the initial firm distribution.

In a next step, in order to emphasize the possible effects of M&As, Cefis et al. restrict their dataset to the firms that are directly involved in a M&A event during the year 1997. The result is a final distribution that is slightly above the starting distribution, showing that the M&As did have an effect on the distribution (cf. figure 4 there). Again, their result is in good correspondence with our theory. Having $N_0 = 3’899$ firms in the beginning that will undergo a M&A event and 2’564 of these firms at the end of 1997, we calculate $\Delta N = 1’335$. According to (5.3) this corresponds to $t = 0.35$. In our model, this results in a distribution that lies also slightly above the initial distribution, as shown in figure 2.

In a very similar fashion, we can explain the observations of Ijiri and Simon (1971, 1974), at least semi-quantitatively. The major difference between their datasets and the one considered by Cefis et al. (2009) is that Ijiri and Simon (1971, 1974) consider only the tail of the distribution, and do not have access to a complete dataset. In their first study, Ijiri and Simon (1971) consider the $N_0 = 500$ largest industrial firms (in terms of sales volume) and the number of firms that they have bought. They do not find any deviation from Zipf’s law through M&A. In our model, this is very natural. Only a total of $\Delta N = 19$ net mergers are observed, resulting in a small time value of $t \approx 0.038$. It is then clear from our calculations that no deviation from Zipf’s law is to be expected.

In a follow-up paper, Ijiri and Simon (1974) consider a much larger data set and come to the conclusion that M&A does change the firm size distribution by shifting it slightly above from what we would expect...
without mergers. They do not have a complete dataset, so they provide an estimate. They estimate that the 831 largest industrial firms in 1969 would actually be 1’002 companies, if it had not been for a large amount of mergers in the preceding decade. Based on their estimate, we determine \( N_0 = 1'002 \) and \( \Delta N = 1'002 - 831 = 171 \), and thus \( t = 0.17 \). The effect that they observe in the data is a tiny shift of the actual curve above the hypothetical curve without mergers (cf. figure 1 there). This is again in agreement with what we would expect from our model with \( t = 0.17 \).

In conclusion, our model accounts for observations (O2)-(O4). Furthermore, we can see that, in order to observe significant deviations from the initial distribution, a large number of mergers and acquisitions are needed, of the order of the number of firms in the considered dataset. This can be traced back to the fact that the heavy tail of the distribution is asymptotically unaffected by the coagulation process with decaying or constant merger kernel. If the firm size distribution was light tailed, deviations in the right end of the distribution would become apparent at an earlier stage, i.e. with fewer M&A observations (see also figure 3).

6. Generality of our results

In the previous sections, we have presented a first order analytical solution to the coagulation equation. Thereby, several assumptions and approximations have been made for mathematical convenience. In this section, we show that our results apply more broadly and are not dependent on the assumptions made. Concretely, we use Lee (2000)’s second-order Runge-Kutta method with variable time step to solve the coagulation equation (3.1) numerically.

Observation (O1) states that smaller firms have a higher chance of being absorbed. We have thus assumed an exponentially decaying merger kernel. While this is a natural choice, other shapes are possible. For instance, we could imagine a Pareto kernel \( A(m, m') = 1/m^\beta + 1/(m')^\beta \) for some \( \beta > 0 \). In figure 3, we show the numerical solution to the coagulation equation for both the exponentially decaying kernel (with different choices of \( \alpha \)) and for the constant kernel \( A(m, m') = 1 \). We see that, for \( t < 1 \), the difference between the two kernels is small. Since the Pareto kernel is in between these two extremes of exponential decay and a constant, we conclude that basing our calculations on the exponential kernel is justified.

To simplify the Laplace transform, we have worked with the fractional exponential distribution with asymptotic power law tail. However, the literature suggests that the lognormal is often also a good approximation of firm size distribution. Empirically, as already mentioned, it can be difficult to distinguish the two. To understand this, note that one can write the log-normal density as

\[
p(m, t = 0) = \frac{1}{\sqrt{2\pi\sigma^2}} m \exp\left( -\frac{\log^2(m/m_0)}{2\sigma^2} \right) = \frac{1}{\sqrt{2\pi\sigma^2}} m_0 (m/m_0)^{-1-\mu(m)} \propto m^{-1-\mu(m)}
\]  

(6.1)

with \( \mu(m) = \frac{1}{2\sigma^2} \log \left( \frac{m}{m_0} \right) \). The second equality in (6.1) follows because \( e^{(\log m)^2} = m^{\log m} \) (cf. section 4.1.3 in Sornette, 2004). Since \( \mu(m) \) is a slowly varying function of \( m \), the log-normal distribution can be mistaken for an apparent power law with an exponent \( \mu \) that is gradually changing. Despite this apparent similarity, the log-normal is very different from a Pareto. According to the Pickands-Balkema-de Haan theorem of extreme value theory (Embrechts et al., 2013), the log-normal law belongs to the domain of attraction of the Gumbel distribution for the distribution of maxima, and the Gumbel distribution has an exponential tail. For the log-normal, all moments exist and we would expect that the coagulation equation shows qualitatively different behavior for those two initial distributions (power law versus log-normal). Indeed, Menon and Pego (2004, 2006) show that (for some specific kernels) the solution of the coagulation equation depends in a subtle way on the number of finite moments of the initial distribution. Nevertheless, as is confirmed in
figure 3, these differences are not visible for small times over the considered range of firm sizes. All our theoretical statements seem to apply equally well for other heavy-tailed initial distributions.

7. Conclusions & outlook

We have examined the effect of mergers and acquisitions on time evolution of firm size distributions. Specifically, we have described the effect of M&A in terms of a partial integro-differential equation, also known in the literature as the coagulation equation. This approach allowed us to account for the inverse relation of firm size and the probability of being acquired. While the resulting coagulation equation cannot be solved analytically in closed form, we have derived a first order solution that can explain empirical observations appearing counterintuitive at first sight, namely that a lot more observations of M&A events are required in order to see a significant deviation from a heavy tailed firm size distribution. Precise numerical solutions have confirmed the validity of our approximations. Hence, we conclude that the coagulation equation is an adequate tool to capture the dynamics of merger and acquisitions on the firm size distribution.

We have made two assumptions that could be relaxed. First, we considered the situation where the considered initial distributions are already heavy tailed equilibrium distributions. Implicitly, this assumes that the effect of mergers and acquisitions is negligible at an early stage of an economy. This assumption is justified insofar as M&As have become especially prominent in the second half of the last century (cf. figure 1 in (Kummer and Steger, 2008)), whereas the observations of heavy tailed firm size distributions date back to Gibrat (1931). However, it may be that the asymptotic long time behavior of the firm size distribution is strongly dependent on the initial distribution, a problem that is in full generality still under investigation, see for instance Da Costa (2015). It is then interesting to ask whether mergers and acquisitions can significantly impact the firm size landscape of a developing economy.

Second, we have only considered M&As and neglected other firm size dynamics such as spin-offs, firm creation, firm exits or sudden bankruptcy. Adding spin-off dynamics to the coagulation equation leads to the so-called coagulation-fragmentation equation. In contrast to pure coagulation, here equilibrium states are possible by offsetting the M&As with spin-offs. An interesting question is then whether such a balance condition is met in real economies, or whether external growth dynamics acts as a driver towards out-of-equilibrium dynamics. Investigation of such questions is part of future research.

Acknowledgment

The research performed by one of us (SCL), which is reported in this article, was conducted at the Future Resilient Systems at the Singapore-ETH Centre (SEC). The SEC was established as a collaboration between ETH Zurich and the National Research Foundation (NRF) Singapore (FI 370074011) under the auspices of the NRF’s Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

Aaronovitch, S. and Sawyer, M. (1975). Big business : theoretical and empirical aspects of concentration and mergers in the United Kingdom. London: Macmillan.
Aldous, D. J. (1999). Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. Bernoulli, 5(1):3–48.
Axtell, R., Gallegati, M., and Palestrini, A. (2008). Common components in firms’ growth and the sectors scaling puzzle. Economics Bulletin, 12:1–8.
Cabral, L. and Mata, J. (2003). On the evolution of the firm size distribution: Facts and theory. American economic review, pages 1075–1090.
Niethammer, B. and Velázquez, J. J. L. (2013). Self-similar solutions with fat tails for Smoluchowski’s coagulation equation with locally bounded kernels. *Communications in Mathematical Physics*, 318(2):505–532.

Pagano, P. and Schivardi, F. (2003). Firm size distribution and growth. *Scandinavian Journal of Economics*, 105(2):255–274.

Repin, O. N. and Saichev, A. I. (2000). Fractional poisson law. *Radiophysics and Quantum Electronics*, 43(9):738–741.

Saichev, A., Malevergne, Y., and Sornette, D. (2010). *Theory of Zipf’s Law and Beyond*. Springer.

Saichev, A. and Zaslavsky, G. (1997). Fractional kinetic equations: solutions and applications. *Chaos*, 7(4):753–764.

Schenk, H. (2005). *Multidisciplinary Economics: The Birth of a New Economics Faculty in the Netherlands*, chapter Organisational economics in an age of restructuring, or: how corporate strategies can harm your economy, pages 333–365. Springer US.

Segarra, A. and Teruel, M. (2012). An appraisal of firm size distribution: Does sample size matter? *Journal of Economic Behavior & Organization*, 82(1):314–328.

Singh, A. (1975). Take-overs, economic natural selections, and the theory of the firm: Evidence from the postwar united kingdom experience. *The Economic Journal*, 85(339):497.

Smoluchowski, M. (1916). Drei vorträge ber diffusion, brownsche molekularbewegung und koagulation von kolloidteilchen. *Zeitschrift für Physik*, 17(1):557–571, 587–599.

Smoluchowski, M. (1917). Versuch einer mathematischen theorie der koagulationskinetik kolloider lösungen. *Zeitschrift für physikalische Chemie*, 92(2):129–168.

Sornette, D. (2004). *Critical Phenomena in Natural Sciences*. Springer.

Steindl, J. (1965). *Random processes and the growth of firms: A study of the Pareto law*. Griffin London.

Sutton, J. (1997). Gibrat’s legacy. *Journal of economic literature*, 35(1):40–59.

Takayasu, M., Watanabe, H., and Takayasu, H. (2014). Generalised central limit theorems for growth rate distribution of complex systems. *Journal of Statistical Physics*, 155(1):47–71.

Wattis, J. A. D. (2006). An introduction to mathematical models of coagulation–fragmentation processes: a discrete deterministic mean-field approach. *Physica D*, 222(1):1–20.

Wyart, M. and Bouchaud, J.-P. (2003). Statistical models for company growth. *Physica A*, 326(1):241–255.

Zambrano, E., Hernando, A., Bariviera, A. F., Hernando, R., and Plastino, A. (2015). Thermodynamics of firms’ growth. *Journal of The Royal Society Interface*, 12(112):20150789.

13
Figure 3: Numerical solution, for the complementary cumulative distribution functions (ccdf) of firm sizes, of the coagulation equation (3.1) at time $t = 1$ for different initial distributions. The merger kernel is the exponential (4.1) with rate $\alpha$. We show the solutions for different values of $\alpha$. In particular, $\alpha = 0$ corresponds to the constant kernel. Figure (a) corresponds to the initial distribution being a fractional exponential distribution with asymptotic Pareto tail with exponent $\mu = 0.5$. We can see that, even for the very large value $\alpha = 0.1$, deviations from the constant kernel solution are small, thus justifying our previous approximations. Similarly, figure (b) corresponds to the initial distribution being the heavy-tailed Cauchy distribution with shape parameter $\gamma = 5$ and asymptotic Pareto tail $\mu = 1$ (Zipf’s law). Figure (c) and (d) correspond to the initial distribution being a log-normal law with different values of $\sigma$ (see definition (6.1)). This confirms that, over a finite size range and for times up to $t = 1$, the behavior of the firm distribution as a function of time for the log-normal initial distribution is similar to the case of the power law distributions. Finally, figure (e) and (f) exemplify that, for light-tailed initial distributions, deviations are more pronounced.
Quantification of the evolution of firm size distributions due to mergers and acquisitions:

Step by step solution of the coagulation equation

Sandro Claudio Lera\(^1,2\) and Didier Sornette\(^12,3\)

\(^1\)Singapore-ETH Centre, ETH Zurich, 1 CREATE Way, \#06-01 CREATE Tower, 138602 Singapore

\(^2\)Department of Management, Technology, and Economics, ETH Zurich, Scheuchzerstrasse 7, 8092 Zurich, Switzerland

\(^3\)Swiss Finance Institute, c/o University of Geneva, Geneva, Switzerland

We solve the coagulation equation

\[
\frac{dp(m,t)}{dt} = \frac{1}{2} \int_0^m dm' A(m',m-m')p(m',t)p(m-m',t) - p(m,t) \int_0^\infty dm' A(m,m')p(m',t). \tag{1}
\]

with coagulation kernel

\[
A(m,m') = e^{-\alpha m} + e^{-\alpha m'} \tag{2}
\]

and initial distribution \(p(m,t=0) = f_\mu(m), \mu \in (0,1)\). See main paper for the definition of \(f_\mu\), motivations and other details.

The kernel (2) is of the form

\[
A(m,m') = F(m) + F(m') \tag{3}
\]

with \(F(m) = \exp(-\alpha m)\). Plugging (3) into the coagulation equation (1) gives

\[
\frac{dp(m,t)}{dt} = \int_0^m dm' F(m')p(m')p(m-m') - F(m)p(m)M_0(t) - p(m) \int_0^\infty dm' F(m')p(m')
= [Fp*p](m) - F(m)p(m)M_0(t) - p(m) \int_0^\infty dm' F(m')p(m') \tag{4}
\]

where we have defined \(M_0(t) \equiv \int_0^\infty dm \ p(m,t)\) and introduced the convolution operator ‘*’.

Now, we make the substitution

\[
q(m,\theta) = \frac{p(m,t)}{M_0(t)}, \quad d\theta = dt \ M_0(t), \tag{5}
\]

\(^{*}\)slera@ethz.ch

\(^{†}\)dsornette@ethz.ch
implying
\[
\frac{d}{dt} [p(m, t)] = \frac{d}{dt} [M_0(t)q(m, \theta(t))] = q \frac{dM_0(t)}{dt} + M_0 \frac{dq}{dt}
\]
\[
= q \frac{dM_0(t)}{dt} + M_0 \frac{d\theta}{dt} \frac{dq}{d\theta}
\]
\[
= q \frac{dM_0(t)}{dt} + M_0^2 \frac{dq}{d\theta}.
\]  
(6a)
(6b)
(6c)
(6d)

Plugging this substitution into (4) gives
\[
q \frac{dM_0}{dt} + M_0^2 \frac{dq}{d\theta} = M_0^2 [Fq * q] - F M_0^2 q - M_0^2 \int_0^\infty \int \tilde{m} F(m') q(m').
\]  
(7)

Matters simplify once we notice that the first term on the lhs of (7) cancels with the third term on the rhs of (7). Indeed,
\[
q \frac{dM_0}{dt} = q \int_0^\infty \int \tilde{m} \frac{dp(m', t)}{dt}
\]
\[
= q \int_0^\infty \int \tilde{m} \left[ M_0^2 [Fq * q] - F M_0^2 q - M_0^2 \int_0^\infty \int \tilde{m} F(m') q(m') \right].
\]  
(8a)
(8b)

We note that the first term of (8b) cancels with the third term of (8b) since
\[
\int_0^\infty \int \tilde{m} [Fq * q](m') = \int_0^\infty \int \tilde{m} [Fq](n) q(m' - n)
\]
\[
= \int_0^\infty \int \tilde{m} [Fq](n) q(m' - n)
\]
\[
= \int_0^\infty \int \tilde{m} q(m' - n)
\]
\[
= \int_0^\infty \int \tilde{m} q(m)
\]  
(9a)
(9b)
(9c)
(9d)

where we have used in (9b) that \(q(m) = 0\) for \(m < 0\) by assumption and Fubini’s theorem in (9c). Hence,
\[
q \frac{dM_0}{dt} = -M_0^2 q \int_0^\infty \int \tilde{m} F(m') q(m')
\]  
(10)

which confirms that the first term in (7) cancels with the third term on the rhs of (7). In conclusion, substitution (5) transforms the coagulation equation (1) into the simpler equation
\[
\frac{dq(m, \theta)}{d\theta} = [Fq * q](m) - F(m) q(m).
\]  
(11)
Taking the Laplace transform of (11), we find
\[
\frac{d\hat{q}(k, \theta)}{d\theta} = [\hat{q}(k, \theta) - 1] \hat{F}(k, \theta).
\] (12)

Equation (12) holds for general coagulation kernels of the form (3), in particular for (2). The function \( F(m) = \exp(-\alpha m) \) induces merely a shift by \( \alpha \) in the Laplace transform, so that (12) becomes
\[
\frac{d\hat{q}(k, \theta)}{d\theta} = [\hat{q}(k, \theta) - 1] \hat{q}(k + \alpha, \theta).
\] (13)

Since \( \alpha \) is small (see main paper), we can approximate \( \hat{q}(k + \alpha) \approx \hat{q}(k) + \alpha \hat{q}'(k) \) and plug this into (13), giving
\[
\frac{\partial \hat{q}(k, \theta)}{\partial \theta} + \alpha (1 - \hat{q}) \frac{\partial \hat{q}}{\partial k} = \hat{q}(\hat{q} - 1).
\] (14)

This quasilinear first-order PDE can be solved with the method of characteristics (see appendix for a summary of the method). We look for a parametrization of the form \( k(s), \theta(s), \hat{q}(k(s), \theta(s)) \), resulting in the three ODEs
\[
\frac{d\theta}{ds} = 1,
\frac{dk}{ds} = \alpha (1 - \hat{q}),
\frac{d\hat{q}}{ds} = \hat{q}(\hat{q} - 1),
\] (15, 16, 17)

with initial conditions
\[
\theta(s = 0) = 0, \quad k(s = 0) = k_0, \quad \hat{q}(s = 0) = g(k_0)
\] (18)

and \( g(\cdot) \) the initial condition. Here, \( k_0 \) is only formally defined, as a function that relates \( k \) to \( s \) (see appendix). Equation (15) is trivial, giving \( \theta = s \). We can thus drop the artificial parameter \( s \) and write \( \theta \) instead. Next, we solve equation (17) by separation of variables:
\[
\frac{d\hat{q}}{\hat{q}(\hat{q} - 1)} = d\theta
\] (19a)

\[
\log \left( \frac{1 - \hat{q}}{\hat{q}} \right) = \theta + \text{const}
\] (19b)

\[
\hat{q}(\theta) = \frac{1}{1 + \text{const} \cdot e^{\theta}}
\] (19c)

According to (18), we require that at \( \theta = 0 \), (19c) satisfies
\[
g(k_0) = \frac{1}{1 + \text{const}}
\] (20)

where \( g \) is our initial distribution. In the main paper, we argue that \( p(m, t = 0) \) is the fractional exponential \( f_\mu(m) \) with Laplace transform
\[
g(k_0) = \mathcal{F}_\mu(k_0) = \frac{1}{1 + k_0^\mu}.
\] (21)

Comparing with (20) then fixes the integration constant to \( k_0^\mu \) so that the final solution to equation (17) reads
\[
\hat{q}(k_0, \theta) = \frac{1}{1 + k_0^\mu e^\theta}.
\] (22)
Of course, this solution holds also for general initial distribution $g$. Solving (20) for ‘const’ just gives

$$\tilde{q}(k_0, \theta) = \frac{1}{1 + \left(\frac{1}{g(k_0)} - 1\right) e^\theta}$$

(23)

which recovers (22) as the special case $g = f_\mu$.

We plug (22) into (16),

$$\frac{dk(\theta)}{d\theta} = \alpha \left(1 - \frac{1}{1 + k_0^\mu e^\theta}\right).$$

(24)

Because the rhs is independent of $k$, the solution of (24) is merely an integration problem:

$$k = \alpha \log \left(1 + k_0^\mu e^\theta\right) + \text{const.}$$

(25)

The initial value (18) determines the integration constant, and we end up with

$$k(k_0, \theta) = \alpha \log \left(1 + k_0^\mu e^\theta\right) + k_0.$$  

(26)

As is explained in the appendix, the final result to (14) is therefore

$$\tilde{q}(k, \theta) = g(k_0(k, \theta), \theta) = \frac{1}{1 + k_0^\mu(k, \theta)e^\theta},$$

(27)

where $k_0(k, \theta)$ is the inversion of (26). Equation (26) can only be inverted numerically. However, note that for some fixed $m$, the relevant contribution of the Laplace transform $\tilde{q}(k)$ to $q(m)$ comes from $k$ values that are in the range $k \sim O\left(\frac{1}{m}\right)$. It is then straightforward to see from (27) that for realistic values of $\alpha$ and $\theta$ (see main paper for what this means), the approximation $k \approx k_0$ holds very well for the entire range of relevant $k$ values (see figure 1). In that case, (27) simplifies to

$$\tilde{q}(k, \theta) = \frac{1}{1 + k^\mu e^\theta},$$

(28)

Figure 1: We show the functional relation (26) between $k$ and $k_0$ for different values of $\alpha$ and $\theta$. We can see that the approximation $k \approx k_0$ holds very well over the entire range of relevant $k$ values.
We notice that (28) is almost exactly of the form (21), only with a prefactor $\exp(\theta)$. Because the inverse Laplace transform is an integration problem, we can relate (28) to the inverse Laplace transform of (21) simply by substitution of a constant. The result reads

$$q(m, \theta) = e^{-\theta} m^{\mu-1} E_{\mu,\mu}\left(-e^{-\theta} m^{\mu}\right). \quad \text{(29)}$$

This finishes the calculation of $q(m, \theta)$. Finally, we have to transform back to $p(m, t)$. First, we integrate the second equation in (5). Using that $\hat{\theta}(t = 0) = 0$, this gives

$$\theta = \int_0^t d\tau \, M_0(\tau). \quad \text{(30)}$$

We note that $M_0(t) = \hat{\rho}(0,t)$. Hence, by taking the Laplace transform of the coagulation equation (1) and then setting $k = 0$ we arrive at an equation for $M_0$. Looking at (4), the Laplace transform of (1) is straight forward,

$$\frac{d\hat{\rho}(k, t)}{dt} = \hat{F} p(k) \hat{\rho}(k) - M_0 \hat{F} p(k) - \hat{\rho}(k) \hat{F} p(0). \quad \text{(31)}$$

With $F(m) = \exp(-m\alpha)$, this translates into

$$\frac{d\hat{\rho}(k, t)}{dt} = \hat{\rho}(k + \alpha, t) \hat{\rho}(k) - M_0 \hat{\rho}(k + \alpha, t) - \hat{\rho}(k) \hat{\rho}(\alpha, t) \quad \text{(32)}$$

and setting $k = 0$ gives

$$\frac{dM_0}{dt} = -\hat{\rho}(\alpha, t) M_0 = -\hat{\rho}(\alpha, \theta(t)) M_0^2 = -\frac{M_0^2}{1 + k e^\theta}. \quad \text{(33)}$$

This last equation involves $\theta$ explicitly via (28). To get rid of this dependency, we self-consistently approximate $\theta(t) = \theta'(0)t + \mathcal{O}(t^2) \approx t$. This approximation is justified a posteriori in the main paper. Plugging $\theta \approx t$ into (33) yields

$$\frac{dM_0}{dt} = -\frac{M_0^2}{1 + \alpha e^t}, \quad M_0(0) = 1. \quad \text{(34)}$$

This Riccati equation is straightforward to solve with separation of variables. We obtain

$$M_0(t) = \left(1 + t + \log\left(\frac{1 + \alpha^2}{1 + \alpha e^t}\right)\right)^{-1} = 1 - \frac{t}{1 + \alpha e^t} + \mathcal{O}(t^2), \quad \text{(35)}$$

and then, via (30),

$$\theta = t - \frac{t^2}{2(1 + \alpha^2)} + \mathcal{O}(t^3). \quad \text{(36)}$$

In conclusion, the first order solution to (1) is given by (29) with $\theta(t)$ determined via (36).

**Appendix: Method of characteristics**

The method of characteristics enables us to solve a large class of first order PDEs. This method is well known and detailed treatments can be found in standard textbooks. Here, we summarize briefly the idea of the method and how it is applied to our problem.

Consider the quasilinear 1 + 1-dimensional PDE

$$a(x, t, f) \frac{\partial f}{\partial x} + b(x, t, f) \frac{\partial f}{\partial t} = c(x, t, f) \quad \text{(A1)}$$

and then, via (30),

$$\theta = t - \frac{t^2}{2(1 + \alpha^2)} + \mathcal{O}(t^3). \quad \text{(36)}$$

In conclusion, the first order solution to (1) is given by (29) with $\theta(t)$ determined via (36).
with initial condition \( f(x, t = 0) = g(x) \). This equation is called ‘quasilinear’ because the functions \( a, b \) and \( c \) may also depend on the function \( f = f(x, t) \) itself. The equation is thus not really linear, but it is linear in its derivatives. Equation (14) is exactly of this form. Suppose that a solution \( f(x, t) \) is known, and consider the surface graph

\[
\begin{pmatrix}
x \\
t \\
f(x, t)
\end{pmatrix} \in \mathbb{R}^3. \tag{A2}
\]

A normal vector to this surface is given by

\[
\begin{pmatrix}
\frac{\partial f}{\partial x} \\
\frac{\partial f}{\partial t} \\
-1
\end{pmatrix}. \tag{A3}
\]

Consequently, equation (A1) is equivalent to the geometrical statement that the vector field

\[
\begin{pmatrix}
a(x, t, f) \\
b(x, t, f) \\
c(x, t, f)
\end{pmatrix} \tag{A4}
\]

is tangent to the surface (A2), because the dot-product of (A4) with the surface normal (A3) is zero. We conclude that the graph of the solution must be a union of integral curves of the vector field (A4). These integral curves are called the characteristic curves of the original PDE (A1). We have reduced the problem of solving (A1) to finding and labeling these characteristic curves.

We can label a specific characteristic curve uniquely by introducing the parametrization \( x = x(x_0; s), \ t = t(x_0; s) \) and \( f = f(x_0; s) = f(x(s), t(s)) \). Here, \( x_0 \) labels the start of the initial curve at \( t = 0 \). Thus, for fixed \( x_0 \), the parameter \( s \) describes a specific curve. Changing \( x_0 \) labels the starting points of different curves. Together, the pair \((x_0, s)\) thus characterizes the entire solution surface graph (A2). For notational simplicity, we do not always write explicitly the \( x_0 \) dependence in the following.

Any tangent vector along any curve \((x(s), t(s), f(s))\) in \( \mathbb{R}^3 \) with parametrization \( s \) is just

\[
\begin{pmatrix}
dx/ds \\
dt/ds \\
df/ds
\end{pmatrix}. \tag{A5}
\]

However, we do not consider any curve \((x(s), t(s), f(s))\), but only the parametrization along a characteristic curve, such that the tangent vector is equal to (A4). This results in the following three ordinary differential equations:

\[
\begin{pmatrix}
dx/ds \\
dt/ds \\
df/ds
\end{pmatrix} = \begin{pmatrix}
a(x, t, f) \\
b(x, t, f) \\
c(x, t, f)
\end{pmatrix} \tag{A6}
\]

We have now effectively reduced the problem of solving a PDE to solving three (in general coupled) ODEs. Note that the starting point of our parametrization, \( s = 0 \), is arbitrary. This allows for a convenient simplification. Let us assume that \( a(x, t, f) \equiv 1 \). This assumption is almost without loss of generality, since often \( a \) is a constant or a non-zero function of \( x \) and \( t \). In that case, we can rescale the entire equation by \( 1/a \) and our assumption is without loss of generality. Of course, the function \( a = a(x, t, f) \) could be such that it is equal to zero at certain points. In this case our rescaling argument does not work, and we have to choose the
parametrization differently. This changes nothing about the general procedure but would make the following discussion notationally and computationally more complicated.

Having set $a = 1$, the first equation in (A6) reads just $dt/ds = 1$, i.e. $t = s + C$. Since the starting point is arbitrary, we may set $C \equiv 0$ so that $t = s$ is our parametrization variable. Thus, $x(s) = x(t)$. Define $x(t = 0) \equiv x_0$. This $x_0$ is somewhat arbitrary and depends only on our choice of initial time $t = 0$. However, in the end result we will get rid of it via inversion. Think of this as follows. We can parametrize the solution surface either as $(x, t, f(x, t))$ or as $(x_0, s, f(x_0, s))$. It is convenient to find the solution surface in terms of $(x_0, s)$, but eventually we invert the bijection $(x_0, s) \leftrightarrow (x, t)$ to get back to our original coordinates.

Finally, we note that $f(x(s), t(s))$ is constant along the characteristics since

$$\frac{d}{ds} f(x(s), t(s)) = \frac{\partial f}{\partial s} + \frac{\partial f}{\partial x} \frac{dx}{ds} + \frac{\partial f}{\partial t} \frac{dt}{ds} = c + a \frac{\partial f}{\partial x} + b \frac{\partial f}{\partial t} = 0. \quad (A7)$$

Thus, $f(x(s), s) = f(x(t), t) = f(x(t = 0), t = 0) = f(x_0, 0) \equiv h(x_0)$ where $h$ denotes an arbitrary function. This yields the following three initial values for the three ODEs (A6),

$$t(s = 0) = 0, \quad x(s = 0) = x_0, \quad f(s = 0) = h(x_0) \quad (A8)$$

and enables us to solve the three (A6) uniquely. We are then mainly interested in the solution $x(t) = x(x_0, t)$. We can invert this relation for $x_0$, i.e. we have $x_0 = x_0(x; t)$. We plug this relation into $h(x_0)$ and have found the general solution $f(x, t) = h(x_0) = h(x, t)$. Here, $h$ is still an arbitrary function. The specific solution is found from the initial condition, by associating $h \equiv g$. 

7
