UCD GEARY INSTITUTE
DISCUSSION PAPER SERIES

Size Metrics and Dynamics of Firms Expansion in the European Pharmaceutical Industry

Franco Mariuzzo
Geary Institute,
University College Dublin

Xiaoheng Zhang
ESRI and Trinity College Dublin

Geary WP2010/37
September 6, 2010

Size Metrics and Dynamics of Firms Expansion in the European Pharmaceutical Industry*

Franco Mariuzzo†
Geary Institute
University College Dublin
Dublin, Ireland.

Xiaoheng Zhang‡
ESRI & Trinity College Dublin
Dublin, Ireland.

September 6, 2010

Abstract

We generalize the empirical growth-of-firms literature by linking a mixture of discrete and continuous alternative metrics of size via a Copula approach. We look at the result of the fitted Copula and justify the metric we base our analysis upon. We then employ the *Amadeus* dataset and investigate the growth dynamics of the European pharmaceutical industry in the *Single Market Programme* era, 1990-2004. Relying on a set of dynamic panel Probit methods that deal with unobserved heterogeneity and initial conditions, we analyze how our units of investigation, multinationals, capture opportunities over time. We find strong evidence of state dependence and mean reversion, as predicated by the theory of maturation - firms face a period of rapid growth, followed by a slow down, or even a stop, in growth. We finish off our exercise by conditioning the fitted Copula on the predicted selected measure of size and simulating the remaining measures. Our methodology has a fair explanatory power.

Keywords: Copula, dynamic nonlinear panel data models with unobserved heterogeneity, firm-growth, pharmaceutical industry, single market programme, Sutton's lower bound.

JEL classifiers: C10, C11, C23, L11, L65.

*This paper was presented at the Irish Economic Association conference, Westport, Ireland (2008), the Small Open Economies in a Globalized World conference, Waterloo, Canada (2008), and the European Economic Association, Barcelona, Spain (2009). We thank Wiji Arulampalam, Silvi Berger, Mike Harrison, Ruud Koning and Paul Walsh for their precious comments.

†Tel. +353-1-716-4613. Fax. +353-1-716-1108. E-mail: franco.mariuzzo@ucd.ie.

‡Email: Kevin.Zhang@esri.ie. A grant from the Institute for International Integration Studies, Trinity College Dublin, is gratefully acknowledged.

1 Introduction

The literature on firm growth has repeatedly relied on a given measure (metric) of size, often being one of the (approximately) continuous variables: employment, assets or sales. To our knowledge, no empirical paper has investigated the possibility that, in principle, there may be systematic differences among the alternative metrics of size available, with the implicit risk that the resulting analysis would suffer from being tailored to the adopted metric. This paper is the first of its kind that conditions the analysis of firm growth on a deep investigation of the relation between alternative measures of size. Given that we are assembling very different measures of size, some discrete and others (approximately) continuous, the Copula is the appropriate approach to study how the alternatives associate to one another. We wish to shed light on the concordance of the variables, i.e. how similar, or different, the alternative metrics might be.[1] If we were to find evidence of highly related metrics, we could send out the message that one does not have to worry much about devising a selected metric. The researcher should, in this case, stick to the most straightforward measure of size available in his dataset. The result of strong concordance between the alternative metrics, in conjunction with a knowledge of the Copula function, could then be exploited in a methodology that allows one to simulate the remaining metrics conditional on the observability of one selected metric.

Upon having a justifiable measure of size at hand, a second purpose of the paper is to investigate the growth dynamics of the European pharmaceutical multinationals during the Single Market Programme era. We compare alternative estimation techniques and select the one that best explains our data. We predict the chosen size metric and repeat the simulation exercise, this time conditioning on the predicted value of the metric. As a digression we investigate how the alternative metrics of firm size available in our data satisfy Sutton's lower bound.

Gibrat's (1931) "law of proportionate effect" (English translation Gibrat (1957)) states that the expected *rate* of growth of a firm's size is, each time period, independent of its current magnitude. Such an innovative theory, capable of explaining important regularities, has motivated an extensive literature in industrial economics, with clear focus on the understanding of skewness in firm size distributions. This "growth-of-firms literature", also termed "stochastic literature on firm size", sees skewness as the result of the relationship between a firm's size and its rate of growth, i.e. if larger

[1] We will use interchangeably the terms "association" and "concordance" and we will be careful to confine the term "correlation" to cases of linear dependence of random variables.

firms tend to grow faster/slower than their smaller rivals then the industry size distribution tends to exhibit more/less skewness, this producing longer/shorter upper tails.[2] Albeit merely statistical, this literature has received considerable attention in industrial economics, due to its contribution in explaining firms' market concentration, one of the three components of the structure-conduct-performance paradigm that started with Mason (1939, 1948) and continued with Bain (1951, 1956).[3]

The lack of economic theory behind this "stochastic literature on firm size" gave room to a new stream of literature, originating in the 1970s, which employed either optimization theory, or game theory, to explain firm's growth.[4] In this approach skewness takes a more deterministic nature, given by the observable role of measurable economic factors.

Sutton (1997a, 1998) bridges the gap between these two streams of literature by proposing a game theoretical "independent submarkets" theory. He rationalizes skewness to be the result of the limiting firm size distribution of an industry. His theory of firm size distribution relies on a simplified discrete metric that counts the number of opportunities the firm has captured, expressed in his terminology as the number of "independent submarkets" the firm has entered. This simple metric suggested originally by Ijiri and Simon (1967) and Simon and Ijiri (1977), was termed by Sutton the number of "independent isolated islands" a firm expands to. Here size is determined by the cumulated sum of (homogeneous) unitary expansions. In his book, Sutton (1998) sketches how to extend the theory of unitary expansions to the more realistic case of unequal (heterogeneous) discrete opportunities (pp. 258-9, 290-1), called "random increments" in Simon and Ijiri (1977). In a nutshell, Sutton's main finding is that a lower bound in a Lorenz curve distribution (i.e. a minimum degree of inequality in firm size) is the best that one can do to unravel the industry distribution of firm size:[5] any distance between the actual data and the lower bound can be justified by, among other things, the degree of heterogeneity in the arrival of opportunities.[6]

[2]By modifying the boundary conditions and underlying assumptions, the literature has derived and discussed several distributions such as: Exponential, Fisher's log series, Geometric, Log-normal, Negative Binomial, Pareto (Zipf), Poisson, and Yule.

[3]A non-exhaustive list of followers includes: Simon (1955), Hart and Prais (1956), Adelman (1958), Simon and Bonini (1958), Hart (1962), Hymer and Pashigian (1962), Mansfield (1962), Ijiri and Simon (1964, 1967), Samuels (1965), Steindl (1965, 1968), Wedervang (1965), Quandt (1966), Singh et al. (1968), Simon and Ijiri (1977), Buldyrev et al. (2007), Cefis et al. (2007).

[4]Here a non-exhaustive list of papers includes: Hjalmarsson (1974), Lucas Jr (1978), Jovanovic (1982), Selten (1983), Hall (1987), Klepper and Graddy (1990), Cohen and Klepper (1992, 1996), Ericson and Pakes (1995), Klepper (1996), Klepper and Simons (2000), Cabral and Mata (2003), Klette and Kortum (2004), Klepper and Thompson (2006).

[5]The Lorenz curve offers a convenient way to represent the size distribution of firms. In the normalized area $[0,1]^2$, it shows the proportion of the measure of size controlled by a given proportion of number of firms. If the curve is a straight line, all firms would be of equal size, and the industry would be completely unconcentrated. On the other side, in the case of asymmetries, the curve would differ from the straight diagonal line. A typical feature here is that the top x percentage of firms control more than the top x percentage of business in that industry. The area between the diagonal and the curve is often utilized as a measure of concentration, and is known as Gini coefficient.

[6]Empirical support for this theory is available in Bottazzi et al. (2001), de Juan (2003), Amisano and Giorgetti (2005),

It is the relation between size measured as cumulated number of unitary opportunities and alternative heterogeneous metrics of size that motivates our work. Our paper considers a firm catching a new opportunity, whenever it expands its number of subsidiaries. In line with the "growth-of-firms literature" we treat arrival of opportunities as stochastic. However, aligned to its competing stream of literature, we investigate the dynamics of a firm's success in capturing opportunities in a more structural way. Do firms that capture more opportunities grow faster and, therefore, end up being the large companies in the industry? Assume, momentarily, that the cumulated number of opportunities satisfies the law of proportionate effects and, as such, gives rise to a skewed distribution. Could one go further and generalize the argument to any metric of size? That is, if a firm is large in cumulated number of opportunities, will the same firm be large in any other measure of size? The heterogeneity of opportunities would advise us to be cautious with a mapping from number of opportunities to alternative measures of size, as one cannot rule out patterns where big companies are those that have won only a small number of large opportunities. While Simon and Ijiri (1977) in a section called "random increments" manage to extend the "growth-of-firms" theory to cover alternative discrete measures of size, their generalization to (approximately) continuous metrics of size, is limited to the statement:

> Whether sales, assets, number of employees, value added, or profits, are used as the size measure, the observed distributions always belong to the class of highly skewed distributions . . . ,

re-emphasized in Sutton (1997b) as:

> Size can be measured in a number of ways, and these arguments have been variously applied to measures of annual sales, of current employment, and of total assets. Though we might in principle expect systematic differences between the several measures, such differences have not been a focus of interest in the literature.

This paper fills in the generalization gap by empirically investigating the association between alternative discrete and (approximately) continuous measures of firm size.

We develop our framework as follows. We first investigate how different metrics of size relate to one another in a joint multivariate nonnormal distribution, where nonnormality is induced by combining different (discrete and quasi-continuous) distributions. Given nonnormality, we rely on a Copula

Bottazzi and Secchi (2005, 2006), Buzzacchi and Valletti (2006).

approach: a methodology that only requires that one knows the empirical marginals of the variates in order to fit the multivariate distribution. Then, we carefully investigate the persistence of firm expansion, relying on the simplest discrete measure of size. To fulfil this goal we make use of dynamic panel Probit models, which account for unobserved heterogeneity and correct for the initial conditions problem. We give robustness to our results by comparing alternative Probit estimation methods. We believe it is fruitful to model firm expansion - the ability of firms at capturing opportunities - in a simple way, and then have a tool that bridges from that simplified measure into more sophisticated, but related, metrics of size.

The European pharmaceutical industry during the early period of the EU enlargement, 1990-2004, offers us a valid firm-level dataset. We focus on this industry for two reasons: firstly, it is an important industry for the European economy in terms of manufacturing value added and R&D;[7] secondly, the industry was heavily regulated at the national level before the Single Market Programme era (see Cecchini et al. (1988)), therefore the enlargement of the European market gives the firms in this industry significant scope to expand, increase production efficiency and enhance R&D capacity through mergers and acquisitions, relocation or external collaboration. In turn, this translates into more business opportunities.[8] Given that multinationals are the unit of investigation of this paper, we shall call firms "multinationals" and employ the abbreviation MNE. We obtain our data from a commercial database called *Amadeus*. It is worth mentioning that our data provide no information on MNE exit, meaning that exit dynamics cannot be studied.[9]

The rest of the paper is organized as follows: in Section 2 we present a simple conceptual model; in Section 3 we outline the dynamic panel Probit approach with unobserved heterogeneity and compare alternative econometric techniques. Section 4 describes our data. Our results are discussed in Section 5. Section 6 concludes.

2 The Conceptual Model

We generalize Steindl's (1965) formulation of Gibrat's (1931) law and allow MNE size to be measured by any metric $h \in \mathcal{H}$. We call the size of the MNE i at time t, s_{it}^h. We denote the *iid* random variable

[7]It is the 5^{th} largest industry in the European Union in terms of manufacturing value added, amounting to 3.5 per cent, and it accounts for about 17 per cent of total EU business R&D expenditures (EFPIA (2005)).

[8]Studies on the pharmaceutical industry which are relevant to this paper can be found in Howells (1992), Matraves (1999), Bottazzi et al. (2001), Kotzian (2004), Bottazzi and Secchi (2005, 2006), Buldyrev et al. (2007) and Cefis et al. (2007).

[9]For an investigation of the sample selection bias generated by nonrandom firm exit, refer to Hall (1987), Evans (1987a,b), Dunne et al. (1988, 1989) and Geroski and Machin (1991).

of proportionate rate of growth between $t-1$ and t, with ε_{it}^{h}. Absolute growth is expressed as

$$s_{it}^{h} - s_{i,t-1}^{h} = s_{i,t-1}^{h}\varepsilon_{it}^{h}. \tag{1}$$

Under the assumption of discrete time periods short enough to make the variance of ε_{it} small, log transformations of Eq. (1) lead to the random walk on a logarithmic scale approximation

$$logs_{it}^{h} \simeq logs_{i0}^{h} + \varepsilon_{i1}^{h} + \varepsilon_{i2}^{h} + \cdots + \varepsilon_{it}^{h}. \tag{2}$$

In the limit as $t \to \infty$, $logs_{i0}^{h}$ can be omitted from Eq. (2), being small compared to $logs_{it}^{h}$.[10] If we apply the Lindeberg-Levy central limit theorem we get that $logs_{it}^{h}$ has a limiting normal distribution. Hence, s_{it}^{h} has a skewed limiting lognormal distribution.

We wish to give Eq. (1) a general form. We do that in two parts. First, we decompose the original stochastic term, ε_{it}^{h}, into observable and unobservable components. Second, we express the right hand side of the equation with a function f. We assume f to depend on lagged size, and the observable and unobservable effects. We incorporate the observable effects into the vector $\mathbf{z}$. This modification, brings in a set of new parameters $\boldsymbol{\theta}^{h}$, and mitigates the role of the unobservable random component (random variable), re-labeled henceforth as u_{it}^{h}. Absolute growth becomes

$$s_{it}^{h} - s_{i,t-1}^{h} = f\left(s_{i,t-1}^{h}, \mathbf{z}_{it}^{h}, u_{it}^{h}; \boldsymbol{\theta}^{h}\right), \tag{3}$$

[10] Eq. (2) is easily obtained in the following way. Define the stochastic growth rate as

$$\varepsilon_{it}^{h} \equiv \frac{s_{it}^{h} - s_{i,t-1}^{h}}{s_{i,t-1}^{h}}.$$

We wish to compound the growth rates over time. Starting at an initial time 0, we have

$$\varepsilon_{i1}^{h} = \frac{s_{i1}^{h} - s_{i0}^{h}}{s_{i0}^{h}}.$$

Solving this for s_{i1}^{h} yields to

$$s_{i1}^{h} = \left(1 + \varepsilon_{i1}^{h}\right) s_{i0}^{h}.$$

Similarly,

$$s_{i2}^{h} = \left(1 + \varepsilon_{i2}^{h}\right) s_{i1}^{h} = \left(1 + \varepsilon_{i2}^{h}\right)\left(1 + \varepsilon_{i1}^{h}\right) s_{i0}^{h}$$

and for any future time t we get

$$s_{it}^{h} = s_{i0}^{h}\left(1 + \varepsilon_{i1}^{h}\right)\left(1 + \varepsilon_{i2}^{h}\right)\cdots\left(1 + \varepsilon_{it}^{h}\right).$$

Now, if we take the log of both sides of the above equation and avail of the fact that the approximation $log\left(1+\varepsilon\right) \approx \varepsilon$ (true for small variance of ε), we reach Eq. (2).

which is nothing more than a generalization of Eq. (1).[11]

We denote this absolute growth of the MNE i in period t, with $y_{it}^h \equiv s_{it}^h - s_{i,t-1}^h$. If we were willing to assume the error component u_{it}^h to be *iid*, we could claim invertibility of the function f in u_{it}^h; meaning that a relation $u_{it}^h = f^{-1}\left(y_{it}^h, s_{i,t-1}^h, \mathbf{z}_{it}^h; \boldsymbol{\theta}^h\right)$ exists. This assumption, along with the further requirement that the conditional distribution $P\left(u_{it}^h | s_{i,t-1}^h, \mathbf{z}_{it}^h; \boldsymbol{\theta}^h\right)$ is known - or $P\left(y_{it}^h | s_{i,t-1}^h, \mathbf{z}_{it}^h; \boldsymbol{\theta}^h\right)$ given the invertibility of f is known - would allow us to access the maximum likelihood estimator. However, as we wish to test the theory of maturation (Rostow (1959)), which states that: firms face a period of rapid growth, followed by a slow down or even a stop in growth, we model autocorrelation in the error term. So, the probability that a firm experiences an increment in size the next period, is proportional to a weighted sum of the increments it had in the past (Ijiri and Simon (1967), Simon and Ijiri (1977)). By introducing autocorrelation we compromise, giving away the simple invertibility of f in u_{it}. We will outline how to recover appropriate estimation techniques in Section 3 and detail the procedure in Appendix A.

Among the elements of $\mathcal{H}$ we have a simple discrete measure of size: the number of new opportunities (which we interchangeably call expansions) an MNE has captured. We denote this metric with $h = op$ and update Eq. (3) to be

$$y_{it}^{op} = f\left(s_{i,t-1}^{op}, \mathbf{z}_{it}^{op}, u_{it}^{op}; \boldsymbol{\theta}^{op}\right), \tag{4}$$

where $y_{it}^{op} \in \{0, 1\}$, with a value of 1 indicating an expansion, which happens if the MNE expands the number of its subsidiaries by *at least* one unit.

Our decision to have lagged number of opportunities, $s_{i,t-1}^{op}$, directly influencing the absolute growth is motivated by Gibrat's law, as re-interpreted in Simon and Ijiri (1977) and Sutton (1998). Each active MNE in the industry has a certain probability of capturing an opportunity, i.e. a certain probability of expanding. The number of opportunities captured has a direct influence on each MNE's propensity to expand, thus conforming to Gibrat's law: the probability that an MNE captures a new opportunity is proportionate to its size. As an MNE becomes larger it has more chances to catch new opportunities, and the continuation of this process is reflected in the skewness of the industry distribution.

Behind our structural model there is an underlying economic story. In every period of time, MNEs

[11] If we parameterize f to be

$$f\left(s_{i,t-1}^h, \mathbf{z}_{it}^h, u_{it}^h; \boldsymbol{\theta}^h\right) = s_{i,t-1}^h\left(\mathbf{z}_{it}^h \boldsymbol{\theta}^h + u_{it}^h\right)$$

and then define $\varepsilon_{it}^h \equiv \mathbf{z}_{it}^h \boldsymbol{\theta}^h + u_{it}^h$, we are back to Eq. (1), assuming u_{it}^h is *iid*.

are given the chance to expand their size by capturing one of the opportunities available in the market. We motivate the "race" for opportunities in the following way. Depending on their profit realization, our agents succeed, or not, in capturing one of the available opportunities. However, opportunities can have a different degree of intensity that is, we count opportunities as single homogeneous units, but recognize that each opportunity spreads effects into alternative heterogeneous metrics of size. Understanding how sum of opportunities and sum of alternative measures of size relate one another is a target that we aim to. Our data allow for three alternative metrics of size. Two are discrete: i) the aforementioned number of opportunities captured, s^{op}; ii) number of subsidiaries established or acquired, s^{su}, and one is (approximately) continuous: iii) operational revenues, s^{or}. A more exhaustive list of single measures of size adopted in the literature includes: assets, capital, employment, inputs, output, plants and equipment, profit, sales and turnover, as documented in Table 1.

We assume the existence of a positive monotonic relation between absolute growth and profit. We formulate the underlying profit function as the sum of the following components:

$$\pi_{it} = q\left(s^h_{i,t-1}, \mathbf{z}^h_{it}\right) + u^h_{it}. \tag{5}$$

The algebraic expression in (5) tells us that MNE i's profit in period t, π_{it}, is given by the sum of an observable part, expressed by the deterministic function of pre-determined variables and lagged size, $q\left(s^h_{i,t-1}, \mathbf{z}^h_{it}\right)$, and the unobservable variable, u^h_{it}.[12] The lagged observable $s^h_{i,t-1}$ influences the absolute growth either directly, as discussed earlier, or in an indirect way; through scope economies, economies of scale in terms of production and R&D, or superior market power that can be gained as MNEs evolve.

The unobservable part of Eq. (5) is a composite error that groups a time-specific function of unobservable multinational shifters and a pure idiosyncratic error component, say

$$u^h_{it} = c^h_{it} + \epsilon^h_{it}. \tag{6}$$

We turn our attention back to the simplest discrete metric of size, s^{op}. Considering that one firm can only expand by zero or one opportunity, we relate the absolute growth and profit in the following

[12] One could model the observable term $\mathbf{z}^h_{it}$ as a "portmanteau" variable, or alternatively as inclusive of factors related to the demand and cost side, as discussed in Quandt (1966). The latter would be an attempt to model profit as the result of a combination of demand shocks, level of competition, scope economies and success of R&D investments.

binary way:

$$y_{it}^{op} = \mathbb{1}\left(\pi_{it} \geq 0\right), \tag{7}$$

where $\mathbb{1}$ is the indicator function.[13]

With an obvious interest in the dynamics of expansion, we choose the variables entering $\mathbf{z}_{it}^{op}$ to be the lagged dependent variable $y_{i,t-1}^{op}$, along with a set of time dummies and observable MNE-specific characteristics such as a vector of time-invariant characteristics. The lagged dependent variable is aimed at capturing any form of persistence in the growth process. Its inclusion is in line with Ijiri and Simon's (1967) model of autocorrelated growth, whose simulations are the core of Ijiri and Simon (1964).[14]

Although not the main focus of the paper, we believe that applying this model to the pharmaceutical data contributes toward a better understanding of European pharmaceutical industry dynamics and the resulting industry concentration. Evidence of a positive and significant coefficient, attached to an expansion that occurred at time $t-1$, would indicate the sustainability of this industry. In other words, it would suggest that the industry is able to generate profits that are large enough to support its expensive innovative activities. In this way, past expansions exert a behavioral effect on current expansion and this effect is termed "true state dependency" by Heckman (1981b). However, as is often the case, the examination of dynamics can be blurred by what Heckman (1981b) has termed "spurious state dependency". This occurs if unobservable multinational-level effects are serially correlated over time, or are correlated with initial expansion, and these correlations have not been properly controlled for. In those cases, the lagged expansion will incorrectly capture this unobserved effect, behaving as if it is a driving force behind the current expansion, even if there is no state dependence at all. To tackle this problem we adopt a dynamic panel random effects model. We will discuss the mechanism and detail of this model in the next section.

Eq. (7) suggests an econometric relation to study size growth, when size is measured by number of opportunities. What happens if size is measured by any other metric? Here, one might be tempted to extend the mapping between profit and expansion to accommodate each alternative metric of size. However, this would be a tedious and difficult task, as we would have to assume a mapping function

[13] We have arbitrarily set the profit cutoff at zero. Alternatively, we could have set it at $\underline{\pi}$, or at $\underline{\pi}_t$ if it varied over time. The implication of not observing the cutoff point is that in our econometrics we shall not be able to identify either the constant, if $\underline{\pi} \neq 0$, or the time dummies, if $\underline{\pi}_t \neq 0$.

[14] The serial correlation assumption states that the probability of growth of an existing MNE is proportional to the weighted sum of past increments of size, and the weight is decreasing the further the occurrence of each increment from the current period. Such a carry-over effect can be triggered by successful innovation in production or marketing processes.

for each metric, and subsequently estimate each resulting (linear or nonlinear) dynamic model. Given the intricacy of the procedure, we opt for an alternative approach, which we believe simplifies and accelerates the process. We define the joint density function of the various metrics of size, with $G(\mathbf{s})$, where $\mathbf{s} \equiv \left[s^1, s^2, \cdots, s^H\right]$. Next, given that we are mixing together discrete and continuous measures of size, it would be inappropriate to assume, as is often the case, $G(\cdot)$ to be multivariate normal, so the functional form of $G(\cdot)$ has to be fitted. We exploit the information on the marginal distribution of each metric of size, $G_h(\cdot)$, and employ the Copula approach to recover the multivariate CDF $G(\cdot)$ from the H marginals. Knowing $G(\cdot)$ we undertake the following exercise. We test the Copula by simulating $\tilde{\mathbf{s}}^{-[\mathbf{h}]}$ from the conditional distribution $G(\mathbf{s}|s^h)$, where s^h denotes the values for the selected size metric h and the tilde vector $\tilde{\mathbf{s}}^{-[\mathbf{h}]}$ indicates the simulated values for the remaining size metrics.[15] We evaluate the goodness-of-fit of the simulated metrics. We continue the exercise by simulating a new vector $\tilde{\tilde{\mathbf{s}}}^{-[h]}$, this time from the conditional Copula $G(\mathbf{s}|\hat{s}^h)$, where $\hat{s}^h$ is the predicted value of the size metric. We evaluate, once more, the quality of the new simulations. We present the Copula approach in Appendix B.

The next section outlines the econometric methodologies we make use of.

3 Econometrics

Given that this section only discusses methods to estimate the absolute growth of the variable number of opportunities, we ease the notation by omitting the superscript *op*. We specify the latent profit function for our dynamic model as

$$\pi_{it} = \gamma y_{i,t-1} + \mathbf{x}_{1it}\boldsymbol{\beta}_1 + \mathbf{x}_{2it}\boldsymbol{\beta}_2 + c_{it} + \epsilon_{it} \qquad i = 1, \cdots, M; \qquad t = 1, \cdots, T, \tag{8}$$

where $y_{i,t-1}$ is the lagged version of the binary variable. We partition the explanatory covariates into a row vector of strictly exogenous variables $\mathbf{x}_{1it}$ and a row vector of sequentially exogenous variables $\mathbf{x}_{2it}$. We formulate the c_{it} component as a time varying function of the unobserved MNE-level heterogeneity

$$c_{it} = \delta_t \tilde{\alpha}_i, \tag{9}$$

whose δ_t parameters pertain to cases of free correlation in the composite error, as we will discuss below. The ϵ_{it} term is an idiosyncratic error, which we assume to be identically distributed and independent

[15] In our data $\tilde{\mathbf{s}}^{-[\mathbf{h}]}$ is the vector inclusive of $\tilde{s}^{su}$ and $\tilde{s}^{or}$.

of unobserved heterogeneity and the covariates. Our data share a pattern common in firm-level data: the number of MNEs M is large relative to the number of periods T, so asymptotics rely on $M \to \infty$.

The presence of a large cross-section in a nonlinear panel model rules out the possibility of modeling the $\tilde{\alpha}_i$ as parameters. In fact, because of the "incidental parameters" problem a fixed effects analysis would produce inconsistent estimates of the parameters (Heckman (1981a,b)). So, the rest of the paper will treat the unobserved heterogeneity as a random variable drawn along with $(y_i, \mathbf{X}_i)$ where $\mathbf{X}_i \equiv [\mathbf{x}'_{i1}, \mathbf{x}'_{i2}, \cdots, \mathbf{x}'_{iT}]'$, $\mathbf{x}_{it} \equiv [\mathbf{x}_{1it}, \mathbf{x}_{2it}]$ and $t = \tau_i, \tau_{i+1}, \cdots, T$. We indicate with τ_i the period in which MNE i appears in the sample for the first time. So, for the balanced panel (the incumbents) we have $\tau_i = 1$ (and initial conditions at time $\tau_{i-1} = 0$) and for the new entrants $\tau_i > 1$.

What we observe in our data is not the latent profit function shown in Eq. (8), but rather the binary outcome of an MNE expansion, whose relation with profits has been represented in Eq. (7).

We assume the idiosyncratic error, ϵ_{it}, to be distributed as $NID(0, \sigma_\epsilon^2)$ and given that y_{it} is a binary variable, we standardize the idiosyncratic error as $NID(0,1)$. The implication is that all parameters, as well as the function of unobserved heterogeneity, will be re-scaled by σ_ϵ, as shown in Arulampalam (1999). Hence, the conditional probability that an MNE i expands in period $t = \tau_i, \tau_i + 1, \cdots, T$ is

$$P\left(y_{it} = 1 | y_{i,t-1}, \cdots, y_{i,\tau_{i-1}}, \mathbf{X}_i, \tilde{\alpha}_i; \boldsymbol{\theta}\right) = \Phi\left(\gamma y_{i,t-1} + \mathbf{x}_{1it}\boldsymbol{\beta}_1 + \mathbf{x}_{2it}\boldsymbol{\beta}_2 + c_{it}\right), \tag{10}$$

where Φ is the standard normal cumulative density function and $\boldsymbol{\theta}$ are the parameters to be estimated. The joint conditional density for $(y_{i,\tau_i}, y_{i,\tau_i+1}, \cdots, y_{iT})$ results in the following dynamic unobserved effects Probit model,

$$P\left(y_{i,\tau_i}, y_{i,\tau_i+1}, \cdots, y_{iT} | y_{i,\tau_{i-1}}, \mathbf{X}_i, \tilde{\alpha}_i; \boldsymbol{\theta}\right) = \prod_{t=\tau_i}^{T} \Phi\left[(\gamma y_{i,t-1} + \mathbf{x}_{1it}\boldsymbol{\beta}_1 + \mathbf{x}_{2it}\boldsymbol{\beta}_2 + c_{it})(2y_{it} - 1)\right]. \tag{11}$$

The presence of unobserved heterogeneity makes the log-likelihood function of the above density not suitable to estimate the $\boldsymbol{\theta}$ parameters consistently, unless one has a way to integrate the unobserved heterogeneity out. In order to do so we need, first, to ensure that we account for any possible correlation between the unobserved heterogeneity and the regressors - given that not all our regressors, and surely not the lagged dependent variable, are orthogonal to the unobserved heterogeneity. Secondly, for the *balanced* sub-sample we shall have a way to cope with the initial conditions problem, i.e., an existing relation between the initial observations of the dependent variable y_{i0} and the unobserved heterogeneity $\tilde{\alpha}_i$. This is an effect that is induced by the fact that the stochastic process that has determined an expansion in the initially observed period, which corresponds to the first period we have data available (period 0 in our notation), has been ongoing prior to that date and, as such, we cannot take it as exogenous. The initial conditions problem is particularly severe for small T.

Both of the above issues can be tackled. We make use of the Mundlak (1978)-Chamberlain (1984) approach and account for the correlation between the unobserved heterogeneity, the subset of sequentially exogenous regressors $\mathbf{x}_{2it}$ and the lagged dependent variable, so that the latent profit function displayed in Eq. (8) is augmented to be

$$\pi_{it} = \gamma y_{i,t-1} + \mathbf{x}_{1it}\boldsymbol{\beta}_1 + \mathbf{x}_{2it}\boldsymbol{\beta}_2 + \bar{\mathbf{x}}_{2i}\boldsymbol{\lambda}_1 + c_{it} + \epsilon_{it}. \tag{12}$$

The terms on its right hand side are in following order: the lagged dependent variable; the vector of strictly exogenous variables, $\mathbf{x}_{1it}$, which includes a constant, MNE headquarters area dummies (eu_i and us_i) and $T-1$ time dummies; the vector of sequentially exogenous variables, $\mathbf{x}_{2it}$, which incorporates the lagged size variable measured by number of opportunities, $s^{op}_{i,t-1}$; the vector $\bar{\mathbf{x}}_{2i}$, which corresponds to the average number of opportunities $\bar{s}^{op}_i$ - employed along with the lagged dependent variable and lagged size to study the theory of maturation (Rostow (1959)); the unobserved heterogeneity, which we have parameterized in Eq. (9) as $c_{it} = \delta_t\tilde{\alpha}_i$; and to complete the list, the idiosyncratic error term ε_{it}. The term $\bar{\mathbf{x}}_{2i}$ in Eq. (12) is computed as

$$\bar{\mathbf{x}}_{2i} \equiv \frac{1}{T-\tau_i+1}\sum_{t=\tau_i}^{T}\mathbf{x}_{2it}. \tag{13}$$

Turning to the initial conditions problem, the panel data econometric literature has developed alternative ways to deal with it. Heckman (1981a,b) suggests recovering a full conditional density for $(y_{i0}, y_{i1}, \cdots, y_{iT}|\mathbf{X}_i)$, by extending the original density to the initial period and integrating out the unobserved heterogeneity. To fulfill his idea he specifies first a parametric density for y_{i0} given $(\mathbf{X}_i, \tilde{\alpha}_i)$, thus extending the density to the initial period, and then a parametric density for $\tilde{\alpha}_i$ given $\mathbf{X}_i$, so as to integrate out the unobserved heterogeneity. The lack of an existing program to estimate Heckman's full model has given rise to alternative estimation methods. Orme (2001) introduces a more immediate two-step procedure, which is suitable to cases of low correlation between the initial conditions and the unobserved heterogeneity. Wooldridge (2005) proposes parameterizing a conditional density for the unobserved heterogeneity only, so as to integrate out the unobserved heterogeneity, leaving the density for $(y_{i1}, y_{i2}, \cdots, y_{iT})$ conditional on $(y_{i0}, \mathbf{X}_i)$.

As mentioned, the initial conditions problem concerns only the MNEs that are in the sample the entire time period, the balanced panel, not those that enter at some time $\tau_i > 1$, the new entrants. In Appendix A we internalize this distinction in our modeling of probabilities for the two groups. Each of the three solutions can be estimated in `Stata` either directly, or via an add-on program called `gllamm`. In Appendix A we detail the three methods and discuss the estimation procedures that have been implemented in `Stata`.

4 Data

We gathered information on pharmaceutical MNEs from a commercial database called *Amadeus*, which is published by Bureau van Dijk (BvD) Electronic Publishing. The publisher collects firms' account data from official or commercial sources of individual European countries and processes the data in order to achieve maximum comparability across countries. The database contains balance sheets, profits and losses tables, and other information among which is business activity, date of corporation, location, ownership, etc. This database is by far the most comprehensive source of financial information on European firms.[16] A firm is recorded by *Amadeus* as an EU pharmaceutical MNE, if at some time t it has pharmaceutical-related subsidiaries (either production, research or marketing driven) in more than one country, and if at least one of these countries belongs to the EU-15.[17] Acquired or new own-established subsidiaries are treated in the same way, as we believe that both greenfield investments and acquisitions reflect MNEs' willingness and action to expand. Expansions through both platforms are influenced by industry-level trends, such as relocation of business activities, or adjustments to the Single Market. Consequently, acquired or new own-subsidiaries are components of the same industrial dynamics.

The starting year of the sample period is chosen to be 1991 ($t = 1$ marks the end of the year in our notation), which is just two years before to the implementation of the European Single Market, which occurred on the first day of 1993.[18] The last period is the year 2004 (T in our notation). Information on firm exit is not available in *Amadeus*. Also, mergers complicate the identification of the MNEs in

[16] At the time when we prepared the sample of the pharmaceutical industry, year 2004, *Amadeus* covered approximately 11 million public and private firms in 41 European countries.

[17] Following the EU statistical classification of economic activities, NACE Revision 1.1, they are classified as: 2441 (manufacture of basic pharmaceutical products), 2442 (manufacture of pharmaceutical preparations), 5146 (wholesale of pharmaceutical goods), 5231 (dispensing chemists), 5232 (retail sale of medical and orthopaedic goods) and 7310 (research and experimental development on natural sciences and engineering).

[18] We expect that the pharmaceutical MNEs have taken into account some of the potential gains from the integrated market in their business strategy, triggering their expansion pattern already in early-nineties.

the sample, as the *Amadeus* database drops the MNEs that have been acquired, and only records the acquirer. This makes it impossible to construct a proper panel of expansion history. To get around this difficulty, we adopt Bottazzi and Secchi's (2005; 2006) procedure, which treats the merged enterprises as single entities throughout the period of investigation.

The total number of MNEs in the dataset is 265. For these 265 MNEs we count 827 new subsidiaries established or acquired and 930 existing subsidiaries, for a total of 1757 subsidiaries. These are the result of either greenfield investments by MNE parent companies themselves, or acquisitions through mergers or acquisitions.

Given that our paper generalizes the study of firm growth to alternative metrics of size, in the rest of this section we describe the behavior of each measure that our database allows for.

We directly use the information on subsidiaries to generate an initial discrete measure of MNE size: the cumulated number of subsidiaries acquired over time. We indirectly utilize the data on subsidiaries to construct a second discrete variable of MNE size: the cumulated number of opportunities, *alias* expansions, captured over time.[19]

All in all the size metrics we make use of, are the following:

1. Discrete metrics:
    (a) s_{it}^{op} is the stock of business opportunities captured by the MNE i up to period t. It serves as our most straightforward measure of size, that we use to investigate the dynamics of MNE expansion.
    (b) s_{it}^{su} is the total number of pharmaceutical-related subsidiaries the MNE i has established or acquired up to year t.
2. Quasi-continuous metric:
    (a) ls_{it}^{or} is the log transformation of MNE i's operational revenues.[20] This is a stock variable obtained by summing up, each year, operational revenues of all pharmaceutical-related subsidiaries belonging to that MNE, and taking its logarithm.

[19] Our dataset allows us to trace the absolute growth (the flow) of the two aforementioned discrete metrics, back to the year 1983. One issue we had to cope with is that size (the stock) at year 1983 was only available for the metric "number of subsidiaries". In order to get around this issue, we proxied size measured by the total "number of opportunities" at year 1983, in the following way. For the time span 1983-2004, we obtained the cumulative absolute growth of both number of opportunities and number of subsidiaries. For the specified period, we calculated the ratio between cumulated number of opportunities and cumulated number of subsidiaries. Next, under the assumption of regularity in the expansion process, we approximated the cumulated number of opportunities for 1983 by multiplying our computed ratio with the number of subsidiaries acquired by the MNEs up to 1983.

[20] As operational revenues is a (quasi-) continuous random variable, we take its log as we wish to test if the distribution is lognormal.

Table 2 provides valuable information on our metrics. The first row displays several statistics that should facilitate an understanding of the expansion process, y_{it}^{op}. We see that the average absolute growth of opportunities captured by the cross-section of MNEs, i.e. the share of MNEs that have expanded in a given period, ranges between 14 and 22 per cent most of the time, before shrinking to lower values after 2001. The second row documents the statistics for the stock of opportunities, s_{it}^{op}. We note that the US company Schering Plough (S.P.) holds the largest number of opportunities, a number that rises from 23 to 24, before being overtaken in 2004 by the Swiss company Novartis, with its 25 opportunities. The MNEs' average number of opportunities increases about a unit during the period of investigation, but that comes along with a rise in the dispersion. Skewness does not follow the same trend as it slightly contracts during the period. A similar set of statistics is documented for number of subsidiaries. Once more, we describe first the absolute growth of this metric. We notice that the largest growth occurs in year 2000 when the largest MNE establishes or acquires 17 subsidiaries. Year 2000 is certainly a boom year, as it displays the largest increase in the industry, with 84 new subsidiaries attained in total. Average growth in number of subsidiaries shares a downward trend with the average growth in number of opportunities. What differs here is the dispersion, in this case marked by a positive trend with a rather oscillating behavior. If we convey our attention to the stock of subsidiaries, s_{it}^{su}, we directly observe the evolution of the largest company Novartis, which steadily grows over time, reaching a count of 49 subsidiaries in 2000. From year 2003 Pfizer takes over the lead, eventually holding 55 subsidiaries in 2004.[21] MNEs hold, on average, 4.86 subsidiaries in 1990 and that number grows to 6.63 in 2004. It also happens that the inequality in subsidiary ownership increases over time, as indicated by the upward movement of standard deviation and skewness.

Table 2 additionally provides statistics for our (quasi-) continuous metric of size, operational revenues, s_{it}^{or}. Here, the quality of the *Amadeus* dataset limits the time span of availability to the period 1995 to 2003, as too many missing values were present prior to 1995, and there was no usable information in 2004. Missing values were also present in the time span 1995-2003, starting with a total of 67 observations (out of 226) in year 1995 and dropping to a negligible 5 observations (out of 265) in year 2003.[22] It is worthwhile mentioning that in order to make the table more readable all statistics of operational revenues, aside from skewness, are expressed in millions of dollars. Skewness is related to the log transformation of operational revenue, which is the variable that we use in the Copula approach

[21] Pfizer (US) acquired Pharmacia & Upjohn (UK) in 2003.

[22] At the beginning of this section we praise the *Amadeus* dataset but one serious problem we came across is the large number of missing values that relevant variables such as profit, employment etc. presented. Due to this issue, we found operational revenues to be the only continuous variable that was worth using.

and in the simulations.[23] The table reveals that the leading MNE in terms of operational revenues is Bayer, which retains the leading position for most of the period of interest. Operational revenues, dispersion and skewness all grow over time, and almost double their values from 1995 to 2003.

We provide further evidence on the behavior of the three size metrics by inspecting the frequencies and distributions. We plot the histograms of the three size metrics in Figure 1 for the first and last year of the sample in which all metrics are available, i.e 1995 and 2003. An examination of the two discrete metrics of size, s^{op} and s^{su}, confirms that the size distribution of the European pharmaceutical industry is skewed towards small firms. The number of MNEs increases from 226 to 265, between 1995 and 2003, and this makes the visual comparison of the two time periods more difficult. The values of skewness reported in Table 2 help our reading of the figures, suggesting an almost unnoticeable contraction in skewness, if we compare Sub-figures 1(a) with 1(d), and an increase in skewness, if the comparison is between Sub-figures 1(b) and 1(e). The remaining Sub-figures, 1(c) and 1(f), depict the frequency of the log of operational revenues. The histogram resembles more of a normal distribution (which would approximate lognormality of operational revenues) but still presents a certain degree of skewness, as confirmed by the statistics reported in Table 2. At the lower end of the distribution of 1995 the prominent bar of zeros indicates that there are several firms with zero operational revenues. By 2003 the number of MNEs with zero operational revenue reduces markedly.

We have emphasized the role of skewness above, as this statistic has been at the core of study of the growth-of-firms literature.

The last rows of Table 2 add information on the break-down of the number of MNEs in our sample by the geographical location of the headquarters (EU, US, Other). Furthermore, statistics are included indicating the number of new entrants each year. We note that the number of new entrants reaches a peak of 8 MNEs in 1996 and 1998 and decreases thereafter, vanishing by 2004. Relative to the full sample of MNEs, new entrants have a very marginal role, never going beyond 3.5 per cent. The fact that this particular industry displays "weak" dynamics of entry is convenient, because it allows one to ignore the aforementioned concern that Gibrat's law might be rendered invalid in light of excessive entry and exit, recalling that exit is not documented in our dataset. In the last line, we document the frequency that an opportunity is captured by a new entrant, and we see an oscillating trend that goes from 0 to 19 per cent.

In addition to the frequency histograms, we plot in Figure 2 the empirical cumulative density func-

[23] We define the log of a transformation of s^{or} as: $ls^{or} \equiv log(1 + s^{or})$.

tions (ECDF) for the normalized measures of size. We observe that in 1995 the ECDFs of the discrete metrics s^{op} and s^{su} track one another, highlighting the high proportion of firms with a low value of these discrete measures of size. Such a pattern is not shared by the ECDF of the *log* of s^{or}. The latter is less skewed and closely resembles a normal distribution. In 2003 the ECDF lines for s^{op} and s^{su} depart from one another - a trend that we expected, given the skewness statistics reported in Table 2. As for the ECDF of ls^{or}, we do not notice much difference from the 1995 line.

The cross-analysis of the descriptive statistics from Table 2, and the visual inspections of Figure 1 and Figure 2, convey the following message. While the discrete measures of size share, with the exception of skewness, all their trends, the (quasi-) continuous metric of size seems to exhibit a different behavior. If we were to draw conclusions from these descriptive statistics we would state that only the discrete measures of size co-move. This would be a shallow statement as we need further investigation. In order to produce a more reliable study of the cross-relation of the variables, we appeal to a Copula approach. If we can show with this method that all the metrics of size concord, then it would be enough to analyze the growth dynamics of any metric, and in that case, why not choose the most straightforward one?

Given that the second purpose of the paper is to estimate a dynamic model of pharmaceutical MNE growth relying on size measured by number of opportunities, we now discuss the explanatory variables that we use in our estimations:

1. $y^{op}_{i,t-1}$ is a one-year lagged dependent variable, which equals 1 if some subsidiaries were established or acquired between $t-2$ and $t-1$, i.e., the MNE i captured an opportunity in the previous period.

2. Two dummy variables for ownership: eu_i and us_i. They take value 1 if the MNE i has headquarters located in the EU or in the US, respectively, and zero otherwise. The reference group is *Other*.

3. $s^{op}_{i,t-1}$ is the stock of business opportunities captured by the MNE i up to period $t-1$.

4. Instruments $\bar{s}^{op}_i$ and $age_{i,89}$.

   (a) $\bar{s}^{op}_i$ is the time mean of $s^{op}_{i,t-1}$.

   (b) $age_{i,89}$ is the pre-sample MNE age in 1989 (t=-1 in our notation), postulated to be exogenous.

The next section discusses our results.

5 Results

As mentioned in previous sections, the proper way to link a mixture of discrete and continuous variables or, generally speaking, variables that come from different distributions, is to utilize a multivariate distribution that has marginals uniform over [0,1], i.e. to employ a Copula approach. We start this section by discussing the results obtained from applying this technique to our data.

In order to be able to estimate the Copula, we first fit the parametric distributions of the various measures of size. Table 3 exhibits, for each time period, the estimated parameters of distributions with positive support. We investigate four discrete distributions - Exponential, Geometric, Negative-Binomial and Poisson - and four continuous distributions - Normal, Exponential, Gamma and Weibull. Most of the distributions rely on one parameter, except for the Negative-Binomial, Normal and Gamma, which rely on two.[24] We only fit the distributions of the various metrics of size for the period 1995-2003 as the data on operational revenues are limited to that time span. The first panel of Table 3, investigates size measured by cumulated number of opportunities. Here, it is the Geometric distribution that gives the best fit, as one can see from the Pearson χ^2 goodness-of-fit statistic.[25] The same parametric distribution also provides the best approximation for size measured by cumulated number of subsidiaries, though in this instance we point out the fact that the statistic is only accepted at the 1 per cent level in some cases. As for the (quasi-) continuous metric, the log of a transformation of operational revenues, we select the Weibull distribution and remark on the fact that in two periods the statistic is accepted only at the 1 per cent level. The last column of Table 3 shows the results for the pooled data. None of the distributions that we investigate describes satisfactorily the entire period.

With the parametric distributions on hand for each time period, we have all the information required to fit a parametric Copula. We utilize the package `copula` written in `R`.[26] We fit, by maximum likelihood, the most popular classes of Copulas: Elliptical (Normal) and Archimedean (Clayton, Frank and Gumbel). For all time periods, Table 4 provides the estimated parameters which are: the marginal

[24]We have utilized libraries in `R` to fit the distributions. Refer to Ricci (2005) for details on the procedures adopted.

[25]The Pearson χ^2 goodness-of-fit computation, divides the data into J bins and tests the following statistic:

$$\chi^2 = \sum_{j=1}^{J} \frac{(O_j - E_j)^2}{E_j},$$

where O_j is the observed frequency for bin j and E_j is the expected frequency of bin j (calculated from the chosen CDF).

[26]Consult Yan (2007) and Kojadinovic and Yan (2010) for a detailed exposition of the technique.

parameters and, either the concordance coefficients for the Elliptical Copula, or the Copula parameter for the Archimedean Copula. The arguments of the likelihood maximization are: the concordance coefficients, the Copula parameter and the new marginal parameters. We look at the log-likelihood values and choose as best fitting Copula the Elliptical Normal Copula.

Relevant to our paper is the evidence of strong concordance of the three variates, as confirmed by the high positive values of the Spearman's ϱ estimated parameters.[27] This result relaxes the choice of our metric. So, we choose as our measure of size the cumulated number of expansions/opportunities, s^{op}. We utilize this metric, not only because we believe it is the most straightforward measure, but also because we think it is a building block of the remaining metrics.[28]

In order to evaluate the accuracy of the employed Copula we simulate, for all time periods and metrics of size, 1000 observations.[29] Of those, each period we retain the M_t observations that have the simulated cumulated number of opportunities, $\tilde{s}^{op}$, closest to the true observed values s^{op}, i.e. we condition the Copula simulations on the values of size measured by number of opportunities.[30] That is, we get conditionally on the observed s^{op}, the simulated values for the remaining two metrics. We repeat this procedure $bs = 100$ times. Table 5 displays statistics on the goodness-of-fit of the Copula, inclusive of a 95 per cent empirical confidence interval. In its first column we have the average (over the bs repetitions) simulated coefficient of determination, which tells us that conditioning on the number of opportunities, on average we are able to simulate correctly almost 80 per cent of number of subsidiaries and over 60 per cent of the log of operational revenues. As can be seen from the 95 per cent empirical confidence intervals, both coefficients of determination are highly concentrated. This signals that we have a fair methodology to move from one metric to another. To validate further the estimated Copula, in its second column Table 5 reports the average simulated p-value of the Pearson χ^2 test for the equality of the simulated distributions and the distributions observed in our data. With a probability value of 5 per cent we accept, in all time periods, the equality of distributions of the number of subsidiaries. We only accept the first period for the log of operational revenues, but we notice that all time periods satisfy a 1 per cent level test. Overall we judge positively the goodness-of-fit of the conditional simulations.

At this point, with a selected metric at hand, and a consolidated mapping of metric to metrics at

[27] Alternatively we could have used Kendall's τ, as a measure of association.

[28] Even though technically in our data we derived opportunities from the number of subsidiaries.

[29] We employed the `mvdc` function, which is part of the `Copula` package in `R`, to obtain our simulations. The `mvdc` function requires as inputs the parametric distributions for all variables (size metrics) and their estimated parameters.

[30] In case of multiplicity of simulated observations close to a certain value of the observed metric, we randomly pick one of them.

our disposal, we proceed to the second purpose of the paper: investigating the growth dynamics of European pharmaceutical MNEs.

We commence with a brief recap. In Section 2 we have related the profit function to a functional form of any metric of lagged size and some shifters (observable and unobservable). We have further developed the profit relation into the econometric Eq. (12). In the econometrics section, Section 3, we have set out the conditions required to estimate the parameters of the binary relation expressed in Eq. (7) via a dynamic random coefficients Probit estimator. We have pointed out the issue of possible correlation between the unobserved heterogeneity and a subset of regressors that include the lagged dependent variable and the sequentially exogenous variables. We have suggested the Mundlak-Chamberlain correction, as a solution. Now, we analyze the benefit of introducing this correction, in either a pooled dynamic Probit model or a random effects dynamic Probit model. We impose equicorrelation, which means that in Eq. (9) we fix $\delta_t = 1$ for all $t = 1, 2, \cdots, T$ and $\delta_t = \delta_0$ for $t = 0$. The results are documented in Table 6. The introduction of the correction term on the one hand invigorates the effect of state dependence, on the other hand it produces a significant mean reversion effect on size. This latter consequence brings into the framework rich trajectories of growth that depend, among other things, on how far away the MNE is from its long run average size. Another interesting result we spot in the table is the reduced weight that unobserved heterogeneity carries upon the inclusion of the Mundlak-Chamberlain correction term in the random effects estimator; as confirmed by the lower values of the estimated coefficient of intertemporal correlation, $\hat{\rho}$. A final important finding that we highlight is the lack of predictive power of the estimations without the Mundlak-Chamberlain adjustment. It strikes us to see how unsuccessful the estimates without the correction term are in predicting the positive outcomes of expansion. The introduction of the correction term brings the percentages up to figures above 12 per cent, which is still not high but certainly is encouraging.[31]

Also, we have stressed that the combination of the balanced part of the panel commencing prior to our initial period of observation, and a nonlinear dynamic panel model with unobserved heterogeneity, brings in the problem of possible correlation between the initial conditions and the unobserved heterogeneity. In Appendix A we present and review solutions to this initial conditions problem, and set out several estimation techniques that deal with it directly in `Stata`. In the same appendix, we show that all of the solutions, but the "Wooldridge" one, require an initial conditions equation. We specify such an equation to depend on the constant, area dummies of the headquarters, the pre-sample information

[31] This is not surprising, given that the binary dependent variable displays zeros 84 per cent of the time.

on age and the average number of opportunities.

An exhaustive comparison of alternative techniques to estimate a dynamic Probit model with unobserved heterogeneity is offered in Table 7. All estimation methods use the Mundlak-Chamberlain correction term. Our first estimate is a pooled Probit, see column one of the table. The point estimate of the coefficient associated with the lagged dependent variable $y_{i,t-1}$ is positive and statistically significant at the 1 per cent level, providing evidence of the theory of maturation: if an MNE has expanded in the previous period it faces, *ceteris paribus*, a higher probability of expanding in the current period. Such an effect can be either reinforced, reduced or even wiped out, by the sum of the product between the estimated coefficients and their variables, average size $\bar{s}_i^{op}$ (a variable that ranges in the interval of $[0.5, 23.6]$) and lagged size $s_{i,t-1}^{op}$ (a variable that ranges in the interval of $[0, 24]$). The coefficients of $\bar{s}_i^{op}$ and $s_{i,t-1}^{op}$ are close to one another, but have opposite sign. This implies that if a firm is at its early stage of growth, with size below its average long-run level, the overall effect on size growth is positive. As size reaches its average level, the positive boost on growth vanishes, and the effect is reversed after size overtakes that average value. The area dummies are non-significant, suggesting that the physical location of the MNE headquarters gives no comparative advantage on growth. This is an effect that we believe is triggered by the strong deregulation that took place during the *Single Market Programme* era.

Now consider the random effects estimator. The point estimate of the parameter associated with $y_{i,t-1}$ is smaller than that of the pooled Probit model, though the probability distributions of the two estimators are not significantly different from one another at the 5 per cent level. The non-significant discrepancy between the two estimates is explained further if one accounts for the different normalization of the error term, which imposes $\sigma_u^2 = 1$ in the pooled Probit and $\sigma_\epsilon^2 = 1$ in the random effects estimation. The random effects coefficients need to be multiplied by a factor of $\sqrt{1-\rho}$ to be comparable to those of the pooled Probit (see Arulampalam (1999)), with ρ denoting the equicorrelation between the composite error in two time periods.[32] However, in our context ρ is estimated to be very low at $\widehat{\rho} = 0.066$, suggesting that one has a good enough approximation if the effect of the normalization is disregarded; this makes the two sets of results directly comparable.[33] Although not too high, we note that the estimated coefficient $\hat{\rho}$ is significant at the 1 per cent level, which confirms our suspicion that some intertemporal correlation between u_{it} and u_{is}, with $s \neq t$, exists.

To deepen our understanding of the effect of state dependence, we calculate the Average Partial

[32] $\rho \equiv Corr\left(u_{it}, u_{is}\right) = \frac{\sigma_{\bar{\alpha}}^2}{\sigma_{\bar{\alpha}}^2 + \sigma_\epsilon^2}$, for $s, t = 1, 2, ..., T; s \neq t$. As ρ is bounded between 0 and 1, $\sqrt{1-\rho}$ is smaller than 1.

[33] Because $\sqrt{1-\widehat{\rho}}$ is ≈ 1.

Effect (APE) from the counterfactual outcome probabilities that rely, for all i and t, on the two extreme states: complete expansion, $y_{i,t-1} = 1$, denoted with subscript (1), and absence of expansion, $y_{i,t-1} = 0$, denoted with subscript (0). We compute the expected growth in the two states, as follows:

$$
\begin{aligned}
\widehat{y}^{op}_{(1)} &= \frac{1}{M}\sum_{i=1}^{M}\frac{1}{T-\tau_i+1}\sum_{t=\tau_i}^{T}\Phi[(\widehat{\gamma} + \mathbf{x}_{1it}\widehat{\boldsymbol{\beta}}_1 + \mathbf{x}_{2it(1)}\widehat{\boldsymbol{\beta}}_2 + \bar{\mathbf{x}}_{2i(1)}\widehat{\boldsymbol{\lambda}}_1 + \widehat{c}_{it})\sqrt{1-\widehat{\rho}}] \\
\widehat{y}^{op}_{(0)} &= \frac{1}{M}\sum_{i=1}^{M}\frac{1}{T-\tau_i+1}\sum_{t=\tau_i}^{T}\Phi[(\quad \mathbf{x}_{1it}\widehat{\boldsymbol{\beta}}_1 + \mathbf{x}_{2i(0)}\left(\widehat{\boldsymbol{\beta}}_2 + \widehat{\boldsymbol{\lambda}}_1\right) + \widehat{c}_{it})\sqrt{1-\widehat{\rho}}]
\end{aligned}
\tag{14}
$$

where the term $\sqrt{1-\widehat{\rho}}$ is used to make the random effects APE comparable to the pooled Probit - a term that at low values of $\hat{\rho}$ we suggest can be disregarded. The explanatory variables with subscript (1) are defined as $\mathbf{x}_{2it(1)} \equiv \mathbf{x}_{2i0} + t$ and $\bar{\mathbf{x}}_{2i(1)} \equiv \frac{2\mathbf{x}_{2i0}+T-\tau_i}{2}$, respectively. The explanatory variables with subscript (0) are defined as $\mathbf{x}_{2i(0)} \equiv \mathbf{x}_{2it} = \bar{\mathbf{x}}_{2i}$. The APE is computed as the difference between $\widehat{y}^{op}_{(1)}$ and $\widehat{y}^{op}_{(0)}$. The results of the APE are reported in percentages at the bottom of Table 7. The APE of the pooled Probit model is larger than that of the random effects model by 1.6 percentage points, suggesting a mild "spurious" state dependence.

We move forward and set out the results now based on an analysis with correction for the initial conditions problem, starting with Heckman's solution. We run two alternative econometric estimation techniques. One relies on Arulampalam and Stewart (2009) and the other on Stewart (2006). We outline their technicalities in Appendix A.[34] We estimate a simplified initial conditions equation, whose notation is displayed in Eqs. (19) and (20). As mentioned earlier, we use pre-sample age, area dummies and average (lagged) size as instruments for the correlation between the initial conditions and the unobserved heterogeneity. Controlling for average size in the initial conditions equation makes the *additional* exogenous variable, age, non-significant, but due to data limitations we have no alternative available. The implication is that we might incur some bias in integrating the initial conditions out. As for the coefficient of intertemporal equicorrelation, ρ, it is still low and significant at 5 per cent using Arulampalam and Stewart's (2009) procedure, but is nearly doubled and highly significant using Stewart's (2006) procedure. The discrepancy could be attributed to the different integration methodologies adopted by the competing methods. Whatever the reason might be for the difference, we see that about 6 to 12 per cent of the intertemporal correlation is explained by the unobserved heterogeneity. Wooldridge (2005) offers a second approach to solving the initial conditions problem.

[34] Due to a failure of convergence of the estimation algorithm, we do not report or discuss the free correlation cases related to these two estimation techniques.

Again, we confine our interest to the case of equicorrelation. The results are documented in column five. We have a new coefficient that is associated with MNE initial expansion. It has a highly significant positive parameter, hinting that initial expansion produces a persistent long-run effect on absolute growth. The last attempt to solve the initial conditions problem is offered by Orme's (2001) two-step estimator, which is suitable to environments with weak correlation. His methodology brings in a new variable: the inverse Mill's ratio, imr. We start by explaining the equicorrelated estimation that is reported in column six. The inverse Mill's ratio is significant at the 10 per cent, and the remaining parameters are very much in line with the estimates indicated by the previous techniques, so we omit discussing them further. Our last estimation is dedicated to the case of free correlation, i.e. the instance where we only impose $\delta_T = 1$, leaving all other δ_t parameters unconstrained. We study free correlation under Orme's (2001) two-step estimator. Table 7, column seven, reports the estimates of the unconstrained coefficients δ_t, next to those of the time dummies.

We select as the best estimation technique, the one that predicts correctly the highest percentage of expansions. This is Arulampalam and Stewart's (2009) methodology, with an accuracy just above 16 per cent. The results that are going to be discussed in the rest of the section will focus on this estimation.

In addition to the analysis of state dependence discussed above, the effect of size on expansion merits closer examination. We are interested in answering the following questions:

1. Does initial size matter in explaining MNEs' absolute growth?
2. Do MNEs grow along different paths according to their initial sizes?
3. Does Gibrat's law hold?

In order to respond to these questions we assign the MNEs to five size strata. We construct the strata from the MNE's initial size, $s^{op}_{i,90}$.[35] For each stratum, the average expansion is estimated in each period t as:

$$\widehat{y}^{op}_{rt} = \frac{1}{M_{rt}} \sum_{i \in \mathcal{M}_{rt}} \Phi[(\widehat{\gamma} + \mathbf{x}_{1it}\widehat{\boldsymbol{\beta}}_1 + \mathbf{x}_{2it}\widehat{\boldsymbol{\beta}}_2 + \bar{\mathbf{x}}_{2i}\widehat{\boldsymbol{\lambda}}_1 + \widehat{c}_{it})\sqrt{1-\widehat{\rho}}] \tag{15}$$

$$r = NE, 1, 2, 3, 4 \quad t = \tau_i, \tau_i + 1, \cdots, T. \tag{16}$$

[35] Because of the discrete nature of $s^{op}_{i,90}$, each size stratum does not have an equal number of MNEs in it. The first stratum includes the MNEs that have captured only one opportunity up to year 1990 (27.2 per cent of the sample); the second stratum those that have $s^{op}_{i,90} \in \{2, 3\}$ (20 per cent of the sample); the third stratum have $s^{op}_{i,90} \in \{4, 5, 6\}$ (12.8 per cent of the sample); and the fourth stratum with $s^{op}_{i,90} \geq 7$ (14.7 per cent of the sample). A final stratum contains the subsample of new entrants (25.3 per cent of the sample)

where $\mathcal{M}_{rt}$ is the set of MNEs in the rth stratum that are active in period t and M_{rt} is its number of elements.

Figure 3 displays the estimated, and observed, average expansion (and its cumulated version) of the absolute and proportional growth for s^{op} over time and by strata. Sub-figure 3(a) plots the average expected absolute expansion, i.e. the average probability of expansion. The figure sheds light on the first question. If we momentarily exclude the new entrants (NE) from the analysis, we observe at the beginning of the period a clear pattern of positive monotonicity between initial size and absolute growth. For example, by looking at stratum 4, which includes the largest MNEs in year 1990, we note that in the early nineties its MNEs had more than double the average probability of success of MNEs belonging to the second largest stratum, stratum 3. This monotonicity vanishes over time, up to the point that a common converging path arises for all strata after year 2000. By the end of the period, year 2004, the average probabilities of expansion are clustered at a level of approximately 10 per cent. During the period the largest MNEs have lost a good bit of their positive momentum. We have purposely excluded from the discussion the role of new entrants, as these are known to have an atypical pattern that can undermine Gibrat's law, as demonstrated in Simon and Ijiri (1977).[36] From Sub-figure 3(a) we observe that the stratum that includes the new entrants carries, on average, a higher probability of success than stratum 1 and 2, and in certain years even the stratum 3. More importantly, the NE stratum always exhibits a higher probability of expansion than stratum 1, indicating that a portion of new entrants moves away from the lowest part of the distribution over time, a trend that is confirmed in Ijiri and Simon (1964). The fact that all strata converge to a similar value of growth by the end of the time period, can certainly be explained by the Mundlak-Chamberlain correction which, being significant, introduces a mean reversion effect in the growth dynamics.

Sub-figure 3(b) represents the cumulated version of Sub-figure 3(a). The particular trends of expected cumulated expansions that are graphed in the figure, can be used to answer the second question posed above. Here the strong role of initial size is neatly visible. By the end of the period of investigation, year 2004, we find that an MNE from the largest stratum (stratum 4) is, on average, expected to gain about 2.5 new opportunities, versus the 1.5 opportunities of an MNE that belongs to the second largest stratum (stratum 3). A similar comparison could be extended to the remaining strata. Strata with different initial sizes seem to converge to different steady state levels, where the convergence to a steady state here is suggested by the flattening of the paths toward the end of the period. Abstracting

[36] Even though new entry does not play a major role in this industry as highlighted in Table 2.

from time effects that illustrate yearly changes common to all MNEs, the strict concavity in the paths, followed by convergence to a steady state (more marked in the largest stratum (stratum 4)), could be signalling underlying diseconomies of scale, which might be induced by rising managerial costs or by transaction costs in large enterprises.

We compare the predicted average absolute growth plotted in Sub-figure 3(a) with the observed absolute growth graphed in Sub-figure 3(c). While the predicted values emphasize a converging trend common to all strata, the observed data exhibits a diverging drift for the two largest strata (strata 3 and 4) in the year 2003. Due to the scaling of the data this anomaly is less visible in the cumulative values plotted in Sub-figures 3(b),(d).

In order to answer the third question, i.e. to have a proper say on Gibrat's law, we look at growth defined in proportional terms. Sub-figures 3(e)-(f) graph this information for us. We analyze Sub-figure 3(e). Not only is Gibrat's law violated for the stratum of new entrants, a result that confirms the empirical findings of Lotti et al. (2003) and Calvo (2006), but its violation is also extended to the stratum of small MNEs that are part of the balanced part of the panel. A violation that might happen to be worsened if we had data on MNE exit.

We have largely investigated the dynamics of MNEs growth under the metric "number of opportunities". We wish now to push the analysis further. We exploit our estimated Copula and simulate the predicted size values for the remaining metrics - a method that saves us from estimating linear and nonlinear dynamic panel models again and again. We proceed as follows. We utilize the primitives of our chosen dynamic Probit estimation and predict the size measure "cumulated number of opportunities", $\hat{s}^{op}$.[37] We repeat the exercise of drawing 1000 simulations for all periods of time and each metric of size, but this time we condition the simulations to the predicted number of opportunities $\hat{s}^{op}$. Given that we have conditioned the simulations to the estimated value of number of opportunities, rather than the observed number of opportunities, the first column of Table 8 adds a coefficient of determination for the distance between this predicted value and its observed value. The coefficient of determination of $\hat{s}^{op}$ depicts a high explanatory power of the predicted metric, with a modest downtrend over time. Similarly to Table 5, we report the coefficient of determination and the *p-value* of the Pearson χ^2 test of equality of distributions for the two conditionally simulated metrics. The results are similar to those discussed in Table 5, suggesting that it is also true in this case that the simulated

[37] We construct the variable by adding to its initial size $\hat{s}^{op}_{i,\tau_i-1}$ (set to 0 for new entrants) the cumulated unitary expansions that are obtained from tagging an expansion as occurring ($=1$), if the predicted probability is above 0.5, as not occurring ($=0$), otherwise.

distributions are not too far off their corresponding observed distributions. An encouraging result.

We finish this section with an ultimate result on inequalities. For the entire period of investigation, in Table 9 we report the Gini coefficient of concentration for the Lorenz curves of Sutton's lower bound, the observed metrics and the predicted and simulated metrics. The table highlights a few important facts: i) Sutton's lower bound is not satisfied for the size metrics s^{op}, s^{su} (except for years 2002 and 2003) and ls^{or}, but is satisfied for s^{or}; ii) the predicted and simulated metrics generate a pattern comparable to the original metrics; iii) not surprisingly, the log transformation of operational revenues gives the lowest inequality. We plot the Lorenz curves in Figure 4. We only present the figure for year 2003. In Sub-figure 4(a) the 45 degree line exhibits the pattern of an equally-distributed size. The figure confirms that the larger the inequality carried by the metric, the farther the corresponding Lorenz curve drifts away from the 45 degree line. This behavior is confirmed by all our metrics: s^{op} (dashed line), s^{su} (dotted line), s^{or} (short dotted-dashed line) and ls^{or} (dotted-dashed line). Sub-figure 4(b) repeats the plot for the predicted and simulated size metrics, and the concentration pattern is not much different from the one just discussed.

6 Conclusions

Prior to selecting a measure (metric) upon which to study the dynamics of firm growth, this paper undertakes a diligent study of the association between different measures of size. We rely on a Copula approach and a dataset on pharmaceutical multinationals that allows us to construct three measures of size, two discrete and one (quasi-) continuous. Only having three metrics to work with is a limitation dictated by the quality of the data. We have no reliable information that allows us to extend the analysis further. With this restriction in mind, we believe that investigating the relation between alternative metrics is a step forward in the literature on firm growth, and generally speaking, in any literature that utilizes metrics of size. A dataset holding a more exhaustive list of metrics of size could be exploited to put forward a more complete analysis of association among variables. It would be informative to discover that there are metrics with very little, or null, concordance.

The results from the Copula estimation provide evidence of concordance in the behavior of the alternative variates. We base this statement not only on the high values of the estimated concordance coefficients (always above 75 per cent), but also on the ability of the Copula conditioned on one metric, to simulate the remaining metrics - as confirmed by the statistics documented in Table 5. Evidence of strong association in the variates is the "no impediment to" that we needed to freely select the metric that we believed to be the most straightforward measure to employ: the cumulated number of expansions.

With this metric at hand, we proceed further with an analysis which studies the dynamics of the European pharmaceutical industry within a period of enlargement: the Single Market Programme (SMP), 1990-2004. Under appropriate assumptions outlined in the paper, we estimate a dynamic panel random effects Probit model of expansion. In our estimations we did our best to account, on the one hand for the correlation between the unobserved heterogeneity and the regressors, and on the other hand for the correlation between the unobserved heterogeneity and the initial conditions. We present results from alternative estimation techniques. We pick the one that predicts, correctly, the highest number of opportunities (expansions).

Understanding the evolution of the European pharmaceutical industry during the SMP period, is key to learning about the resulting configuration of market structure. Disentangling the MNE size dynamics and testing the theory of firm growth are the essential requirements. Relative to the former, one of the main findings of the paper is that there is a considerable positive relation between past and current expansion, i.e. we find evidence of strong state dependence. In a counterfactual, where we compare the two extreme scenarios of continuous expansion and lack of expansion, we quantify the average partial effect exerted by the state dependence to be about 25 per cent. We read this percentage as a sign of the healthy growth that the European pharmaceutical industry carries through the period. In addition, we uncover that size has a significant mean reversion effect on MNE growth. This effect is confirmed in our analysis of absolute growth by strata. Initial size matters at the beginning of the period of investigation, but over time due to the mean reversion impact, such an initial advantage vanishes. This confirms the theory of maturation, which states that firms face a period of rapid growth, followed by a slow down, or even a stop in growth. We add onto the analysis the role of proportional growth by strata, and find out that new entrants have a higher probability of expansion than the subset of incumbents with low initial size. This finding suggests that Gibrat's law is invalid for firms at the lower end of the size distribution. Ultimately, we find that concentration measures based on the two least heterogeneous size metrics, number of opportunities and number of subsidiaries do not satisfy Sutton's lower bound, while the most heterogenous metric, operational revenue leads to a concentration curve which is well above the lower bound.

At last, it is worth stressing that we have limited the Copula analysis to its static case. We leave to future research a dynamic investigation of alternative metrics of size.

Table 1: Size metrics used in previous studies

Size Measure	Papers
Assets	Hart and Prais (1956), Adelman (1958), Simon and Bonini (1958), Ijiri and Simon (1964), Samuels (1965), Scherer (1965), Quandt (1966), Samuels and Smyth (1968), Singh et al. (1968), Radice (1971), Smyth et al. (1975), Dunne and Hughes (1994), Amaral et al. (1997).
Capital employed	Samuels and Chesher (1972), Chesher (1979).
Capital invested	Hall and Weiss (1967), Smyth et al. (1975).
Employment	Simon and Bonini (1958), Mansfield (1962), Saving (1965), Scherer (1965), Smyth et al. (1975), Evans (1987b), Hall (1987), Pavitt et al. (1987), Dunne et al. (1989), Dunne and Hughes (1994), Variyam and Kraybill (1994), Amaral et al. (1997), Axtell (2001), Cabral and Mata (2003), Lotti et al. (2003), Calvo (2006).
Input	Amaral et al. (1997).
Market valuation	Samuels and Smyth (1968), Singh et al. (1968), Radice (1971), Smyth et al. (1975).
Output	Mansfield (1962).
Plants and equipments	Amaral et al. (1997).
Profit	Hart (1962).
Property	Amaral et al. (1997).
Sales	Hall and Weiss (1967), Samuels and Smyth (1968), Singh et al. (1968), Radice (1971), Smyth et al. (1975), Hart and Oulton (1996), Amaral et al. (1997), Geroski et al. (1997), Sutton (1997b), Bottazzi et al. (2001), Higson et al. (2002), Bottazzi and Secchi (2005), Bottazzi and Secchi (2006), Buldyrev et al. (2007), Cefis et al. (2007), Amisano and Giorgetti (2005).
Turnover	Freeman (1986), Dunne et al. (1988).

Table 2: Statistics of size metrics and absolute growth

Metric	statistics	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
y_{it}^{op}	max	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	mean	(0.21)	(0.20)	(0.16)	(0.14)	(0.22)	(0.17)	(0.20)	(0.17)	(0.17)	(0.17)	(0.18)	(0.16)	(0.11)	(0.09)	(0.06)
	s.d.	[0.41]	[0.40]	[0.37]	[0.35]	[0.42]	[0.38]	[0.40]	[0.38]	[0.38]	[0.37]	[0.39]	[0.37]	[0.31]	[0.29]	[0.25]
	Σ_i	{43}	{42}	{35}	{30}	{50}	{39}	{46}	{41}	{43}	{42}	{47}	{42}	{29}	{25}	{17}
s_{it}^{op}	max	23	23	23	23	23	23	24	24	24	24	24	24	24	24	25
	mean	(3.81)	(3.94)	(3.99)	(4.06)	(4.16)	(4.29)	(4.33)	(4.41)	(4.44)	(4.53)	(4.63)	(4.74)	(4.76)	(4.83)	(4.90)
	s.d.	[4.01]	[4.07]	[4.11]	[4.15]	[4.20]	[4.30]	[4.40]	[4.48]	[4.54]	[4.62]	[4.69]	[4.77]	[4.81]	[4.89]	[4.96]
	Skewness	{2.07}	{2.02}	{1.99}	{1.95}	{1.89}	{1.82}	{1.82}	{1.80}	{1.80}	{1.77}	{1.76}	{1.73}	{1.71}	{1.72}	{1.73}
	$\underset{i}{argmax}$	S.P.[†]	S.P.	S.P.	S.P.	S.P.	S.P.	S.P.	S.P.	S.P.	S.P.	S.P.	S.P.	S.P.	Novartis	Novartis
y_{it}^{su}	max	4	8	3	1	3	4	5	4	6	4	17	7	4	15	15
	mean	(0.27)	(0.27)	(0.17)	(0.12)	(0.25)	(0.22)	(0.27)	(0.22)	(0.22)	(0.20)	(0.33)	(0.24)	(0.14)	(0.16)	(0.14)
	s.d.	[0.68]	[0.81]	[0.48]	[0.33]	[0.56]	[0.58]	[0.72]	[0.62]	[0.67]	[0.53]	[1.45]	[0.72]	[0.49]	[0.99]	[1.01]
	Σ_i	{54}	{56}	{36}	{26}	{54}	{49}	{60}	{52}	{53 }	{49}	{84}	{62}	{35}	{43}	{38}
s_{it}^{su}	max	32	32	32	33	34	34	39	41	47	47	49	49	49	54	55
	mean	(4.86)	(5.05)	(5.11)	(5.17)	(5.28)	(5.46)	(5.58)	(5.71)	(5.78)	(5.90)	(6.13)	(6.31)	(6.34)	(6.49)	(6.63)
	s.d.	[5.64]	[5.84]	[5.94]	[5.98]	[6.10]	[6.34]	[6.62]	[6.84]	[7.04]	[7.17]	[7.72]	[7.88]	[7.95]	[8.31]	[8.58]
	Skewness	{2.05}	{1.98}	{1.97}	{1.99}	{1.99}	{1.98}	{2.12}	{2.14}	{2.33}	{2.31}	{2.51}	{2.48}	{2.46}	{2.69}	{2.70}
	$\underset{i}{argmax}$	Novartis	Novartis	Novartis	Novartis	Novartis	Novartis	Novartis	Novartis	Novartis	Novartis	Novartis	Novartis	Novartis	Pfizer	Pfizer
y_{it}^{or}	max							3,805	19,353	3,602	5,420	3,970	10,569	7,173	15,192	
	mean	na	na	na	na	na	na	(92)	(68)	(79)	(5)	(45)	(113)	(216)	(289)	na
	s.d.							[384]	[1,391]	[323]	[593]	[304]	[798]	[767]	[1,199]	
	Σ_i							{18,304}	{34,641}	{17,134}	{1,045}	{10,466}	{27,281}	{53,624}	{74,344}	
s_{it}^{or}	max						27,991	30,436	34,620	34,737	28,155	28,235	27,211	34,384	38,788	
	mean						(773)	(836)	(960)	(1,017)	(971)	(979)	(1,066)	(1,238)	(1,510)	
	s.d.						[2,528]	[2,780]	[3,260]	[3,317]	[3,071]	[3,123]	[3,320]	[3,883]	[4,827]	
	Skewness	na	na	na	na	na	{0.36}	{0.36}	{0.43}	{0.44}	{0.50}	{0.45}	{0.50}	{0.50}	{0.51}	na
	$\underset{i}{argmax}$						Bayer	Bayer	Bayer	Bayer	Bayer	Bayer	Bayer	Bayer	Bayer	
	#missing						67	60	50	44	32	23	17	8	5	
	# EU	140	141	143	143	148	149	153	154	159	161	162	165	168	169	169
	# US	33	36	38	39	40	40	42	43	43	44	46	46	47	47	47
	# Other	31	31	33	35	36	37	39	42	45	46	48	48	49	49	49
	Total	204	208	214	217	224	226	234	239	247	251	256	259	264	265	265
	#NE	na	4	6	3	7	2	8	5	8	4	5	3	5	1	0
	$\frac{\#NE_t}{\sum_i y_{it}^{op}}$	na	0.10	0.17	0.10	0.14	0.05	0.17	0.12	0.19	0.10	0.11	0.07	0.17	0.04	0.00

† S.P.=Schering Plough.

Geographic coverage: EU 15 including Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden, and the UK (excluding Luxembourg).

Table 3: Fitting distributions with positive support

Distribution	**Param.**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**95-03**
						$\mathbf{s^{op}}$					
	rate	0.233	0.231	0.227	0.225	0.221	0.216	0.211	0.210	0.207	0.219
Exponential	**s.e.**	(0.016)	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)	(0.013)	(0.013)	(0.013)	(0.005)
	$\mathbf{p-val\ \chi^2}$	[0.168]	[0.111]	[0.097]	[0.043]	[0.038]	[0.015]	[0.008]	[0.005]	[0.006]	[0.000]
	prob.	0.189	0.188	0.185	0.184	0.181	0.178	0.174	0.174	0.171	0.180
Geometric†	**s.e.**	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.003)
	$\mathbf{p-val\ \chi^2}$	[0.989]	[0.938]	[0.789]	[0.648]	[0.685]	[0.573]	[0.517]	[0.535]	[0.496]	[0.013]
	size	1.722	1.658	1.632	1.617	1.622	1.628	1.636	1.611	1.592	1.628
	s.e.	(0.218)	(0.203)	(0.196)	(0.190)	(0.188)	(0.186)	(0.185)	(0.179)	(0.176)	(0.063)
Negative-Binomial	$\boldsymbol{\mu}$	4.283	4.329	4.406	4.437	4.534	4.628	4.737	4.757	4.834	4.559
	s.e.	(0.257)	(0.258)	(0.261)	(0.259)	(0.262)	(0.264)	(0.267)	(0.267)	(0.271)	(0.088)
	$\mathbf{p-val\ \chi^2}$	[0.024]	[0.011]	[0.018]	[0.009]	[0.006]	[0.002]	[0.001]	[0.001]	[0.001]	[0.000]
	$\boldsymbol{\lambda}$	4.283	4.329	4.406	4.437	4.534	4.629	4.737	4.758	4.834	4.559
Poisson	**s.e.**	(0.138)	(0.136)	(0.136)	(0.134)	(0.134)	(0.134)	(0.135)	(0.134)	(0.135)	(0.045)
	$\mathbf{p-val\ \chi^2}$	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
						$\mathbf{s^{su}}$					
	rate	0.183	0.179	0.175	0.173	0.170	0.163	0.158	0.158	0.154	0.167
Exponential	**s.e.**	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)	(0.009)	(0.004)
	$\mathbf{p-val\ \chi^2}$	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	prob.	0.155	0.152	0.149	0.148	0.145	0.140	0.137	0.136	0.134	0.143
Geometric†	**s.e.**	(0.010)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.003)
	$\mathbf{p-val\ \chi^2}$	[0.046]	[0.072]	[0.110]	[0.133]	[0.086]	[0.029]	[0.065]	[0.071]	[0.027]	[0.000]
	size	1.213	1.173	1.157	1.146	1.151	1.118	1.111	1.094	1.074	1.130
	s.e.	(0.131)	(0.123)	(0.119)	(0.115)	(0.114)	(0.108)	(0.107)	(0.104)	(0.101)	(0.037)
Negative-Binomial	$\boldsymbol{\mu}$	5.460	5.577	5.707	5.777	5.896	6.133	6.313	6.345	6.487	5.983
	s.e.	(0.365)	(0.371)	(0.376)	(0.376)	(0.379)	(0.394)	(0.403)	(0.404)	(0.415)	(0.130)
	$\mathbf{p-val\ \chi^2}$	[0.001]	[0.005]	[0.007]	[0.008]	[0.004]	[0.002]	[0.004]	[0.009]	[0.003]	[0.000]
	$\boldsymbol{\lambda}$	5.460	5.577	5.707	5.777	5.896	6.133	6.313	6.345	6.487	5.983
Poisson	**s.e.**	(0.155)	(0.154)	(0.155)	(0.153)	(0.153)	(0.155)	(0.156)	(0.155)	(0.156)	(0.052)
	$\mathbf{p-val\ \chi^2}$	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
						$\mathbf{ls^{or*}}$					
	mean	4.575	4.618	4.696	4.785	4.726	4.664	4.734	4.897	5.120	4.768
	s.e.	(0.143)	(0.141)	(0.137)	(0.135)	(0.131)	(0.132)	(0.129)	(0.127)	(0.124)	(0.044)
Normal	**s.d.**	2.007	2.014	2.009	2.001	2.007	2.051	2.039	2.031	2.001	2.025
	s.e.	(0.101)	(0.099)	(0.097)	(0.095)	(0.093)	(0.093)	(0.092)	(0.090)	(0.088)	(0.031)
	$\mathbf{p-val\ \chi^2}$	[0.006]	[0.002]	[0.001]	[0.001]	[0.001]	[0.000]	[0.000]	[0.012]	[0.002]	[0.000]
	rate	0.219	0.217	0.213	0.209	0.212	0.214	0.211	0.204	0.195	0.210
Exponential	**s.e.**	(0.016)	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)	(0.013)	(0.013)	(0.012)	(0.005)
	$\mathbf{p-val\ \chi^2}$	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	shape	3.889	3.883	4.283	4.778	4.919	4.177	4.390	4.646	5.197	4.428
	s.e.	(0.375)	(0.368)	(0.398)	(0.440)	(0.441)	(0.366)	(0.380)	(0.396)	(0.442)	(0.132)
Gamma	**rate**	0.850	0.841	0.912	0.998	1.041	0.895	0.927	0.949	1.015	0.929
	s.e.	(0.088)	(0.085)	(0.090)	(0.097)	(0.098)	(0.083)	(0.085)	(0.085)	(0.091)	(0.029)
	$\mathbf{p-val\ \chi^2}$	[0.041]	[0.033]	[0.105]	[0.079]	[0.112]	[0.028]	[0.040]	[0.196]	[0.365]	[0.000]
	shape	2.378	2.386	2.451	2.535	2.515	2.391	2.438	2.526	2.675	2.475
	s.e.	(0.133)	(0.131)	(0.130)	(0.132)	(0.127)	(0.120)	(0.120)	(0.121)	(0.127)	(0.042)
Weibull†	**scale**	5.139	5.183	5.277	5.384	5.329	5.249	5.324	5.499	5.735	5.360
	s.e.	(0.161)	(0.159)	(0.154)	(0.150)	(0.146)	(0.148)	(0.146)	(0.143)	(0.140)	(0.050)
	$\mathbf{p-val\ \chi^2}$	[0.152]	[0.105]	[0.114]	[0.080]	[0.090]	[0.031]	[0.020]	[0.374]	[0.116]	[0.000]

† Selected distribution based on the Pearson χ^2 test.

* Due to a lack of data the variable presents, in subsequent period, the following number of missing values: 67, 60, 50, 44, 32, 23, 17, 8 and 5.

Table 4: Fitting Copulas

Copula		**Marginal parameter**				**Concordance coefficient**			**Copula parameter**	**Log-likelihood**
		prob$_{op}$	**prob**$_{su}$	**shape**$_{or}$	**scale**$_{or}$	$\varrho_{op,su}$	$\varrho_{op,or}$	$\varrho_{su,or}$	ϱ	**ll**
						1995				
Elliptical	**Normal**†	0.204 (0.006)	0.172 (0.005)	2.078 (0.044)	4.654 (0.071)	0.929 (0.003)	0.780 (0.011)	0.787 (0.011)		-6,108
Archimedean	**Clayton**	0.204 (0.005)	0.168 (0.004)	1.669 (0.032)	4.549 (0.073)				2.681 (0.090)	-6,678
	Frank	0.193 (0.005)	0.162 (0.004)	2.198 (0.056)	4.944 (0.071)				2.697 (0.154)	-7,306
	Gumbel	0.194 (0.005)	0.165 (0.005)	2.062 (0.035)	4.655 (0.071)				2.023 (0.050)	-6,590
						1996				
Elliptical	**Normal**†	0.193 (0.006)	0.162 (0.005)	2.039 (0.042)	4.646 (0.072)	0.925 (0.004)	0.789 (0.011)	0.780 (0.011)		-6,226
Archimedean	**Clayton**	0.204 (0.005)	0.182 (0.005)	1.706 (0.036)	4.551 (0.072)				2.574 (0.092)	-6,644
	Frank	0.192 (0.005)	0.160 (0.004)	2.196 (0.056)	4.977 (0.072)				2.404 (0.152)	-7,360
	Gumbel	0.191 (0.005)	0.160 (0.004)	2.128 (0.035)	4.814 (0.070)				2.113 (0.052)	-6,609
						1997				
Elliptical	**Normal**†	0.186 (0.005)	0.157 (0.005)	2.088 (0.042)	4.855 (0.073)	0.934 (0.003)	0.805 (0.010)	0.802 (0.010)		-6,229
Archimedean	**Clayton**	0.202 (0.005)	0.164 (0.004)	1.821 (0.036)	4.676 (0.069)				2.528 (0.088)	-6,720
	Frank	0.190 (0.005)	0.151 (0.004)	2.254 (0.058)	5.013 (0.071)				2.351 (0.149)	-7,444
	Gumbel	0.196 (0.005)	0.164 (0.005)	2.188 (0.041)	4.786 (0.068)				1.970 (0.048)	-6,642
						1998				
Elliptical	**Normal**†	0.188 (0.005)	0.158 (0.005)	2.112 (0.043)	4.924 (0.074)	0.926 (0.004)	0.797 (0.010)	0.798 (0.010)		-6,295
Archimedean	**Clayton**	0.201 (0.005)	0.171 (0.004)	1.758 (0.034)	4.827 (0.074)				2.623 (0.090)	-6,708
	Frank	0.186 (0.005)	0.160 (0.004)	2.464 (0.062)	5.151 (0.066)				2.482 (0.152)	-7,329
	Gumbel	0.201 (0.005)	0.163 (0.004)	2.259 (0.039)	4.860 (0.067)				2.025 (0.050)	-6,580
						1999				
Elliptical	**Normal**†	0.199 (0.006)	0.162 (0.005)	2.555 (0.046)	4.863 (0.069)	0.930 (0.003)	0.782 (0.011)	0.769 (0.011)		-6,145
Archimedean	**Clayton**	0.196 (0.005)	0.164 (0.004)	1.783 (0.036)	4.743 (0.071)				2.407 (0.084)	-6,788
	Frank	0.197 (0.005)	0.152 (0.004)	2.268 (0.058)	5.061 (0.071)				2.319 (0.147)	-7,370
	Gumbel	0.183 (0.005)	0.148 (0.004)	2.275 (0.040)	5.027 (0.069)				2.053 (0.050)	-6,758

Table 4: Fitting Copulas (cont.)

Copula		**Marginal parameter**				**Concordance coefficient**			**Copula parameter**	**Log-likelihood**
		prob$_{op}$	**prob**$_{su}$	**shape**$_{or}$	**scale**$_{or}$	$\varrho_{op,su}$	$\varrho_{op,or}$	$\varrho_{su,or}$	ϱ	**ll**
						2000				
Elliptical	**Normal**†	0.197 (0.006)	0.159 (0.005)	2.125 (0.044)	4.711 (0.070)	0.902 (0.005)	0.788 (0.011)	0.779 (0.011)		-6,329
Archimedean	**Clayton**	0.192 (0.005)	0.158 (0.004)	1.739 (0.035)	4.636 (0.072)				2.382 (0.085)	-6,899
	Frank	0.182 (0.005)	0.147 (0.004)	2.272 (0.057)	5.064 (0.070)				2.562 (0.148)	-7,490
	Gumbel	0.186 (0.005)	0.147 (0.004)	2.118 (0.036)	4.867 (0.072)				2.055 (0.050)	-6,779
						2001				
Elliptical	**Normal**†	0.183 (0.005)	0.147 (0.004)	2.126 (0.041)	4.877 (0.073)	0.911 (0.004)	0.813 (0.010)	0.807 (0.010)		-6,404
Archimedean	**Clayton**	0.186 (0.005)	0.156 (0.004)	1.780 (0.034)	4.744 (0.071)				2.475 (0.084)	-6,865
	Frank	0.180 (0.005)	0.136 (0.004)	2.277 (0.058)	5.113 (0.071)				2.505 (0.152)	-7,581
	Gumbel	0.178 (0.005)	0.148 (0.004)	2.151 (0.038)	4.890 (0.071)				2.052 (0.050)	-6,806
						2002				
Elliptical	**Normal**†	0.183 (0.005)	0.146 (0.004)	2.195 (0.043)	5.101 (0.073)	0.920 (0.004)	0.822 (0.009)	0.802 (0.010)		-6,402
Archimedean	**Clayton**	0.186 (0.005)	0.148 (0.004)	1.868 (0.036)	5.024 (0.072)				2.519 (0.087)	-6,976
	Frank	0.181 (0.005)	0.140 (0.004)	2.408 (0.061)	5.279 (0.070)				2.180 (0.147)	-7,553
	Gumbel	0.180 (0.005)	0.143 (0.004)	2.343 (0.040)	5.187 (0.069)				2.057 (0.050)	-6,847
						2003				
Elliptical	**Normal**†	0.177 (0.005)	0.144 (0.004)	2.278 (0.043)	5.263 (0.073)	0.893 (0.005)	0.833 (0.009)	0.800 (0.010)		-6,525
Archimedean	**Clayton**	0.179 (0.004)	0.147 (0.004)	2.006 (0.040)	5.263 (0.070)				2.407 (0.085)	-7,054
	Frank	0.173 (0.005)	0.142 (0.004)	2.507 (0.063)	5.479 (0.070)				2.329 (0.154)	-7,561
	Gumbel	0.180 (0.005)	0.145 (0.004)	4.463 (0.042)	5.395 (0.068)				2.016 (0.049)	-6,848

† Selected Copula based on the log-likelihood value reported in the last column.

Table 5: Coefficient of determination and Pearson's χ^2 test of equality of distributions

Year	$\tilde{\mathbf{s}}^{\mathbf{su}\dagger}$	$\tilde{\mathbf{ls}}^{\mathbf{or}\dagger}$	$\tilde{\mathbf{s}}^{\mathbf{su}\dagger}$	$\tilde{\mathbf{ls}}^{\mathbf{or}\dagger}$
	$\mathbf{R^2}$		$\mathbf{p-value}\ \chi^2$	
1995	0.778 [0.771,0.786]	0.597 [0.593,0.601]	0.211 [0.001,0.853]	0.072 [0.000,0.579]
1996	0.773 [0.765,0.783]	0.596 [0.591,0.600]	0.304 [0.003,0.930]	0.043 [0.000,0.432]
1997	0.778 [0.770,0.787]	0.605 [0.602,0.608]	0.329 [0.008,0.913]	0.012 [0.000,0.127]
1998	0.770 [0.761,0.781]	0.584 [0.582,0.587]	0.362 [0.006,0.942]	0.013 [0.000,0.158]
1999	0.779 [0.771,0.789]	0.600 [0.597,0.602]	0.291 [0.005,0.832]	0.049 [0.000,0.451]
2000	0.726 [0.716,0.737]	0.607 [0.602,0.611]	0.122 [0.000,0.631]	0.038 [0.000,0.350]
2001	0.726 [0.716,0.737]	0.628 [0.623,0.632]	0.143 [0.000,0.684]	0.030 [0.000,0.277]
2002	0.732 [0.722,0.741]	0.653 [0.652,0.655]	0.147 [0.000,0.696]	0.037 [0.000,0.378]
2003	0.703 [0.693,0.715]	0.661 [0.657,0.663]	0.066 [0.000,0.456]	0.029 [0.000,0.289]

In square brackets 95 per cent empirical confidence interval.
† Simulations from Copula, conditioning on s^{op}.

Table 6: Dynamic panel Probit estimation. Dependent variable y_{it}. Impact of Mundlak-Chamberlain (MC) correction.

Variable	**Pooled**† no MC correction	**Random Effect**†† no MC correction	**Pooled**† MC correction	**Random Effect**†† MC correction
$Const\ [\beta_{1,0}]$	-1.755***	-1.804***	-1.339***	-1.371***
	(0.151)	(0.162)	(0.144)	(0.147)
$y_{i,t-1}\ [\gamma]$	0.369***	0.144*	0.571***	0.456***
	(0.065)	(0.079)	(0.070)	(0.083)
$s^{op}_{i,t-1}\ [\beta_{2,1}]$	0.039***	0.032***	-0.538***	-0.530***
	(0.008)	(0.009)	(0.074)	(0.040)
$\bar{s}^{op}_i\ [\lambda_{1,1}]$			0.594***	0.591***
			(0.072)	(0.041)
$eu_i\ [\beta_{1,1}]$	-0.112	-0.137	-0.096	-0.103
	(0.079)	(0.103)	(0.073)	(0.089)
$us_i\ [\beta_{1,2}]$	0.068	0.076	0.062	0.058
	(0.107)	(0.128)	(0.103)	(0.109)
Year dummies	yes	yes	yes	yes
ρ		0.149***		0.066***
		(0.033)		(0.025)
Statistics				
APE of $y_{i,t-1}$	17.3	8.5	34.4	32.8
Percentage of positive predicted outcomes	0.96	0.00	14.2	12.1
Log-likelihood	-1370.2	-1351.7	-1242.3	-1237.6
BIC	2886.5	2857.7	2638.9	2637.6
N. of observations	3369	3369	3369	3369

† Clustered standard errors in brackets.
†† Robust standard errors in brackets, computed using Fisher's information matrix.

Table 7: Dynamic panel Probit estimation. Dependent variable y_{it}

Variable	Pooled†	Random Effect††	Arulampalam & Stewart EC††	Stewart EC††	Wooldridge EC††	Orme EC††	Orme FC††,†††
Main equation							
$Const\ [\beta_{1,0}]$	-1.339***	-1.371***	-1.279***	-1.500***	-1.464***	-1.020***	-0.985***
	(0.144)	(0.147)	(0.138)	(0.186)	(0.152)	(0.149)	(0.194)
$y_{i,t-1}\ [\gamma]$	0.571***	0.456***	0.462***	0.474***	0.415***	0.420***	0.422***
	(0.070)	(0.083)	(0.077)	(0.093)	(0.083)	(0.083)	(0.084)
$eu_i\ [\beta_{1,1}]$	-0.096	-0.103	-0.093	-0.106	-0.092	-0.094	-0.092
	(0.073)	(0.089)	(0.078)	(0.118)	(0.090)	(0.089)	(0.090)
$us_i\ [\beta_{1,2}]$	0.062	0.058	0.060	0.046	0.046	0.015	0.016
	(0.103)	(0.109)	(0.107)	(0.145)	(0.111)	(0.109)	(0.111)
$s^{op}_{i,t-1}\ [\beta_{2,1}]$	-0.538***	-0.530***	-0.549***	-0.435***	-0.512***	-0.520***	-0.520***
	(0.074)	(0.040)	(0.075)	(0.043)	(0.040)	(0.040)	(0.041)
$\bar{s}^{op}_i\ [\lambda_{1,1}]$	0.594***	0.591***	0.610***	0.510***	0.571***	0.572***	0.572***
	(0.072)	(0.041)	(0.073)	(0.044)	(0.041)	(0.041)	(0.042)
$y_{i,90}\ [\mu_0]$					0.214***		
					(0.074)		
Initial condition equation (1990)							
$Const\ [\eta_1]$			-2.246***	-2.266***		-2.219***	-2.219***
			(0.432)	(0.421)		(0.408)	(0.408)
$eu_i\ [\eta_2]$			0.469	0.467		0.461	0.461
			(0.404)	(0.392)		(0.389)	(0.389)
$us_i\ [\eta_3]$			0.433	0.432		0.433	0.441
			(0.506)	(0.445)		(0.433)	(0.441)
$age_{i,89}\ [\eta_4]$			-0.004	-0.004		-0.004	-0.004
			(0.005)	(0.005)		(0.005)	(0.005)
$\bar{s}^{op}_i\ [\lambda_{0,1}]$			0.178***	0.180***		0.178***	0.178***
			(0.033)	(0.033)		(0.032)	(0.032)
$imr\ \alpha_i\ [\delta_0]$			1.483	0.375		-0.207***	-0.244*
			(0.902)	(0.518)		(0.044)	(0.133)
$t_{91}\ [\beta_{1,3}]\ \|\ \alpha_i\ [\delta_1]$	-0.679***	-0.675***	-0.804***	-0.614***	-0.610***	-0.672***	-1.023*** | 0.272
	(0.186)	(0.198)	(0.192)	(0.223)	(0.200)	(0.194)	(0.355) | (0.225)
$t_{92}\ [\beta_{1,4}]\ \|\ \alpha_i\ [\delta_2]$	-0.707***	-0.719***	-0.843***	-0.783***	-0.665**	-0.729***	-0.642*** | -0.068
	(0.182)	(0.195)	(0.187)	(0.225)	(0.197)	(0.191)	(0.324) | (0.222)
$t_{93}\ [\beta_{1,5}]\ \|\ \alpha_i\ [\delta_3]$	-0.710***	-0.735***	-0.855***	-0.776***	-0.688***	-0.752***	-0.823** | 0.063
	(0.174)	(0.195)	(0.177)	(0.225)	(0.197)	(0.191)	(0.323) | (0.223)
$t_{94}\ [\beta_{1,6}]\ \|\ \alpha_i\ [\delta_4]$	-0.182	-0.184	-0.298*	-0.237	-0.143	-0.206	-0.091 | -0.086
	(0.159)	(0.176)	(0.157)	(0.204)	(0.177)	(0.172)	(0.285) | (0.197)
$t_{95}\ [\beta_{1,7}]\ \|\ \alpha_i\ [\delta_5]$	-0.326**	-0.329*	-0.439***	-0.349*	-0.288	-0.348**	-0.274 | -0.055
	(0.147)	(0.178)	(0.146)	(0.205)	(0.180)	(0.174)	(0.291) | (0.205)
$t_{96}\ [\beta_{1,8}]\ \|\ \alpha_i\ [\delta_6]$	-0.067	-0.066	-0.171	-0.157	-0.034	-0.098	0.132 | -0.190
	(0.148)	(0.168)	(0.146)	(0.196)	(0.169)	(0.164)	(0.265) | (0.191)
$t_{97}\ [\beta_{1,9}]\ \|\ \alpha_i\ [\delta_7]$	-0.077	-0.071	-0.173	-0.109	-0.046	-0.108	-0.094 | -0.007
	(0.149)	(0.167)	(0.144)	(0.193)	(0.168)	(0.162)	(0.263) | (0.189)
$t_{98}\ [\beta_{1,10}]\ \|\ \alpha_i\ [\delta_8]$	0.070	0.074	-0.023	-0.025	0.093	0.026	0.068 | -0.032
	(0.148)	(0.162)	(0.143)	(0.190)	(0.163)	(0.157)	(0.249) | (0.181)
$t_{99}\ [\beta_{1,11}]\ \|\ \alpha_i\ [\delta_9]$	0.134	0.140	0.046	0.164	0.158	0.093	-0.099 | 0.166
	(0.144)	(0.160)	(0.139)	(0.184)	(0.161)	(0.155)	(0.251) | (0.179)
$t_{00}\ [\beta_{1,12}]\ \|\ \alpha_i\ [\delta_{10}]$	0.295**	0.303*	0.212	0.277	0.315**	0.250*	0.176 | 0.067
	(0.145)	(0.157)	(0.137)	(0.181)	(0.157)	(0.152)	(0.241) | (0.175)
$t_{01}\ [\beta_{1,13}]\ \|\ \alpha_i\ [\delta_{11}]$	0.312**	0.322**	0.234*	0.315*	0.328**	0.265*	0.064 | 0.175
	(0.145)	(0.156)	(0.139)	(0.180)	(0.157)	(0.151)	(0.243) | (0.174)
$t_{02}\ [\beta_{1,14}]\ \|\ \alpha_i\ [\delta_{12}]$	0.164	0.163	0.079	0.082	0.164	0.092	-0.014 | 0.098
	(0.151)	(0.161)	(0.145)	(0.189)	(0.162)	(0.157)	(0.246) | (0.179)
$t_{03}\ [\beta_{1,15}]\ \|\ \alpha_i\ [\delta_{13}]$	0.164	0.161	0.079	0.159	0.158	0.085	0.059 | 0.026
	(0.152)	(0.163)	(0.145)	(0.188)	(0.164)	(0.159)	(0.245) | (0.183)
ρ		0.066***	0.064**	0.116***	0.072***	0.063***	0.067***
		(0.025)	(0.017)	(0.033)	(0.025)	(0.024)	(0.025)
APE of $y_{i,t-1}$	34.4	32.8	25.0	39.3	31.6	30.9	30.4
Percentage of positive predicted outcomes	14.2	12.1	16.7	13.9	12.5	13.7	14.6
Statistics							
Log-likelihood	-1242.3	-1237.6	-1312.5	-1046.0	-1233.1	-1232.4	-1227.8
BIC	2638.9	2637.6	2837.7	2304.8	2636.8	2635.5	2731.8
N. of the main equation	3369	3369	3375	3375	3369	3375	3375
N. of the initial equation	na	na	198	198	na	198	198

† Clustered standard errors in brackets.
†† Robust standard errors in brackets, computed using Fisher's information matrix.
††† Free correlation parameters and robust standard errors after the slash punctuation.

Table 8: Coefficient of determination and Pearson's χ^2 test of equality of distributions

Year	$\hat{\mathbf{s}}^{\mathbf{op}}$	$\tilde{\tilde{\mathbf{s}}}^{\mathbf{su}}\dagger$	$\tilde{\tilde{\mathbf{ls}}}^{\mathbf{or}}\dagger$	$\tilde{\tilde{\mathbf{s}}}^{\mathbf{su}}\dagger$	$\tilde{\tilde{\mathbf{ls}}}^{\mathbf{or}}\dagger$
		$\mathbf{R^2}$		$\mathbf{p-value}\ \chi^2$	
1995	0.951 [0.950,0.952]	0.764 [0.757,0.772]	0.575 [0.570,0.580]	0.366 [0.003,0.954]	0.234 [0.000,0.938]
1996	0.936 [0.935,0.937]	0.733 [0.725,0.743]	0.574 [0.567,0.580]	0.359 [0.003,0.915]	0.181 [0.000,0.861]
1997	0.930 [0.929,0.931]	0.723 [0.713,0.732]	0.559 [0.548,0.569]	0.320 [0.003,0.862]	0.070 [0.000,0.555]
1998	0.920 [0.919,0.921]	0.704 [0.694,0.716]	0.536 [0.523,0.549]	0.274 [0.001,0.804]	0.070 [0.000,0.640]
1999	0.910 [0.908,0.912]	0.693 [0.682,0.705]	0.523 [0.508,0.538]	0.201 [0.000,0.731]	0.094 [0.000,0.644]
2000	0.897 [0.895,0.899]	0.614 [0.601,0.629]	0.507 [0.492,0.522]	0.302 [0.002,0.889]	0.055 [0.000,0.449]
2001	0.881 [0.879,0.884]	0.589 [0.574,0.604]	0.496 [0.480,0.512]	0.138 [0.000,0.611]	0.042 [0.000,0.359]
2002	0.871 [0.869,0.874]	0.578 [0.566,0.593]	0.473 [0.454,0.493]	0.072 [0.000,0.435]	0.042 [0.000,0.389]
2003	0.863 [0.860,0.865]	0.541 [0.525,0.558]	0.449 [0.427,0.473]	0.095 [0.000,0.523]	0.043 [0.000,0.363]

In square brackets 95 per cent empirical confidence interval.

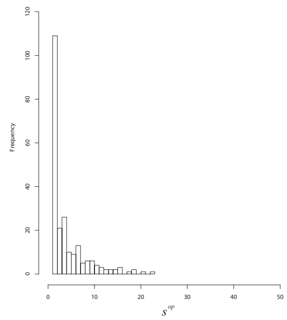
† Simulations from Copula, conditional to the estimated $\hat{s}^{op}$.

Table 9: Gini Coefficient

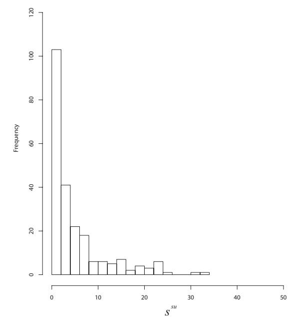
Year	Sutton's lb	$\mathbf{s}^{\mathbf{op}}$	$\mathbf{s}^{\mathbf{su}}$	$\mathbf{ls}^{\mathbf{or}}$	$\mathbf{s}^{\mathbf{or}}$	$\hat{\mathbf{s}}^{\mathbf{op}}$	$\tilde{\tilde{\mathbf{s}}}^{\mathbf{su}}\dagger$	$\tilde{\tilde{\mathbf{ls}}}^{\mathbf{or}}\dagger$	$\tilde{\tilde{\mathbf{s}}}^{\mathbf{or}}\dagger$
1995	0.556	0.473	0.528	0.247	0.846	0.521	0.475	0.158	0.771
1996	0.556	0.475	0.533	0.245	0.851	0.528	0.473	0.160	0.790
1997	0.556	0.481	0.540	0.240	0.859	0.536	0.484	0.157	0.794
1998	0.556	0.481	0.540	0.235	0.854	0.540	0.481	0.150	0.795
1999	0.556	0.488	0.547	0.239	0.855	0.544	0.487	0.148	0.795
2000	0.556	0.488	0.554	0.248	0.860	0.547	0.472	0.156	0.819
2001	0.556	0.487	0.554	0.242	0.862	0.548	0.476	0.158	0.830
2002	0.556	0.493	0.560	0.233	0.861	0.550	0.483	0.150	0.825
2003	0.556	0.496	0.564	0.218	0.862	0.550	0.462	0.143	0.821

† Simulations from Copula, conditional to the estimated $\hat{s}^{op}$.

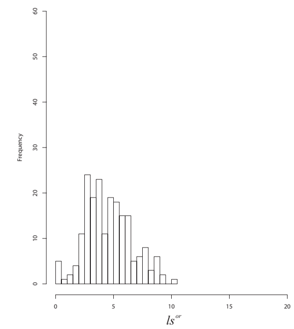
Figure 1: Histograms



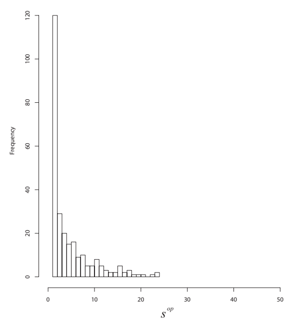
(a) Year 1995



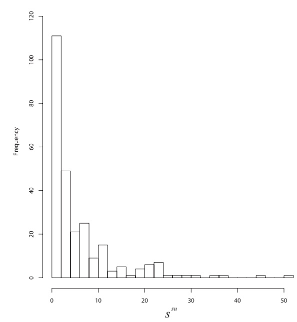
(b) Year 1995



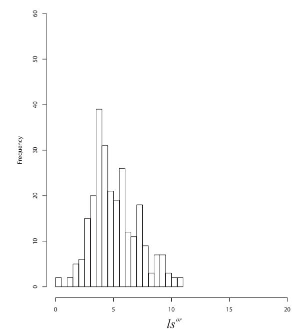
(c) Year 1995



(d) Year 2003

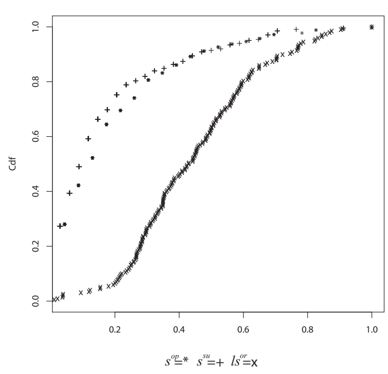


(e) Year 2003

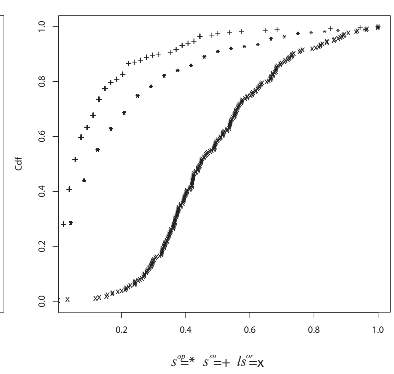


(f) Year 2003

Figure 2: Empirical cumulative density function for normalized measures of size

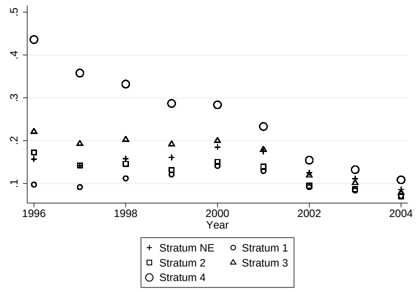


(a) Year 1995

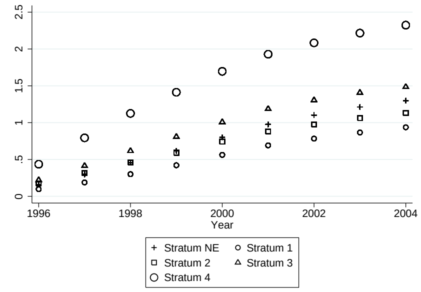


(b) Year 2003

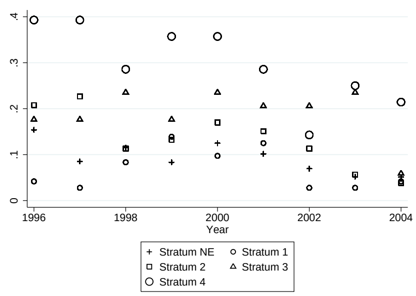
Figure 3: Absolute and proportional growth of estimated and observed number of opportunities



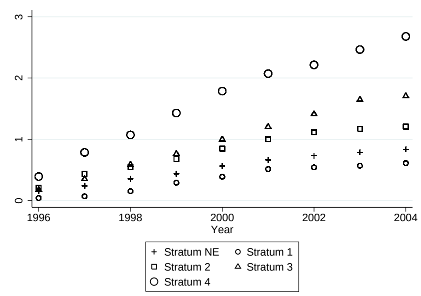
(a) $\hat{s}^{op}$: Avg. abs. growth



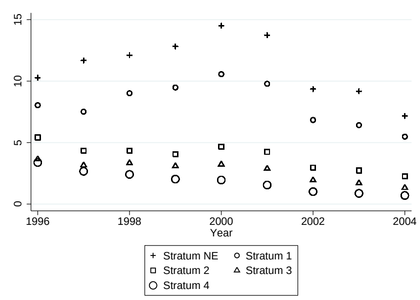
(b) $\hat{s}^{op}$: Cum. avg. abs. growth



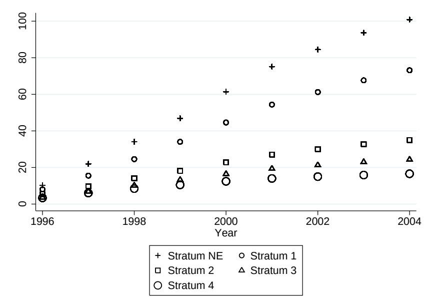
(c) s^{op} : Avg. abs. growth



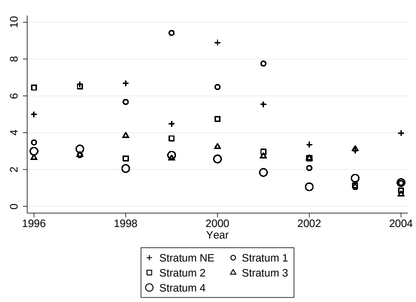
(d) s^{op} : Cum. avg. abs. growth



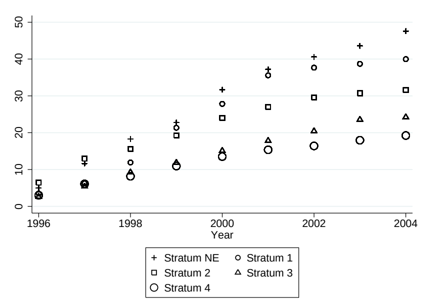
(e) $\hat{s}^{op}$: Avg. prop. growth (%)



(f) $\hat{s}^{op}$: Cum. avg. prop. growth (%)

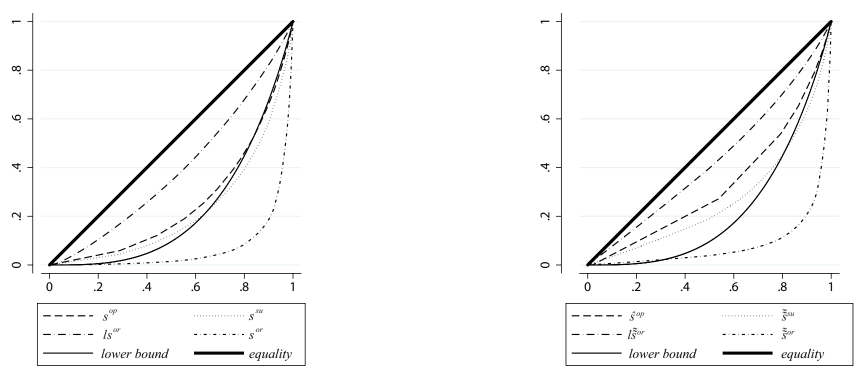


(g) s^{op} : Avg. prop. growth (%)



(h) s^{op} : Cum. avg. prop. growth (%)

Figure 4: Lorenz concentration curves (year=2003)



(a) Actual size metrics

(b) Predicted or simulated size metrics

A Solutions to the Initial Conditions Problem

A.1 Heckman's Procedure

Heckman (1981a,b) suggests a way to account for the initial conditions and integrate the unobserved heterogeneity out of the density function displayed in Eq. (11)

$$P\left(y_{i0}, y_{i1}, \cdots, y_{iT} | \mathbf{X}_i; \boldsymbol{\theta}\right) = \int G\left(y_{i0} | \mathbf{X}_i, \alpha\right) \prod_{t=1}^{T} \Phi\left[\left(\mathbf{z}_{it}\boldsymbol{\vartheta} + c_{it}\right)\left(2y_{it} - 1\right)\right] h\left(\alpha | \mathbf{X}_i\right) d\alpha, \tag{17}$$

where $G(y_{i0}|\mathbf{X}_i, \alpha)$ is the conditional distribution of the initial value of the dependent variable, and $h(\alpha|\mathbf{X}_i)$ is the conditional distribution of the unobserved heterogeneity. For the sub-sample of MNEs that enter in the industry at time $\tau_i > 1$, the new entrants, Eq. (17) simplifies to

$$P\left(y_{i\tau_i}, \cdots, y_{iT} | y_{i,\tau_i - 1} = 0, \mathbf{X}_i; \boldsymbol{\theta}\right) = \int \prod_{t=\tau_i}^{T} \Phi\left[\left(\mathbf{z}_{it}\boldsymbol{\vartheta} + c_{it}\right)\left(2y_{it} - 1\right)\right] h\left(\alpha | \mathbf{X}_i\right) d\alpha. \tag{18}$$

The remainder of the section concentrates on the balanced panel, as it is this one that suffers the initial conditions problem.

We recall from Section 3 the time varying function of unobserved heterogeneity $c_{it} = \delta_t \tilde{\alpha}_i$ and assume $\tilde{\alpha}_i | \mathbf{X}_i \sim NID(0, \sigma^2_{\tilde{\alpha}})$.[38] Next, we formulate the underlying profit function for the initial period as

$$\pi_{i0} = \mathbf{w}_i \boldsymbol{\eta} + \bar{\mathbf{x}}_{2i} \boldsymbol{\lambda_0} + u_{i0}, \tag{19}$$

where $\mathbf{w}_i$ as suggested in Heckman (1981a,b) includes a constant and any exogenous pre-sample regressor. The initial period composite error u_{i0}, consists of

$$u_{i0} = \delta_0 \tilde{\alpha}_i + \epsilon_{i0}, \tag{20}$$

with $\epsilon_{i0} \sim N(0, 1)$ assumed to be orthogonal to $\tilde{\alpha}_i$, as well as to $\mathbf{X}_i$ and $\mathbf{w}_i$.

The initial conditions problem arises because of the correlation between u_{i0} and $\tilde{\alpha}_i$, given $\mathbf{X}_i$ and $\mathbf{w}_i$. In order to grasp the problem, we rewrite the initial profit function of Eq. (19) as

$$\tilde{\pi}_{i0} = \mathbf{w}_i \tilde{\boldsymbol{\eta}} + \bar{\mathbf{x}}_{2i} \tilde{\boldsymbol{\lambda}}_0 + \tilde{\alpha}_i + \tilde{\epsilon}_{i0}, \tag{21}$$

[38] So that $(c_{it}|\mathbf{X}_i) \sim NID(0, \delta^2_t \sigma^2_{\tilde{\alpha}})$ has a variance that is allowed to vary over time.

and specify the joint distribution for $(u_{i0}, \tilde{\alpha}_i | \mathbf{X}_i, \mathbf{w}_i)$ to be the bivariate normal

$$\begin{pmatrix} \tilde{\alpha}_i \\ u_{i0} \end{pmatrix} | \mathbf{X}_i, \mathbf{w}_i \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\tilde{\alpha}} & \rho_0 \sigma_{\tilde{\alpha}} \sigma_{u_0} \\ \rho_0 \sigma_{\tilde{\alpha}} \sigma_{u_0} & \sigma^2_{u_0} \end{pmatrix}\right], \tag{22}$$

which implies that the conditional distribution for u_{i0} given $\tilde{\alpha}_i$ is $\sim N[\rho_0 \frac{\sigma_{u_0}}{\sigma_{\tilde{\alpha}}}, \sigma^2_{u_0}(1-\rho_0^2)]$, where ρ_0 denotes the correlation coefficient between u_{i0} and $\tilde{\alpha}_i$. Because of the assumption of normality of $(u_{i0}, \tilde{\alpha}_i)$, $\tilde{\epsilon}_{i0}$ is itself $\sim N\left[0, \sigma^2_{u_0}\left(1-\rho_0^2\right)\right]$. Thus, we re-formulate the initial latent profit function as

$$\tilde{\pi}_{i0} = \mathbf{w}_i \tilde{\boldsymbol{\eta}} + \bar{\mathbf{x}}_{2i} \tilde{\boldsymbol{\lambda}}_0 + \rho_0 \frac{\sigma_{u_0}}{\sigma_{\tilde{\alpha}}} \tilde{\alpha}_i + \left(\sigma_{u_0} \sqrt{1-\rho_0^2}\right) \epsilon_{i0}, \tag{23}$$

where we assumed $\epsilon_{i0} \sim N(0,1)$. If we divide this equation by $\sigma_{u_0}\sqrt{1-\rho_0^2}$, we obtain the transformed latent profit function

$$\pi_{i0} = \mathbf{w}_i \boldsymbol{\eta} + \bar{\mathbf{x}}_{2i} \boldsymbol{\lambda}_{\mathbf{0}} + \delta_0 \tilde{\alpha}_i + \epsilon_{i0}, \tag{24}$$

where $\pi_{i0} = \frac{\tilde{\pi}_{i0}}{\sigma_{u_0}\sqrt{1-\rho_0^2}}$, $\boldsymbol{\eta} = \frac{\tilde{\boldsymbol{\eta}}}{\sigma_{u_0}\sqrt{1-\rho_0^2}}$, $\boldsymbol{\lambda}_{\mathbf{0}} = \frac{\tilde{\boldsymbol{\lambda}}_0}{\sigma_{u_0}\sqrt{1-\rho_0^2}}$, $\delta_0 = \frac{1}{\sigma_{\tilde{\alpha}}} \cdot \frac{\rho_0}{\sqrt{1-\rho_0^2}}$.

We now have all the components to parameterize the functions $G(\cdot)$ and $h(\cdot)$ in Eq. (17) with $\Phi(\mathbf{w}_i \boldsymbol{\eta} + \bar{\mathbf{x}}_{2i} \boldsymbol{\lambda}_{\mathbf{0}} + \delta_0 \sigma_{\tilde{\alpha}} \alpha)$ and $\phi(\alpha)$, respectively. The density can be reformulated as

$$P\left(y_{i0}, y_{i1}, \cdots, y_{iT} | \mathbf{X}_i, \boldsymbol{\theta}^{\boldsymbol{F}}\right) = \tag{25}$$
$$\int_{-\infty}^{\infty} \Phi\left\{\left[\mathbf{w}_i \boldsymbol{\eta} + \bar{\mathbf{x}}_{2i} \boldsymbol{\lambda}_{\mathbf{0}} + \delta_0 \sigma_{\tilde{\alpha}} \alpha\right](2y_{i0}-1)\right\} \prod_{t=1}^{T} \Phi\left\{\left[\mathbf{z}_{it} \boldsymbol{\vartheta} + \bar{\mathbf{x}}_{2i} \boldsymbol{\lambda}_{\mathbf{1}} + \delta_t \sigma_{\tilde{\alpha}} \alpha\right](2y_{it}-1)\right\} \phi(\alpha) d\alpha,$$

with $\alpha = \frac{\tilde{\alpha}}{\sigma_{\tilde{\alpha}}}$ and δ_T set to one to identify $\sigma_{\tilde{\alpha}}$, and $\boldsymbol{\theta}^F$ indicating the parameters to be estimated in the free correlation scenario.

It is very common in empirical papers, that use dynamic Probit models, to impose equicorrelation in the composite error. Such an assumption sets the δ_t parameters (for $t = 1, \cdots, T$) to one, and consequently the density in Eq. (26) simplifies to

$$P\left(y_{i0}, y_{i1}, \cdots, y_{iT} | \mathbf{X}_i, \boldsymbol{\theta}^{\boldsymbol{E}}\right) = \tag{26}$$
$$\int_{-\infty}^{\infty} \Phi\left\{\left[\mathbf{w}_i \boldsymbol{\eta} + \bar{\mathbf{x}}_{2i} \boldsymbol{\lambda}_{\mathbf{0}} + \delta_0 \sigma_{\tilde{\alpha}} \alpha\right](2y_{i0}-1)\right\} \prod_{t=1}^{T} \Phi\left[\left(\mathbf{z}_{it} \boldsymbol{\vartheta} + \bar{\mathbf{x}}_{2i} \boldsymbol{\lambda}_{\mathbf{1}} + \sigma_{\tilde{\alpha}} \alpha\right)(2y_{it}-1)\right] \phi(\alpha) d\alpha,$$

with $\boldsymbol{\theta}^E$ denoting the set of parameters to be estimated in the equicorrelation case.

The next two sections outline two alternative estimation procedures to estimate Heckman's solution to the initial conditions problem within the `Stata` environment.

A.1.1 Arulampalam and Stewart Standard Random Effects

Arulampalam and Stewart (2009) propose an approach to estimate Heckman's estimator, that can easily be implemented in software that deals with heteroscedasticity, such as the add-on program for `Stata`, `gllamm`. For the free correlated case, the authors suggest employing $T+1$ time dummies to identify the free correlation parameters. The density function of an expansion is modified to

$$P\left(y_{it}=1|y_{i,t-1},\cdots,y_{i0},\mathbf{X}_i,\mathbf{w}_i,\alpha_i;\boldsymbol{\theta}^{\boldsymbol{AS,F}}\right)= \\ \Phi\left[\left(\mathbf{w}_i\boldsymbol{\eta}+\bar{\mathbf{x}}_{2i}\boldsymbol{\lambda_0}+\delta_0\tilde{\alpha}_i\right)d_{it}^0+\left(\mathbf{z}_{it}\boldsymbol{\vartheta}+\bar{\mathbf{x}}_{2i}\boldsymbol{\lambda_1}\right)(1-d_{it}^0)+\tilde{\alpha}_i\sum_{\tau=1}^{T}\delta_\tau d_{it}^\tau\right], \qquad (27)$$

with $t=1,\cdots,T$ and δ_T set to one.

The alternative case of equicorrelation sets the $\delta_\tau$parameters to one, for $\tau=1,\cdots,T$, and subsequently the conditional probability of an expansion changes to

$$P\left(y_{it}=1|y_{i,t-1},\cdots,y_{i0},\mathbf{X}_i,\mathbf{w}_i,\alpha_i;\boldsymbol{\theta}^{\boldsymbol{AS,E}}\right)= \\ \Phi\left[\left(\mathbf{w}_i\boldsymbol{\eta}+\bar{\mathbf{x}}_{2i}\boldsymbol{\lambda_0}+\delta_0\tilde{\alpha}_i\right)d_{it}^0+\left(\mathbf{z}_{it}\boldsymbol{\vartheta}+\bar{\mathbf{x}}_{2i}\boldsymbol{\lambda_1}+\tilde{\alpha}_i\right)(1-d_{it}^0)\right]. \qquad (28)$$

A.1.2 Stewart's Method

Stewart has written a `Stata` code `redprob` that can estimate both cases of free correlation, Eq. (26), and equicorrelation, Eq. (27), using Gaussian-Hermite quadrature. Details on the use of his program are outlined in Stewart (2006).

The next section illustrates Orme's alternative solution to the initial condition problem.

A.2 Orme's Two-Step Procedure

Orme (2001) proposes an easy-to-use two-step estimator that is suitable for cases of low correlation between the initial conditions and unobserved heterogeneity (weak correlation). Under the assumption that y_{it} and y_{i0} are independent of one another, one can integrate the unobserved heterogeneity out of the density in the following way

$$P_i\left(y_{i1},\cdots,y_{iT}|y_{i0};\boldsymbol{\theta}\right)=\frac{\int_{-\infty}^{\infty}F\left(y_{i1},\cdots,y_{iT}|\alpha\right)G\left(y_{i0}|\alpha\right)h\left(\alpha\right)d\alpha}{\int_{-\infty}^{\infty}G\left(y_{i0}|\alpha\right)h\left(\alpha\right)d\alpha}. \tag{29}$$

If we consider y_{i0} as exogenous, the above conditional joint probability simplifies to

$$P_i\left(y_{i1},\cdots,y_{iT}|y_{i0};\boldsymbol{\theta}\right)=\int_{-\infty}^{\infty}F\left(y_{i1},\cdots,y_{iT}|\alpha\right)h\left(\alpha\right)d\alpha, \tag{30}$$

where we have exploited the condition $G\left(y_{i0}|\alpha\right)_{|\rho_0=0}=G\left(y_{i0}\right)$, with ρ_0 already defined as the correlation coefficient between $\tilde{\alpha}_i$ and u_{i0}.

By virtue of local approximation, Orme (2001) extends the logic to low values of ρ_0, and approximates the conditional joint density of Eq. (29) with

$$P_i^a\left(y_{i1},\cdots,y_{iT}|y_{i0};\boldsymbol{\theta}^O\right)=\int_{-\infty}^{\infty}F^a\left(y_{i1},\cdots,y_{iT}|\alpha\right)h\left(\alpha\right)d\alpha, \tag{31}$$

where the superscript a denotes an approximation of the function.

Similarly to Heckman (1981a,b), Orme (2001) assumes bivariate normality for the unobserved heterogeneity and the initial composite error (u_{i0}) but, differently from Heckman, he conditions the unobserved heterogeneity to the initial composite error and formulates

$$\tilde{\alpha}_i=\rho_0\frac{\sigma_{\tilde{\alpha}}}{\sigma_{u_0}}u_{i0}+\sqrt{1-\rho_0^2}\sigma_{\tilde{\alpha}}\epsilon_{i0}, \tag{32}$$

with ϵ_{i0} assumed to be conditionally independent of u_{i0}, as well as $\mathbf{w}_i$ and $\mathbf{X}_i$, and distributed as $NID(0,1)$.

The estimation of the approximated conditional density formulated in Eq. (31) requires a two-step procedure:

Step 1 Normalize $\sigma_{u_0}^2=1$, so that $G\left(y_{i0}\right)$ can be estimated as an ordinary Probit. Having $E\left(\epsilon_{i0}|y_{i0}\right)=0$, by construction, the conditional expectation for the initial composite error yields

$$E\left(u_{i0}|y_{i0}\right)=\frac{(2y_{i0}-1)\phi\left(\mathbf{w}_i\boldsymbol{\eta}+\bar{\mathbf{x}}_{2i}\boldsymbol{\lambda_0}\right)}{\Phi\left[\left(2y_{i0}-1\right)\left(\mathbf{w}_i\boldsymbol{\eta}+\bar{\mathbf{x}}_{2i}\boldsymbol{\lambda_0}\right)\right]}. \tag{33}$$

Step 2 Augment the specification in Eq. (12) as follows:

$$\pi_{it} = \gamma y_{i,t-1} + \mathbf{x}_{1it}\boldsymbol{\beta}_1 + \mathbf{x}_{2it}\boldsymbol{\beta}_2 + \bar{\mathbf{x}}_{2i}\boldsymbol{\lambda}_{\mathbf{1}} + \rho_0\sigma_{\tilde{\alpha}}u_{i0} + \sqrt{1-\rho_0^2}\sigma_{\tilde{\alpha}}\epsilon_{i0} + \epsilon_{it}, \quad t = 1, 2, \cdots, T, \tag{34}$$

where ϵ_{it} is $NID(0,1)$. Next, derive the following density:

$$P_i^a\left(y_{i1}, \cdots, y_{iT}|y_{i0}; \boldsymbol{\theta}^{\boldsymbol{O},\boldsymbol{E}}\right) = \prod_{t=1}^{T} \Phi\left[\left(\mathbf{z}_{it}\boldsymbol{\vartheta} + \bar{\mathbf{x}}_{2i}\boldsymbol{\lambda}_{\mathbf{1}} + \delta_0\hat{u}_{i0}\right)(2y_{it}-1)\right]. \tag{35}$$

where $\hat{u}_{i0}$ is the Probit residual from the first stage, and the parameter δ_0 is the product $\rho_0\sigma_{\tilde{\alpha}}$.

Arulampalam and Stewart (2009), recognize that Orme's two-step procedure can be generalized to allow for free correlation. It is enough to interact the initial composite error with time dummies d_t, yielding the following modified version of Eq. (34):

$$\pi_{it} = \gamma y_{i,t-1} + \mathbf{x}_{1it}\boldsymbol{\beta}_1 + \mathbf{x}_{2it}\boldsymbol{\beta}_2 + \bar{\mathbf{x}}_{2i}\boldsymbol{\lambda}_{\mathbf{1}} + d_t\rho_0\sigma_{\tilde{\alpha}}u_{i0} + \sqrt{1-\rho_0^2}\sigma_{\tilde{\alpha}}\epsilon_{i0} + \epsilon_{it}, \tag{36}$$

which has augmented density

$$P_i^a\left(y_{i1}, \cdots, y_{iT}|y_{i0}; \boldsymbol{\theta}^{\boldsymbol{O},\boldsymbol{F}}\right) = \prod_{t=1}^{T} \Phi\left[\left(\mathbf{z}_{it}\boldsymbol{\vartheta} + \bar{\mathbf{x}}_{2i}\boldsymbol{\lambda}_{\mathbf{1}} + \delta_t\hat{u}_{i0}\right)(2y_{it}-1)\right], \tag{37}$$

where $\delta_t \equiv d_t\rho_0\sigma_{\tilde{\alpha}}$.

Arulampalam et al. (2000) show that Orme's estimator is heteroscedastic for high values of ρ_0, as the conditional variance of the second stage error component is:

$$Var(\epsilon_{i0}|y_{i0}) = \sigma_{\tilde{\alpha}}^2\left\{1 - \rho_0^2\left[\frac{\phi\left(\mathbf{w}_i\boldsymbol{\eta} + \bar{\mathbf{x}}_{2i}\boldsymbol{\lambda}_{\mathbf{0}}\right)}{\Phi\left(\mathbf{w}_i\boldsymbol{\eta} + \bar{\mathbf{x}}_{2i}\boldsymbol{\lambda}_{\mathbf{0}}\right)\Phi\left(-\mathbf{w}_i\boldsymbol{\eta} - \bar{\mathbf{x}}_{2i}\boldsymbol{\lambda}_{\mathbf{0}}\right)}\right]^2\right\}. \tag{38}$$

A.3 Wooldridge's Method

Wooldridge (2005) proposes conditioning the distribution of unobserved heterogeneity on the initial value of the dependent variable and full history of the (exogenous) time-varying covariates.

Wooldridge's idea produces the following density function

$$P\left(y_{i1},\cdots,y_{iT}|y_{i0},\mathbf{X}_i,\boldsymbol{\theta}^{\boldsymbol{W}}\right)=$$
$$\int_{-\infty}^{\infty}\left[\prod_{t=1}^{T}f_t\left(y_{it}=1|y_{i,t-1},\cdots,y_{i0},\mathbf{X}_i,c\right)\right]h\left(c|y_{i0},\mathbf{X}_i\right)dc. \tag{39}$$

Using Mundlak-Chamberlain's approach, we parameterize the general functions in Eq. (39) as

$$\begin{aligned} h\left(c_i|y_{i0},\mathbf{X}_i\right) &\sim N\left(\mu_0 y_{i0}+\bar{\mathbf{x}}_{2i}\boldsymbol{\lambda}_{\mathbf{0}},\sigma_\alpha^2\right) \\ f_t\left(y_{it}=1|y_{i,t-1},\cdots,y_{i0},\mathbf{X}_i,c_i\right) &= \Phi\left[\left(\mathbf{z}_{it}\boldsymbol{\vartheta}+c_i\right)\left(2y_{it}-1\right)\right]. \end{aligned}$$

One limitation of Wooldridge's methodology is that one cannot identify the time-invariant variables that are correlated with the unobserved heterogeneity. Nevertheless they should all be included in the exogenous variables, as they bring about explanatory power to the estimation process. The major advantage of Wooldridge's methodology is that it can be easily estimated in `Stata` using the command `xtprobit`.

The densities to be estimated, in case of equicorrelation are

$$P\left(y_{i1},\cdots,y_{iT}|y_{i0},\mathbf{X}_i,\boldsymbol{\theta}^{\boldsymbol{W,E}}\right)=$$
$$\int_{-\infty}^{\infty}\prod_{t=1}^{T}\Phi\left\{\left[\mathbf{z}_{it}\boldsymbol{\vartheta}+\bar{\mathbf{x}}_{2i}\boldsymbol{\lambda}_{\mathbf{0}}+\mu_0 y_{i0}+\sigma_{\tilde{\alpha}}\alpha\right]\left(2y_{it}-1\right)\right\}\phi(\alpha)d\alpha, \tag{40}$$

and in case of free correlation

$$P\left(y_{i1},\cdots,y_{iT}|y_{i0},\mathbf{X}_i,\boldsymbol{\theta}^{\boldsymbol{W,F}}\right)=$$
$$\int_{-\infty}^{\infty}\prod_{t=1}^{T}\Phi\left[\left(\mathbf{z}_{it}\boldsymbol{\vartheta}+\bar{\mathbf{x}}_{2i}\boldsymbol{\lambda}_{\mathbf{0}}+\mu_0 y_{i0}+\delta_t\sigma_{\tilde{\alpha}}\alpha\right)\left(2y_{it}-1\right)\right]\phi(\alpha)d\alpha, \tag{41}$$

with δ_T set to one to identify $\sigma_{\tilde{\alpha}}$.

B The Copula

The Copula is a function that maps from marginals to the multivariate joint distribution.[39] As such, it can be used to recover multivariate joint distributions from information on marginal distributions and

[39] It was first introduced by Hoeffding (1940) (see collection Hoeffding (1994)), but is commonly associated to Sklar (1973).

is particularly suitable to deal with non-normal data or with marginals coming from different parametric families. Prior to outlining the Copula, we introduce some notation. Let $(R_1, R_2, \cdots, R_H)$ be H random variables, with distribution functions $G_1(r_1) = P(R_1 \leq r_1), G_2(r_2) = P(R_2 \leq r_2), \cdots,$ $G_H(r_H) = P(R_H \leq r_H)$, respectively, and joint distribution function $L(r_1, r_2, \cdots, r_H) = P(R_1 \leq r_1,$ $R_2 \leq r_2, \cdots, R_H \leq r_H)$. Also, let $\bar{\Re}$ denote the extended real line $[-\infty, \infty]$. We define an H-dimensional distribution function L with domain $\bar{\Re}^H$ a function that has the following properties:

1. $L(\infty, \cdots, \infty, r_h, \infty, \cdots, \infty) = G_h(r_h)$ for any$h \leq H$;

2. $L(\infty, \cdots, \infty, \cdots, \infty) = 1$;

3. $L(r_1, \cdots, r_H) = 0$ if $r_h = -\infty$ for any $h \leq H$;

4. L is H-increasing.

Next, if we denote with $\Im$ the unit interval $[0, 1]$, we have all the notation required to formulate the main Copula's theorem:

Theorem B.1 (Sklar's theorem) *Let L be an H-dimensional distribution function with distribution functions $G_1, G_2, \cdots, G_H$ and denote with $\boldsymbol{\varrho}$ the vector of parameters that measures the dependence between the marginals. Then, there exists an H-Copula C such that $\forall \ (r_1, r_2, \cdots, r_H) \in \bar{\Re}^H$,*

$$L(r_1, r_2, \cdots, r_H) = C[G_1(r_1), G_2(r_2), \cdots, G_H(r_H); \boldsymbol{\varrho}]. \tag{42}$$

If $G_1, G_2, \cdots, G_H$ are all continuous, then C is unique; otherwise C is uniquely determined on $[RanG_1 \times RanG_2 \times \cdots \times RanG_H]$, where Ran denotes the range of the distribution function. Conversely, if C is an H-Copula and $G_1, G_2, \cdots, G_H$ are distribution functions, then the function L defined in Eq. (42) is an H-dimensional distribution function with marginal distribution functions $G_1, G_2, \cdots, G_H$.

Sklar's (1973) theorem states that an $H-$dimensional Copula is the function C that maps from the unit space $\Im^H$ to the unit interval $\Im$ and that satisfies the conditions:

1. $C(1, \cdots, 1, u_h, 1, \cdots, 1) = u_h \ \ \forall h \leq H$ and $u_h \in [0, 1]$;

2. $C(u_1, \cdots, u_H) = 0$ if $u_h = 0$ for any $h \leq H$;

3. C is H-increasing.

Hence, an H-dimensional Copula is a multivariate distribution function that has all H one-dimensional marginals over the uniform distribution, $U(0,1)$.

An important corollary of the theorem B.1 sets out a method to construct Copulas directly from joint distribution functions:

Corollary B.2 (From joint distribution functions to Copulas) *Let $L, C, G_1, G_2, \cdots, G_H$ be as in theorem B.1, and let $G_1^{-1}, G_2^{-1}, \cdots, G_H^{-1}$, be inverse (quantile) functions of $G_1, G_2, \cdots, G_H$, respectively. Then for any point $(u_1, u_2, \cdots, u_H) \in \Im^H$,*

$$C(u_1, u_2, \cdots, u_H; \boldsymbol{\varrho}) = L\left[G_1^{-1}(u_1), G_2^{-1}(u_2), \cdots, G_H^{-1}(u_H)\right], \tag{43}$$

with $G_1^{-1} = r_1, G_2^{-1} = r_2, \cdots, G_H^{-1} = r_H$. Corollary B.2 is particularly useful for simulations.

In the rest of the section we introduce the parametric Copula functions that are available for the multivariate case of $H > 2$. We separate them in two classes: Elliptical and Archimedean.

B.1 Elliptical Copula

If we differentiate Eq. (43) we get the density of the Elliptical Copula

$$c(u_1, u_2, \cdots, u_H; \boldsymbol{\varrho}) = \frac{l\left[G_1^{-1}(u_1), G_2^{-1}(u_2), \cdots, G_H^{-1}(u_H)\right]}{\prod_{h=1}^{H} l_h\left[G_h^{-1}(u_h)\right]}, \tag{44}$$

where l is the joint probability density function of the Elliptical distribution, and l_h the marginal density functions. If we assume the marginal distribution functions to be standard normals, $G_h = \Phi$, we have the H-dimensional Copula standard normal (gaussian), whose density is given by

$$\begin{aligned} &c(u_1, u_2, \cdots, u_H; \boldsymbol{\Gamma}) = \\ &\quad |\boldsymbol{\Gamma}|^{-1/2} \exp\left\{-\frac{1}{2}\left[\Phi^{-1}(u_1), \cdots, \Phi^{-1}(u_H)\right]' \left(\mathbf{I}_H - \boldsymbol{\Gamma}^{-1}\right)\left[\Phi^{-1}(u_1), \cdots, \Phi^{-1}(u_H)\right]\right\}, \end{aligned} \tag{45}$$

where $\mathbf{I}_H$ denotes the identity matrix and $\boldsymbol{\Gamma}$ the dispersion matrix.

B.2 Archimedean Copula

Theorem B.3 (Nelsen (2006)'s Archimedean Copula) *Let φ be a continuous strictly decreasing function from $\Im$ to $[0, \infty]$ such that $\varphi(0) = \infty$ and $\varphi(1) = 0$, and let φ^{-1} denote the inverse of φ. If C^H is the function from $\Im^H$ to $\Im$ given by*

$$C^H (u_1, u_2, \cdots, u_H; \varrho) = \varphi^{-1} [\varphi (u_1) + \varphi (u_2) + \cdots + \varphi (u_H)], \tag{46}$$

then C^H is a H-Copula $\forall h \geq 2$, if and only if φ^{-1} is completely monotonic on $[0, \infty)$.

We now parameterize the generation function φ and its inverse φ^{-1} and obtain some well-known Archimedean Copulas:

1. **Clayton Copula**: If we let the generator function be $t = \varphi (u; \varrho) = u^{-\varrho} - 1$ for $\varrho > 0$, then its inverse is $\varphi^{-1} (t; \varrho) = (1 + t)^{-1/\varrho}$, and the resulting multivariate Copula is

$$C^H (u_1, u_2, \cdots, u_H; \varrho) = \left(u_1^{-\varrho} + u_2^{-\varrho} + \cdots + u_H^{-\varrho} - H + 1\right)^{-1/\varrho}. \tag{47}$$

2. **Frank Copula**: If we let the generator function be $t = \varphi (u; \varrho) = -\ln \frac{(e^{-\varrho u} - 1)}{(e^{-\varrho} - 1)}$ for $\varrho > 0$, then its inverse is $\varphi^{-1} (t; \varrho) = -\frac{1}{\varrho} \ln [1 - (1 - e^{-\varrho}) e^{-t}]$, and the resulting multivariate Copula is

$$C^H (u_1, u_2, \cdots, u_H; \varrho) = -\frac{1}{\varrho} \ln \left[1 + \frac{(e^{-\varrho u_1} - 1)(e^{-\varrho u_2} - 1) \cdots (e^{-\varrho u_H} - 1)}{(e^{-\varrho} - 1)^{H-1}}\right]. \tag{48}$$

3. **Gumbel Copula**: If we let the generator function be $t = \varphi (u; \varrho) = (-\ln u)^{\varrho}$ for $\varrho \geq 1$, then its inverse is $\varphi^{-1} (t; \varrho) = \exp\left(-t^{1/\varrho}\right)$, and the resulting multivariate Copula is

$$C^H (u_1, u_2, \cdots, u_H; \varrho) = \exp\left\{- [(-\ln u_1)^{\varrho} + (-\ln u_2)^{\varrho} + \cdots + (-\ln u_H)^{\varrho}]^{1/\varrho}\right\}. \tag{49}$$

Consult Nelsen (2006) for further reading on Copulas.

References

Adelman I. 1958. A stochastic analysis of the size distribution of firms. *Journal of the American Statistical Association* **53**: 893–904.

Amaral L, Buldyrev S, Havlin S, Leschhorn H, Maass P, Salinger M, Stanley H, Stanley M. 1997. Scaling behavior in economics: Empirical results for company growth. *Journal De Physique France* **7**: 621–633.

Amisano G, Giorgetti M. 2005. Entry in pharmaceutical submarkets: a Bayesian panel probit approach. *Working Papers ubs0408, University of Brescia, Department of Economics* .

Arulampalam W. 1999. A note on estimated effects in random effect probit models. *Oxford Bulletin of Economics and Statistics* **61**: 597–602.

Arulampalam W, Booth A, Taylor M. 2000. Unemployment persistence. *Oxford Economic Papers* **52**: 24.

Arulampalam W, Stewart M. 2009. Simplified implementation of the Heckman estimator of the dynamic probit model and a comparison with alternative estimators. *Oxford Bulletin of Economics and Statistics* **71**: 659–681.

Axtell R. 2001. Zipf distribution of US firm sizes. *Science* **293**: 1818–1820.

Bain J. 1951. Relation of profit rate to industry concentration: American manufacturing, 1936-1940. *The Quarterly Journal of Economics* **65**: 293–324.

Bain J. 1956. *Barriers to New Competition*. Harvard University Press: Cambridge, MA.

Bottazzi G, Dosi G, Lippi M, Pammolli F, Riccaboni M. 2001. Innovation and corporate growth in the evolution of the drug industry. *International Journal of Industrial Organization* **19**: 1161–1187.

Bottazzi G, Secchi A. 2005. Growth and diversification patterns of the worldwide pharmaceutical industry. *Review of Industrial Organization* **26**: 195–216.

Bottazzi G, Secchi A. 2006. Gibrat's Law and diversification. *Industrial and Corporate Change* **15**: 847–875.

Buldyrev S, Growiec J, Pammolli F, Riccaboni M, Stanley H. 2007. The growth of business firms: facts and theory. *Journal of the European Economic Association* **5**: 574–584.

Buzzacchi L, Valletti T. 2006. Firm size distribution: testing the "independent submarkets model" in the Italian motor insurance industry. *International Journal of Industrial Organization* **24**: 809–834.

Cabral L, Mata J. 2003. On the evolution of the firm size distribution: facts and theory. *American Economic Review* **93**: 1075–1090.

Calvo J. 2006. Testing Gibrats Law for small, young and innovating firms. *Small Business Economics* **26**: 117–123.

Cecchini P, Catinat M, Jacquemin A. 1988. *The European challenge, 1992 : the benefits of a single market*. Wildwood House: Aldershot.

Cefis E, Ciccarelli M, Orsenigo L. 2007. From Gibrats legacy to Gibrats fallacy: a Bayesian approach to study the growth of firms. *Structural Change and Economic Dynamics* **18**: 348–369.

Chamberlain G. 1984. Panel data. In Griliches Z, Intriligator M (eds.) *Handbook of Econometrics*, volume 2. North Holland: Amsterdam, 1247–1318.

Chesher A. 1979. Testing the law of proportionate effect. *The Journal of Industrial Economics* **27**: 403–411.

Cohen W, Klepper S. 1992. The anatomy of industry R&D intensity distributions. *The American Economic Review* **82**: 773–799.

Cohen W, Klepper S. 1996. A reprise of size and R&D. *The Economic Journal* **106**: 925–951.

de Juan R. 2003. The independent submarkets model: an application to the Spanish retail banking market. *International Journal of Industrial Organization* **21**: 1461–1487.

Dunne P, Hughes A. 1994. Age, size, growth and survival: UK companies in the 1980s. *The Journal of Industrial Economics* **42**: 115–140.

Dunne T, Roberts M, Samuelson L. 1988. Patterns of firm entry and exit in US manufacturing industries. *The RAND Journal of Economics* **19**: 495–515.

Dunne T, Roberts M, Samuelson L. 1989. The growth and failure of US manufacturing plants. *The Quarterly Journal of Economics* **104**: 671–698.

EFPIA. 2005. Pharmaceutical industry in figures. The European Federation of Pharmaceutical Industries and Associations (EFPIA).

Ericson R, Pakes A. 1995. Markov-perfect industry dynamics: a framework for empirical work. *The Review of Economic Studies* **62**: 53–82.

Evans D. 1987a. Tests of alternative theories of firm growth. *The Journal of Political Economy* **95**: 657–674.

Evans D. 1987b. The relationship between firm growth, size, and age: estimates for 100 manufacturing industries. *The Journal of Industrial Economics* **35**: 567–581.

Freeman J. 1986. A piecewise regression analysis for firm-size distributions. *The Statistician* **35**: 547–559.

Geroski P, Machin S. 1991. *Market Dynamics and Entry*. Blackwell: Oxford.

Geroski P, Machin S, Walters C. 1997. Corporate growth and profitability. *The Journal of Industrial Economics* **45**: 171–189.

Gibrat R. 1931. *Les Inégalités Économiques*. Librairie du Recueil Sirey: Paris.

Gibrat R. 1957. On economic inequality. *International Economic Papers* **7**: 53–70.

Hall B. 1987. The relationship between firm size and firm growth in the US manufacturing sector. *The Journal of Industrial Economics* **35**: 583–606.

Hall M, Weiss L. 1967. Firm size and profitability. *The Review of Economics and Statistics* **49**: 319–331.

Hart P. 1962. The size and growth of firms. *Economica* **29**: 29–39.

Hart P, Oulton N. 1996. Growth and size of firms. *The Economic Journal* **106**: 1242–1252.

Hart P, Prais S. 1956. The analysis of business concentration: a statistical approach. *Journal of the Royal Statistical Society. Series A (General)* **119**: 150–191.

Heckman J. 1981a. Heterogeneity and state dependence. *Studies in Labor Markets* **31**: 91–140.

Heckman J. 1981b. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In Manski C, McFadden D (eds.) *Structural Analysis of Discrete Data and Econometric Applications*. MIT Press: Cambridge, MA, 179–195.

Higson C, Holly S, Kattuman P. 2002. The cross sectional dynamics of the US business cycle: 1950-1999. *Journal of Economic Dynamics and Control* : 1539–1555.

Hjalmarsson L. 1974. The size distribution of establishments and firms derived from an optimal process of capacity expansion. *European Economic Review* **5**: 123–140.

Hoeffding W. 1940. Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin* **5**: 179–233.

Hoeffding W. 1994. Scale-invariant correlation theory. In Fisher N, Sen P (eds.) *The Collected Works of Wassily Hoeffding*. Springer-Verlag: New York, 57–107.

Howells J. 1992. Pharmaceuticals and europe 1992: the dynamics of industrial change. *Environment and Planning A* **24**: 33–48.

Hymer S, Pashigian P. 1962. Firm size and rate of growth. *The Journal of Political Economy* **70**: 556–569.

Ijiri Y, Simon H. 1964. Business firm growth and size. *The American Economic Review* **54**: 77–89.

Ijiri Y, Simon H. 1967. A model of business firm growth. *Econometrica: Journal of the Econometric Society* **35**: 348–355.

Jovanovic B. 1982. Selection and the evolution of industry. *Econometrica: Journal of the Econometric Society* **50**: 649–670.

Klepper S. 1996. Entry, exit, growth, and innovation over the product life cycle. *The American Economic Review* **86**: 562–583.

Klepper S, Graddy E. 1990. The evolution of new industries and the determinants of market structure. *The RAND Journal of Economics* **21**: 27–44.

Klepper S, Simons K. 2000. The making of an oligopoly: firm survival and technological change in the evolution of the US tire industry. *The Journal of Political Economy* **108**: 728–760.

Klepper S, Thompson P. 2006. Submarkets and the evolution of market structure. *The RAND Journal of Economics* **37**: 861–886.

Klette T, Kortum S. 2004. Innovating firms and aggregate innovation. *The Journal of Political Economy* **112**: 986–1018.

Kojadinovic I, Yan J. 2010. Modeling multivariate distributions with continuous margins using the copula R package. *Journal of Statistical Software* **34**: 1–20.

Kotzian P. 2004. Pharmaceutical r&d in the setting of incomplete european integration. *International Journal of the Economics of Business* **11**: 175–195.

Lotti F, Santarelli E, Vivarelli M. 2003. Does Gibrat's Law hold among young, small firms? *The Journal of Evolutionary Economics* **13**: 213–235.

Lucas Jr R. 1978. On the size distribution of business firms. *The Bell Journal of Economics* **9**: 508–523.

Mansfield E. 1962. Entry, Gibrat's Law, innovation, and the growth of firms. *The American Economic Review* **52**: 1023–1051.

Mason E. 1939. Price and production policies of large-scale enterprise. *The American Economic Review* **29**: 61–74.

Mason E. 1948. The current status of the monopoly problem in the United States. *Harvard Law Review* **62**: 1265–1285.

Matraves C. 1999. Market structure, r&d and advertising in the pharmaceutical industry. *The Journal of Industrial Economics* **47**: 169–194.

Mundlak Y. 1978. On the pooling of time series and cross section data. *Econometrica* **46**: 69–85.

Nelsen R. 2006. *An Introduction to Copulas*. Springer: New York.

Orme C. 2001. Two-step inference in dynamic non-linear panel data models. *Unpublished paper, University of Manchester* .
URL `http://personalpages.manchester.ac.uk/staff/chris.orme/documents/Research%20Papers/initcondlast.pdf`

Pavitt K, Robson M, Townsend J. 1987. The size distribution of innovating firms in the UK: 1945-1983. *The Journal of Industrial Economics* **35**: 297–316.

Quandt R. 1966. On the size distribution of firms. *The American Economic Review* **56**: 416–432.

Radice H. 1971. Control type, profitability and growth in large firms: an empirical study. *The Economic Journal* **81**: 547–562.

Ricci V. 2005. Fitting distributions with R. *R project website* . URL `http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf`

Rostow W. 1959. The stages of economic growth. *Economic History Review* : 1–16.

Samuels J, Chesher A. 1972. Company growth, survival and size. In Cowling K (ed.) *Market Structure and Corporate Behaviour*. Gray Mills: London.

Samuels JM. 1965. Size and the growth of firms. *The Review of Economic Studies* **32**: 105–112.

Samuels JM, Smyth DJ. 1968. Profits, variability of profits and firm size. *Economica* **35**: 127–139.

Saving T. 1965. The four-parameter lognormal, diseconomies of scale and the size distribution of manufacturing establishments. *International Economic Review* **6**: 105–114.

Scherer F. 1965. Firm size, market structure, opportunity, and the output of patented inventions. *The American Economic Review* **55**: 1097–1125.

Selten R. 1983. A model of oligopolistic size structure and profitability. *European Economic Review* **22**: 33–57.

Simon H. 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics* **69**: 99–118.

Simon H, Bonini C. 1958. The size distribution of business firms. *The American Economic Review* **48**: 607–617.

Simon H, Ijiri Y. 1977. *Skew Distributions and the Sizes of Business Firms*. North Holland: Amsterdam.

Singh A, Whittington G, Burley H. 1968. *Growth, Profitability and Valuation: a Study of United Kingdom Quoted Companies*. Cambridge University Press: Cambridge.

Sklar A. 1973. Random variables: joint distribution functions and copulas. *Kybernetika* **9**: 449–460.

Smyth D, Boyes W, Peseau D. 1975. The measurement of firm size: theory and evidence for the United States and the United Kingdom. *The Review of Economics and Statistics* **57**: 111–114.

Steindl J. 1965. *Random Process and the Growth of Firms: a study of the Pareto Law.* Hafner: New York.

Steindl J. 1968. Size distributions in economics. In *International Encyclopedia of the Social Sciences*, volume 14. New York: MacMillan and Free Press, 294–300.

Stewart MB. 2006. -redprob- A Stata program for the Heckman estimator of the random effects dynamic probit model. *mimeo, University of Warwick* .

Sutton J. 1997a. Game-theoretic models of market structure. In Kreps DM, Wallis KF (eds.) *Advances in Economics and Econometrics: Theory and Applications. Seventh World Congress*, 26. Econometric Society, Cambridge University Press: Cambridge, 66–86.

Sutton J. 1997b. Gibrat's legacy. *Journal of Economic Literature* **35**: 40–59.

Sutton J. 1998. *Technology and Market Structure: Theory and History.*

Variyam J, Kraybill D. 1994. Managerial inputs and the growth of rural small firms. *American Journal of Agricultural Economics* **76**: 568–575.

Wedervang F. 1965. *Development of a Population of Industrial Firms.* Universitetsforlaget: Blindern, Oslo.

Wooldridge J. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Jouranal of Applied Econometrics* **20**: 39–54.

Yan J. 2007. Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software* **21**. URL `http://ideas.repec.org/a/jss/jstsof/21i04.html`