Australian School of Business Research Paper No. 2012 ECON 24

Small Business Redefined: A Quasi-Linear Fuzzy Classification of Firm Size

Sasan Bakhtiari

Small Business Redefined:
A Quasi-Linear Fuzzy Classification of Firm Size

Sasan Bakhtiari
Australian School of Business
University of New South Wales
Sydney NSW 2052,
Australia

April 4, 2012

Abstract

The quasi-linear fuzzy modeling of Filev (1991) is used to estimate the relationship between the number of managers and employees in a firm. The results form the basis for the classification of firms into small and large businesses. Application to a data of Australian firms shows an evolution episode during which firms are driven by various transitional forces. The composition of the transition region suggests that the 2011 small business tax-break cap set by Australian Taxation Office falls short of fully supporting growth as intended. The implications pave the way for improvement to the business tax code aiming at growth and job creation.

Keywords: fuzzy logic, small business, job creation, business taxation.

JEL Code: C38, C61, D23, H25.

1 Introduction

The importance of small businesses cannot be emphasized enough. They are vital to the process of creative destruction. Young small firms, in particular, contribute greatly to job creation (Haltiwanger et al., 2010). Many large companies are supported by a vast network of small suppliers. But, what is a small business?

* Author's contact is Phone: (+61 2) 9385 3962, Fax: (+61 2) 9313 6337, Email: s.bakhtiari@unsw.edu.au.

Country	Source	Definition
Institutions		
USA	IRS (2011)	Less than 25 employees
UK	HMRC (2005)	Less than 50 employees
Australia	ATO (2011)	Less than 2 mil. (AUD) annual revenue
Australia	ABS (2009)	Less than 20 employees
Economic Research		
USA	Acs & Audretsch (1988)	Less than 500 employees
USA	Haltiwanger et al. (2010)	Less than 500 employees
USA	Neumark et al. (2011)	Less than 20 employees
UK	Barnes & Haskel (2002)	Less than 100 employees
UK	Storey (1994)	Manufacturing: Less than 200 employees Construction and Mining: Less than 25 employees Retail: Less than £50,000 Wholesale Trades: Less than £200,000 Road Transport: Less than 6 vehicles
Dutch	Broersma & Gautier (1997)	Less than 100 employees

Table 1: The definition of small business from different sources.

Prima facie, there is no universal consensus on the definition. In the US, firms with less than 25 employees can claim the New Health Care Tax Credit (IRS, 2011). Firms less than 50 employees can benefit from reduced paperwork in the UK (HMRC, 2005). For an Australian business to qualify for small business tax credit on new investments, there is a two million dollar limit on annual turnover (ATO, 2011). Even economists tend to vary over the definition. Table 1 lists a variety of definitions that have been used to describe small businesses. The interesting feature of the classification system in Storey (1994) is the dependence between the definition of small business and the industry of operation, also the adaptive use of multiple size instruments such as sales, employment, and even the number of vehicles.

This paper offers a structural approach to defining small business which utilizes one fundamental difference between small and large businesses. Borrowing from a vast literature on firm management and organization, small businesses are defined as those that are small enough in their scale of operation to be manageable by the owner(s) without any assistance from professional managers and supervisors. In a large firm, on the other hand, the operation will falter and break down in the absence of enough monitoring and coordination by a management team. The average management ability of an owner is assumed innate and time-invariant. As a result, the definition of small business is time invariant and independent of prices and inflation, contrary to the use of sales as such instrument. The classification is also robust to changes in the distribution of size as firms align themselves with government and taxation policies to benefit from small business tax breaks. Section 2 elaborates on this issue further and demonstrates the suitability of piecewise log-linear models in approximating the employee–management relationship.

The employee-management relation is estimated using a Quasi-Linear Fuzzy Modeling (QLFM) introduced by Filev (1991), and the transition region where the evolution from small to large business occurs is identified in the process. The fuzzy modeling, in particular, offers a conceptual classification of size with soft and seamless transition between the classes of firms that well mimics the graduality of transition and also accounts for heterogeneity across firms. Statistical classification methods, such as mixture models (McLachlan & Peel, 2000), similarly estimate the degree of classification but on an individual basis, hence the classification is noisy in the sense that proximate observations do not necessarily classify the same way. The QLFM overcomes this issue by effectively filtering noise through its use of membership functions to offer a "clean" view of classes. Moreover, the running time for the QLFM is faster on average, because the estimation method is OLS, although the structure being estimated is inherently nonlinear. In comparison to non-parametric threshold models, namely, regression splines Friedman (1991), the QLFM is theoretically able to use fewer classes yet produce better fits, owing to its smooth transition properties. In fact, the QLFM is flexible enough to estimate near threshold models by a change of parameter, thus threshold models are only a subset of models handled by the QLFM. Finally, a full set theory exists for fuzzy sets that is utilized to study the properties of firms as they transition from small to large business.

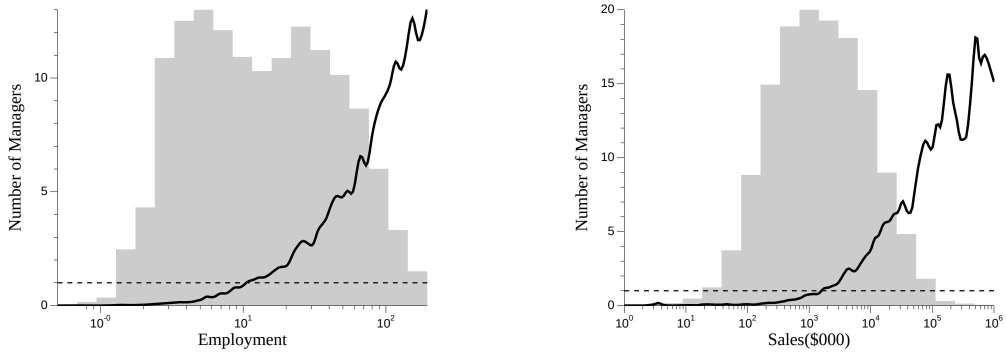
In the remainder, Section 3 describes the principals of fuzzy logic and the QLFM algorithm. Section 3.3 outlines the numerical implementation of the algorithm. Section 4 describes the source of data and the measurement of size. Section 5 reports the identification results using the Australian data. Section 6 compares the findings to those from a mixture model and also serves as robustness test. In Sections 7, the QLFM is applied to identify a threshold model and make some sharp predictions about the transition of firms from small to large size. Paper is concluded afterwards.

2 A Basis for Size Classification

Storey (1989) observes that "small firms are not simply 'scaled down' versions of large firms." Some studies have already documented fundamental differences in the performance, formation, and the business objectives of small and large businesses (Acs & Audretsch, 1988; Storey, 1989; Haltiwanger et al., 2010). In one instance, Storey (1989) associates small businesses with a persona: incentives in a small business are not necessarily directed towards expanding profit and markets, but more often accentuate the owner's desire to maintain control and achieve job satisfaction. Once the business grows into a large size, however, the management tends to pursue incentives that are fairly objective in nature.

In a broader sense, the management structure and organization of firms is a nonscalable feature of size. A small business, the one that possesses a persona, can effectively be managed and directed entirely by the owner(s), with no need for assistance from professional managers and supervisors. In larger firms, on the other hand, the same management strategy is ineffective because of the "free-rider" problem (Alchian & Demsetz, 1972): in a large firm, workers have less incentive to contribute to production and tend to hide behind other workers' efforts. In a Nash equilibrium with identical workers, every worker provides the lowest effort and production suffers tremendously. At some point, the workforce needs to be broken into "teams" and "divisions", and each group of workers is operated and supervised by a manager. But, with the size of the businesses growing, the same free-riding rule now applies to the managers themselves, and another management tier is required to oversee the performance of the lower-ranking managers. The story goes on as size keeps growing. Through a formulation of this process, Williamson (1967) is able to establish a strictly positive relationship between employment and the number of managers, also the number of management hierarchies, within a firm.

Data readily testifies to the existence of such piecewise relation between the number of



(a) BW=0.1 (b) BW=0.2

Figure 1: The non-parametric estimation of the relation between the number of managers and employment or real sales. A kernel regression with Gaussian kernel is used. The bandwidth for log values is indicated under the plot. Sales are deflated to 2011 dollars using March GDP deflators. Management numbers exclude owners. The gray bars illustrate the distribution of observations over the horizontal axis. The dashed line indicates one manager. The sample is 25,553 firm–years from the Business Longitudinal Survey.

employees and the number of managers (Figure 1).[1] Smaller firms typically operate without any managers and are run entirely by their owner(s). But some firms appear to be hiring between zero and one manager on average. These firms mark a transition phase partly induced by the heterogeneity of owners' abilities in effective management, but also caused by some transitional forces. Beyond this range, the number of managers clearly rises with size, as predicted by Williamson (1967). Ignoring those transitioning firms for the moment, the relation can be described in the simplest from as

$$MAN(EMP) = \begin{cases} 0, & EMP < \bar{E}, \\ M_1 \log(EMP/\bar{E}), & EMP \geq \bar{E}, \end{cases} \qquad M_1 > 0, \tag{1}$$

in which MAN is the number of managers, and EMP is the number of employees. $\bar{E}$ is the maximum manageable number of employees by an average owner, above which a firm has to be characterized as large business. The goal of the classification is to estimate $\bar{E}$ using the data that generated Figure 1(a).

Lastly, as opposed to sales and employment, the maximum manageable size is a more robust classifier of size. Sales are affected by inflation and the overall economic performance.

[1] For the description and details of the data see Section 4.

Employment size, by itself, can also be endogenous to government policies. For instance, if firms below 50 employees receive tax breaks, the size distribution of firm will be affected by firm repositioning themselves to take the full advantage of tax breaks. The maximum number of manageable workers by an owner, however, is unaffected by these forces.

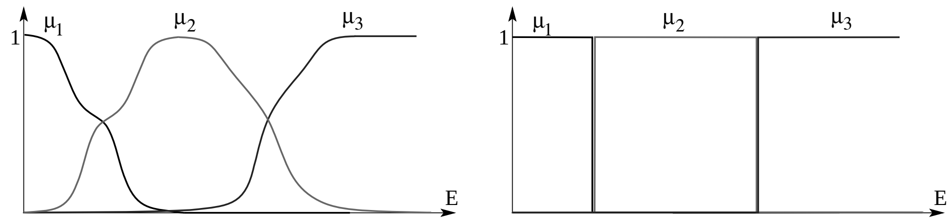
3 Fuzzy Classification Method

3.1 Basics of Fuzzy Logic

Introduced by Zadeh (1965), fuzzy logic mimics a conceptual, rather than mechanical, approach to classifying objects or values. In a "crisp" classification, attributed to machines, the membership of an object in a class is dichotomous, i.e., zero–one or yes–no type. For continuous variables, the border or threshold value between the two neighboring classes is precisely defined. For instance, dealing with positive and negative real numbers, it is agreed that the threshold separating one class from the other is zero. Conversely, any real number can be exactly classified into positive or negative.

However, in most standard situations such crisp classification could be impossible because 1) everyone cannot agree on the same definition for classes, and 2) even an individual is unable to provide a precise specification of her classification. For example, in specifying "small business", not only people disagree on a crisp upper limit on size, but also an individual is yet unsure whether a firm with 50 employees is small or large. Instead, one resorts to descriptions such as "somewhat small", "rather small", or "relatively small". With larger employment size, one's belief that the business is small fades and gives place to the belief that the business is large. As a result, we do not envisage a clear threshold between the two classes, rather see them diffusing into each other.

Zadeh's solution to this problem is to extend the zero–one level of membership into a continuum of membership grades in the interval [0,1]. A membership of one for an object or value says that it is undoubtedly a member of the class, and a membership of zero suggests the opposite. Any membership between zero and one reflects one's belief about how strongly the object or value is associated with a certain class. Let $x \in \mathbb{R}^r$ be a vector in space and $\mathcal{X}^{(i)}$, $i = 1, \ldots, c$ be c fuzzy notions specifying classes of interest. For instance, if x



(a) Fuzzy Membership

(b) Boolean Membership

Figure 2: (a) An illustration of fuzzy membership functions for three classes. (b) the illustration of the same three classes, but with Boolean membership functions that only assume $\{0, 1\}$ grades

is employment, then $\mathcal{X}^{(1)}$ =“Small” and $\mathcal{X}^{(2)}$ =“Large”. The strength of the membership relation $x \in \mathcal{X}^{(i)}$ is denoted by the membership function $\mu^{(i)}(x)$ and satisfies

$$0 \leq \mu^{(i)}(x) \leq 1, \quad \forall x, \tag{2}$$

$$\mu^{(i)}(x) \text{ is continuous,} \tag{3}$$

$$\sum_{i=1}^{c} \mu^{(i)}(x) = 1, \quad \forall x. \tag{4}$$

Condition (4), in particular, ensures the completeness of $\mathcal{X}$; there is no outlying class. A sample classification system with three classes and the corresponding membership functions satisfying the above conditions is shown in Figure 2(a). Note that, in this example, some values of x have their membership split between two classes. This is opposed to the Boolean specification of the same three classes, where the membership grades only assume $\{0, 1\}$ values and every point is a member of one and only one class (Figure 2(b)).

Treating fuzzy sets as an extension to the standard set theory, Zadeh (1965) also defines union and intersection among the fuzzy sets as follows:

$$\begin{aligned} \text{Intersection:} \quad & \left(\mu^{(i)} \wedge \mu^{(k)}\right)(x) = \min\left(\mu^{(i)}(x), \mu^{(k)}(x)\right), \\ \text{Union:} \quad & \left(\mu^{(i)} \vee \mu^{(k)}\right)(x) = \max\left(\mu^{(i)}(x), \mu^{(k)}(x)\right). \end{aligned}$$

These set operations are handy in defining new sets of observations (here firms) using the existing fuzzy sets.

3.2 Quasi-Linear Fuzzy Modeling (QLFM)

The QLFM identifies an approximation to nonlinear relations by breaking the support into (fuzzy) regions and describing the relation in each region using a linear model (Filev, 1991). Let the support be composed of r variables and be broken into c fuzzy classes (regions). Let $x = (1, x_1, \ldots, x_r) \in \mathbb{R}^{r+1}$ be a set of features used for clustering and $z = (z_1, \ldots, z_q)$ be a set of controls with coefficients that are common to all classes but not specific to clusters. Then for class i, the specification goes as

$$\text{IF } x \in \mathcal{X}^{(i)} \text{ THEN } \hat{y}^{(i)} = xa^{(i)} + zb, \quad i = 1, \ldots, c, \tag{5}$$

where $a^{(i)} = (a_0^{(i)}, a_1^{(i)}, \ldots, a_r^{(i)})'$ and $b = (b_1, \ldots, b_q)'$. All the rules can be aggregated to find the outcoming prediction of y as

$$\hat{y} = \sum_{i=1}^{c} \mu^{(i)}(x)\hat{y}^{(i)}. \tag{6}$$

Because of the innumerability of the membership functions, a finite parametrization of those functions is required to make identification feasible. The parametrization used here is due to Bezdek (1981) and characterizes the membership function in each class i by a center $v^{(i)} = (1, v_1^{(i)}, \ldots, v_r^{(i)})'$. The level of membership for a point x is assigned according to its relative distance to each of the centers, $v^{(i)}$, in the following way

$$\mu^{(i)}(x) = \frac{||x - v^{(i)}||^{-m}}{\sum_{k=1}^{c} ||x - v^{(k)}||^{-m}}, \qquad m > 0, \tag{7}$$

where $||.||$ is any norm in $\mathbb{R}^{r+1}$. $v^{(i)}$ represents the center of gravity for class i where points close to $v^{(i)}$ assume a membership of one in $\mathcal{X}^{(i)}$, and getting farther away from the center causes the degree of membership in $\mathcal{X}^{(i)}$ to fall. Points in the middle of $v^{(i)}$ and $v^{(k)}$, in principal, have their membership split between $\mathcal{X}^{(i)}$ and $\mathcal{X}^{(k)}$. Parameter m determines the level of fuzziness embedded in the description of classes. With large values of m, small deviations from the center cause an immediate drop in membership, so that membership functions will have steeper and thinner tails. With small m, on the other hand, the tails get longer and thicker and more fuzziness follows (Figure 3). In the limit, when $m = 0$, each

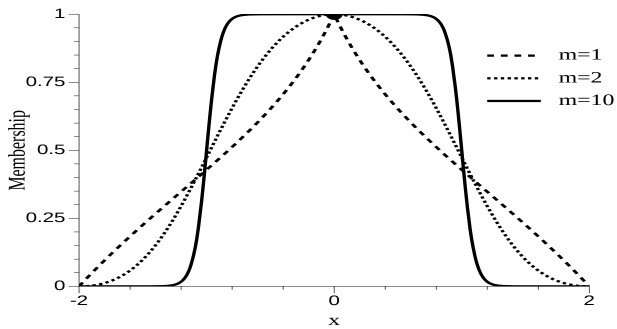


Figure 3: Sample membership functions of the form (7) with different values of m.

point in the space is equally a member of all classes.

When n independent observations are available, the goal of QLFM is to identify $\{v^{(i)}\}$, $\{a^{(i)}\}$, and b that minimize the following least-squared error criteria:

$$J = \sum_{j=1}^{n} w(j)\big(y(j) - \hat{y}(j)\big)^2 = \sum_{j=1}^{n} w(j)\left(y_j - \sum_{i=1}^{c} \mu^{(i)}\big(x(j)\big)x(j)a^{(i)} - zb\right),$$

where $y(j)$ and $x(j)$ denote the jth observation, and $w(j)$ is the corresponding sample weight. In a more compact form

$$J = (Y - XA)'W(Y - XA), \tag{8}$$

where

$$Y = (y(1), \ldots, y(n))', \quad A = (a^{(1)}, \ldots, a^{(c)}, b)',$$

and

$$X = \begin{bmatrix} \mu^{(1)}\big(x(1)\big)x(1) & \mu^{(2)}\big(x(1)\big)x(1) & \ldots & \mu^{(c)}\big(x(1)\big)x(1) & z(1) \\ \mu^{(1)}\big(x(2)\big)x(2) & \mu^{(2)}\big(x(2)\big)x(2) & \ldots & \mu^{(c)}\big(x(2)\big)x(2) & z(2) \\ \vdots & \vdots & & \vdots & \vdots \\ \mu^{(1)}\big(x(n)\big)x(n) & \mu^{(2)}\big(x(n)\big)x(n) & \ldots & \mu^{(c)}\big(x(n)\big)x(n) & z(n) \end{bmatrix}.$$

W is a diagonal matrix with diagonal elements set to $w(j)$'s. Note that, in the above, $x(j)$ is a row vector and $\mu^{(i)}\big(x(j)\big)$ is a scalar, so that $X : n \times \big((r+1)c + q\big)$.

In most cases, there is an acceptable range for the identified centers. If so, the following constraints are enforced:

$$\underline{v}_k^{(i)} \leq v_k^{(i)} \leq \bar{v}_k^{(i)}, \quad k = 1, \ldots, r,\ i = 1, \ldots, c. \tag{9}$$

For the ease of application, the problem of minimizing J within the constraint bounds of (9) can be integrated into minimizing the following augmented objective function

$$\tilde{J} = (Y - XA)'W(Y - XA) - \lambda \sum_{i=1}^{c} \sum_{k=1}^{r} \left(\log\left(\bar{v}_k^{(i)} - v_k^{(i)}\right) + \log\left(v_k^{(i)} - \underline{v}_k^{(i)}\right) \right), \tag{10}$$

where λ is a barrier parameter, normally chosen to be a very small real number. As a result, $\tilde{J} \simeq J$ within the interior of the feasible set. But, $\tilde{J}$ explodes closer to the boundaries of the feasible set, so that the solution is kept away from the boundary lines.

To tackle this problem, a gradient method similar to that of Yager & Filev (1993) is used. To begin with, take the gradients in (8) with respect to $\{v^{(i)}\}$ and $\{a^{(i)}\}$ to get

$$\partial\tilde{J}/\partial A = -2X'W(Y - X'A), \tag{11}$$

$$\partial\tilde{J}/\partial v_k^{(i)} = -2(Y - XA)W \left[\frac{\partial X}{\partial v_k^{(i)}} \right] A + \lambda \left(\frac{1}{\bar{v}_k^{(i)} - v_k^{(i)}} - \frac{1}{v_k^{(i)} - \underline{v}_k^{(i)}} \right), \tag{12}$$

$$k = 1, \ldots, r, \qquad i = 1, \ldots, c.$$

For an interior solution, the gradients above should be set to zero, but the second set of equations do not easily resolve into a closed form solution. Instead, the solution is found by recursively applying the following

$$A = (X'WX)^{-1}(X'WY), \tag{13}$$

$$v^+ = v - \alpha\Delta v, \quad \Delta v = H^{-1}\nabla\tilde{J}, \tag{14}$$

in which plus superscript indicates the next round iteration and

$$v = \begin{pmatrix} v_1^{(1)} & v_2^{(1)} & \ldots & v_r^{(1)} & \ldots & v_1^{(c)} & v_2^{(c)} & \ldots & v_r^{(c)} \end{pmatrix}'$$

$$\nabla\tilde{J} = \begin{pmatrix} \frac{\partial\tilde{J}}{\partial v_1^{(1)}} & \frac{\partial\tilde{J}}{\partial v_2^{(1)}} & \ldots & \frac{\partial\tilde{J}}{\partial v_r^{(1)}} & \ldots & \frac{\partial\tilde{J}}{\partial v_1^{(c)}} & \frac{\partial\tilde{J}}{\partial v_2^{(c)}} & \ldots & \frac{\partial\tilde{J}}{\partial v_r^{(c)}} \end{pmatrix}'.$$

H is a symmetric positive-definite matrix, conventionally set to the Hessian matrix where possible, while the non-convex nature of the problem might prohibit the use of Hessian in some regions of the feasible set.

Condition (13) is clearly the OLS solution when the membership weights are fixed. Condition (14) ensures that v is moved in the direction of reducing $\tilde{J}$, and when H is set to the Hessian matrix and the Hessian is positive definite in the locality of v, the step becomes the Newton recursion formula and moves the center coordinates in the direction of steepest descent. α is the length of the step to be taken in that direction,

3.3 Numerical Implementation

The recursive solution to (13) and (14) is found by applying a Quasi-Newton method in which gradients are computed analytically, but the Hessian matrix is estimated in each iteration using the BFGS update (Nocedal & Wright , 1999). Appendix C details the computation of gradients and the Hessian. The author is also aware of the ill-conditioning that arises in the estimates of Hessian when iterates get too close to the solution in barrier methods in general. However, the severity of numerical errors introduced because of this ill-conditioning is shown to be very benign (Wright, 1998).

The step size, α, is selected subject to the following constraint (Bakhtiari & Tits, 2003):

$$\tilde{J}^{+} < \tilde{J} - \xi\alpha(\nabla\tilde{J})'\Delta v, \tag{15}$$

In (15), ξ is normally chosen a small number and is to ensure that iterates do not get too close to the boundaries of the descent region, otherwise next iterations can slow down to a crawl by forcing α very close to zero. The condition also leads to a monotonic reduction in the objective function at each iteration, which is used in the proof of convergence.

At each iteration, α is adjusted recursively using a simple trust region method. Specifically, when taking a step with size α satisfies (15), trust builds and step size is made larger in the next iteration. However, if taking a step with size α violates (15), the region is assumed uncharted territory and α is repeatedly reduced and tested in (15) until a value of α is found to satisfy the condition. In the course of next iterations, an understanding of the whereabouts is established, and step size is expanded until another hitch is encountered. Compared to typical line search procedures, this mechanism is simpler and requires much fewer evaluations of the objective function per iteration, saving time and still offering fast convergence rate. As a rule of thumb, and through some experimentation, the following adjustment procedure is put into use:

$$\alpha^+ = \begin{cases} \max\{1.25\alpha, \bar{\alpha}\}, & \text{If (15) satisfied,} \\ 0.5\alpha, & \text{Otherwise.} \end{cases} \tag{16}$$

The reduction mechanism above, applicable when condition (15) is not satisfied, basically simulates an Armijo (1966)-type line search. Putting it all together, the algorithm to solve (13) and (14) would be:

Algorithm 1:

1. Initialization: Choose ϵ as the acceptable levels of tolerance for $||\Delta v||$. Choose ξ, $\bar{\alpha}$, the initial value of α, and $\{v^{(i)}\}$.

2. Updating Coefficients: Compute A from (13) and find $\tilde{J}$.

3. Updating Centers: Update the location of centers using (14).

4. Line search:

   - If α achieves (15) in the first try, then scale up α according to (16).

   - Otherwise, scale down α according to (16) and repeat steps (3) and (4) until (15) is satisfied.

5. Termination:

   - If $||\Delta v|| < \epsilon$ then stop.

   - Otherwise, repeat steps 2 to 4.

■

Theorem 1 *If H is a positive-definite symmetric matrix, then Algorithm 1 converges to a local solution characterized by conditions* (11) *and* (12).

Proof: See Appendix A.

An OCTAVE 3.2.4 implementation of the QLFM has been tested on a PC with a 3GHz Intel Core Duo processor, 4G RAM, running on Windows XP operating system. The parameter values are set to $\epsilon = 10^{-6}$, $\xi = 10^{-4}$, $\lambda = 10^{-3}$, $\alpha_0 = 1$ and $\bar{\alpha} = 1$. The initial guess for centroids, $\{v^{(i)}\}$, is generated using a simple fuzzy c-mean algorithm (Bezdek, 1981). The c-mean algorithm applies a simpler version of the QLFM, without the linear model, to identify the class centers solely based on the proximity of points (Appendix B).

4 Data

Data is obtained from the Confidentialised Unit Record File (CURF) version of the Business Longitudinal Survey (BLS) provided by the Australian Bureau of Statistics (ABS). The BLS is composed of four waves of the Business Growth and Performance survey conducted in (fiscal) years 1994–95 to 1997–98, henceforth referred to by the ending year. The unit of observation is a firm defined as a management unit which is “the highest level accounting unit within a business”. By definition, a management unit could be composed of several locations. For the sample of 1995, about 13,000 firms are randomly selected from the Australian Business Register; 8,375 of those firms are kept in the CURF version. The selection covers several broad industries, such as mining, manufacturing, construction, etc., and each firm is weighted to make the number of businesses representative within the corresponding industry×size stratum (Will & Wilson, 2001). In the next years, only 4,543 firms from the initial sample of 1995 are included in the CURF and surveyed. About half of the continuing sample is categorized for having shown innovation, export, or growth activity and the remaining half are selected from the remaining firms. The continuing sample in each year is also supplemented with about 450 new firms to compensate for the exits and non-responses. Sample weights are re-adjusted according to the new stratification accounting for innovation, export, and growth, in addition to size and industry.

In the BLS, each firm reports the number of salary-earning non-managerial employees who are hired on full-time ($FTWORK$) or part-time ($PTWORK$) basis in the pay period ending in June 30 of each year. Firms also report the number of managers (MAN) at the same time point. The BLS defines managers as “those who are in charge of a significant number of employees or who have significant responsibilities in the conduct or operations of business”

(ABS Cat.No. 8141.0.15.001). The number of managers excludes working owners, which are reported separately (OWN). Up to 1996, the number of managers and working owners reported does not distinguish between full-time and part-time positions, while this distinction is made in the last two years of the panel. However, less than 3% of firms report having only one or two part-time managers or owners in those years. To maintain the consistency of counts across years, and without much fallout in accuracy, all managers and working owners in each firm are assumed full-time regardless. As a result, the full-time workforce of a firm is set to the total count of managers, working owners, and full-time non-managerial employees, or

$$FTEMP = MAN + OWN + FTWORK.$$

A measure of total employment (EMP) is built by combining full-time and part-time employment in each firm. For this purpose, the proportion of part-time to full-time hours is obtained from the ABS report on earnings and hours (Cat.No.6306.0) and is used to find the *full-time equivalent* of part-time employment. Total employment is the sum of full-time employment and the prorated part-time employment, or

$$EMP = FTEMP + p \times PTWORK,$$

where p typically ranges between 0.43 to 0.45 in different years.

Finally, firms in the BLS report their number of business locations ($NLOCS$) in June of each year. In 1995, firms are asked to report the number of employees and managers in June 1994 and the number of locations they opened and closed since. This latter information is used to find the number of locations in June 1994; thus year 1994 is also added as a usable set of data.

For this study, all active firms in each year that reported positive total employment and sales during the year are selected. For confidentiality protection, the ABS drops all firms with more than 200 full-time employees. Therefore, the reader should bear in mind that the largest firms in the data are just relatively large. Also, the omission of firms larger than 200 employees does not undermine the quality of analysis done in this study, as the ABS business counts show that less than 1% of Australian businesses are within that size range.

ANZSIC	Description	1994	1995	1996	1997	1998
1	Mining	28	46	40	46	48
2	Manufacturing	2,354	2,765	1,685	1,649	1,632
3	Construction	293	351	219	214	230
4	Retail Trade	810	974	681	692	688
5	Wholesale Trade	689	823	482	472	487
6	Hotels, Cafes & Restaurants	197	285	189	176	187
7	Transport & Storage	213	264	169	172	175
8	Finance & Insurance	139	169	129	122	131
9	Property & Business Services	771	946	551	521	538
10	Cultural & Recreational Services	65	146	90	83	92
11	Personal & Other Services	147	180	88	93	100
	All	5,707	6,959	4,328	4,246	4,313
	Total			25,553		

Table 2: The size of sample by year and industry.

In the introduction to size classification, managers are portrayed as supervisors. To better match the data to this description of management role, all firm-years that show an average number of managers per total employment of one or above are dropped (a total of 620 firm-years). This omission guarantees that the average sphere of control, the number workers under the direct supervision of a manager, is more than one employee. Moreover, any firm-year in the data that has more than 40 managers is a gross outlier. In industries Transport & Storage (ANZSIC 7) and Cultural & Recreational Services (ANZSIC 10), any firm-year with the number of managers more than 25 is a gross outlier. These observations are also dropped (a total of 40 firm-years). Table 2 reports the selected sample size for different years and industries. For most industries, the sample size drops after 1995 due to the ABS sample selection explained earlier.

Table 3 reports the simple statistics for the set of variables used in the empirical exercises of the next sections. Variable $PART$ is the proportion of part-time non-managerial employees; put formally

$$PART = PTWORK/(PTWORK + FTWORK).$$

It is also noted that all of these variables are, to some degree, correlated with each other (Table 4). Expectedly, larger firms have more locations, more full-time, and more part-time employees. The proportion of part-time employees, nevertheless, is lower among larger firms.

Variable	Mean	Std.Dev	1st Decile	Median	3dr Decile
EMP	25.7	31.0	3.0	13.0	67.1
MAN	2.4	4.0	0	1.0	7.0
$FTEMP$	23.6	30.2	2.0	11.0	63.0
$PTEMP$	4.7	12.4	0	1.0	12.0
$NLOCS$	1.7	2.6	1.0	1.0	3.0
$PART$	0.26	0.35	0	0.06	1.0

Table 3: Simple statistics for different variables. The full sample of 25,553 firm-years from the BLS is used. Sample weights are not applied.

	EMP	MAN	$FTEMP$	$PTEMP$	$NLOCS$
MAN	0.720				
$FTEMP$	0.984	0.706			
$PTEMP$	0.219	0.173	0.043		
$NLOCS$	0.326	0.359	0.308	0.147	
$PART$	-0.272	-0.195	-0.353	0.414	-0.054

Table 4: Table of correlations. Using the full sample of 25,553 from the BLS. Sample weights are not applied.

At the same time, part time employment has strong ties with the industry of operation; For instance, Hotels, Restaurants, and Cafes (ANZSIC 6) depend heavily on part-time employment, but Mining (ANZSIC 1) and Manufacturing (ANZSIC 2) mostly rely on full-time workers.

5 A Fuzzy Model of Employment Size

As brought up earlier, in a discrete model of size, firms of types small and large demonstrate distinctive management organizations. The distinction is used to divide firms into groups of observations in such a way that firms belonging to each group follow a different management-employment log-linear relation, as depicted in (1). It is understood that the number of managers in a firm can also be influenced by the abundance of part time workers among the firm's workforce and the number of locations run by the firm, not the least the industry of operation. Accounting for all these factors, a proper QLFM to describe firm size is

$$\begin{aligned}\text{IF } x \in \mathcal{X}^{(i)} \text{ THEN } \widehat{MAN}^{(i)} =& a_0^{(i)} + a_1^{(i)} \log(EMP) + b_1 \log(NLOC) + b_2 PART \\ &+ \sum_{l=1}^{10} c_l ANZSIC_l, \quad i = 1, \ldots, c.\end{aligned}$$

In the above, $NLOC$ controls for the number of business locations and $PART$ controls for the proportion of part-time employees within the firm's workforce. $ANZSIC_l$ controls for the industry effects. The total employment described in Section 4 serves here as the numerical measure of size. In view of Figure 1(a), the relation for both classes is specified as a log-linear one. The following feasibility constraints are also enforced:

$$-1 \leq v^{(i)} \leq 10, \ i = 1, \ldots, c,$$

where the upper and lower bounds are in the logs of employment. The smallest firm in the data has about 0.4 employees (employs only one part-time worker) and can be definitely classified as a small business. The largest firm in the data has around 200 employees, but data is censored and there are larger firms in Australia. Therefore, at this point, what constitutes a large business is not certain, and for that reason the feasible region is kept practically unbounded from above.

5.1 Model Selection

At this stage, the modeler enjoys freedom in choosing two parameters in the QLFM, namely, c (the number of categories), and m (the degree of fuzziness in membership functions). There are some practical considerations, especially, when choosing a value for m. A very small m makes matrix X ill-conditioned and hampers convergence. A very large value for m, on the other hand, makes the membership functions very sharp-edged, and the gradients and Hessian matrix will explode, again, causing numerical instability. In choosing the number of classes, c, over-estimation causes the same issues as choosing a large m; it makes the centers of membership functions fall too close to each other, making the membership functions very sharp-edged.

In choosing m, the literature on fuzzy logic is traditionally fixated on choosing $m = 2$ without offering a theoretical or experimental background. Here, I proceed by examining the Akaike Information Criterion (AIC) for each pair of (c, m).[2] The number of clusters tested

[2]In the data, the number of managers has a monotonically increasing variance by employment size. To account for this heteroscedasticity in the AIC, $E[MAN|EMP]$ and $E[MAN^2|EMP]$ are estimated non-parametrically by kernel regression with a Gaussian kernel and using the bandwidth of 1.0 for log employment. Bayesian Information Criterion is also computed, but the values almost duplicated those of AIC, therefore, they are not reported.

m	$c=2$ #Itr	AIC/n	CSI	$c=3$ #Itr	AIC/n	CSI	$c=4$ #Itr	AIC/n	CSI
2	36	105.0	1.23	48	103.4	1.26	118	103.3	8.05
5	33	119.2	-0.10	17	112.6	1.02	26	106.9	1.41
10	107	105.3	3.24	22	108.1	1.23	29	104.4	2.11
15	21	153.9	0.74	36	107.8	1.34	48	104.4	2.16
20	21	153.5	0.88	47	107.7	1.51	40	107.5	2.25
50	35	147.1	1.33	47	107.8	2.28	non-convergent		

Table 5: Convergence and information criteria for different combinations of c and m.

for this application is $c \in \{2, 3, 4\}$, and each one is tried in combination with m assuming integer values of two and above.[3]

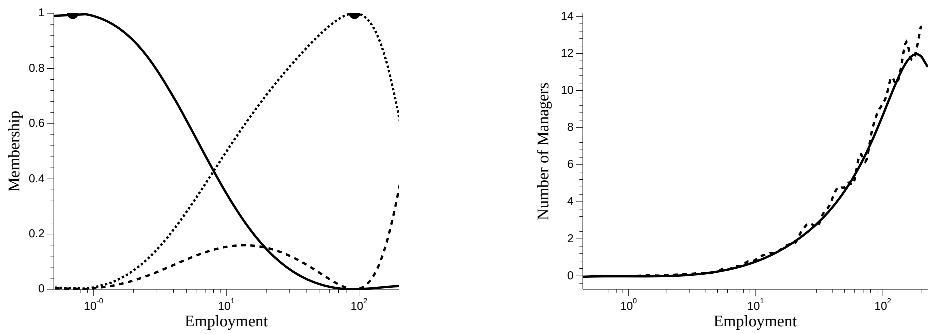
The AIC measure, however, only provides information on the quality of fit for parametric models and lacks any provision to penalize the misallocation of clusters. In the clustering literature it is customary to form measures of cluster similarity by forming euclidean norm of the distance between centroids and opt for those allocation where clusters are distant enough (Davies & Bouldin, 1979). Here, I introduce a Cluster Similarity Index (CSI) whose main purpose is to keep centroids as far apart as possible. The CSI is defined as

$$CSI = -\log\left(\frac{\min_{i,k,i\neq k} ||v_i - v_k||}{\max_i ||v_i||}\right). \tag{17}$$

In the index above, the denominator represents the range, and the index returns large values when any two centroids are relatively positioned too close to each other. The QLFM has an inherent tendency to put centroids very close, practically on top of each other, when there is over-estimation, that is, more classes are requested than required. Using the CSI along with the AIC enables me to choose a model with the best parametric fit, but also the one that does not overspecify. Table 5 reports convergence properties as well as the AIC and CSI for each pair of (c, m). The AIC values are divided by n (the number of observations) to reduce the range of values reported.

Based on the AIC performance, the table indicates that the most preferred (parametric) models require $m = 2$. The emergence of $m = 2$ as part of the best fit particularly redeems

[3]The application of Multivariate Adaptive Regression Splines of Friedman (1991) to the same data identifies four classes. The regression spline approach is a piecewise-linear identification of the relation. Since the QLFM is a nonlinear smooth fit, then it is able to use fewer classes to achieve a similar or even better fit. As a result, choosing $c = 4$ as the upper bound seems appropriate.



(a) (b)

Figure 4: The identified fuzzy membership functions and the quality of fit by the QLFM against that of the non-parametric relation. Using $m = 2$ and three clusters. Bullets are the identified centers within the data range. The sample used is 25,553 firm–years from the BLS.

the previous application of the same value all across the relevant literature. Also, according to the AIC measure, the best parametric fits require either three or four clusters (the AICs are not that different). The resolution comes from investigating the CSI, which points to a possible over-specification when four clusters are involved. Based on the evidence, I proceed by using $m = 2$ and $c = 3$. Figure 4 demonstrates the allocation of the membership functions and the quality of fit using the parametrization just described. Note that the third center identified by the QLFM does not fall in the data range and is not shown.

5.2 Identification Results

The choice of $m = 2$ and $c = 3$, nevertheless, leads to some interesting and unexpected findings. To supplement the graphical presentation, Table 6 reports the estimated coefficients for each fuzzy class. Based on prior expectations, the first two classes associate with "small" and "large" businesses. In particular, firms belonging to the first class (small business) practically hire no managers, whereas firms belonging to class two (large business) hire about seven managers per one unit increase in log employment.

The third class helps trim the relation into a close fit and points to a type of firm that hires fewer managers for larger employment. There is a weak element of this relation that affects firms that are primarily around 10 to 20 employees. This *weak presence* suggests the existence of an evolutionary or transitional force which affects firms passing through the size

$FX^{(i)}$	$v^{(i)}$	$a_0^{(i)}$	$a_1^{(i)}$	b_1	b_2
Small (S)	0.70	-0.017	0.026		
		(0.033)	(0.048)		
Large (L)	92.49	-24.128***	7.120***	0.542***	0.013
		(0.033)	(0.048)	(0.046)	(0.020)
Transition (T)	548.69	45.977***	-6.942***		
		(3.102)	(0.655)		
Average Sum of Squared-Errors		1.18			
Run Time		7.5sec			

Table 6: The identified relation for each fuzzy class. Numbers in parentheses are robust standard errors. *** indicates significance at 1% level. The sample is 25,553 firm-years from the BLS.

of 10 and prepares them to switch to a large business model. The transition effect starts to wear off as firms pass the size of 20; the large business model gradually takes control thereafter. A part of this transitional force is also caused by the heterogeneity among firms and their timing to adopt a more sophisticated management structure. This is especially true for firms hiring up to 10 employees, for which Figure 1(a) suggests that some of those firms have already hired managers while others are coping all by themselves.

Based on the observations made above, I will call the fuzzy description of the third class the "transitional element". There is also a strong association of the same class with the extreme upper tail of size, but that part is created by the algorithm trying to get rid of the peak of the membership function by pushing the center out of range. Further supporting evidence for this claim is presented in Section 6, where a mixture model is estimated to the same effect.

The estimates are also controlled for part-time employment and the number of locations, and the model can be further validated by looking as those effects. The expectation is that having more business locations makes the coordination task more complicated and demands more management resources. The estimated effect has the correct sign with statistical significance: increasing the number of locations by two log units demands on average one more manager.

With part-time employment, the expectation is mixed. On one hand, having a higher proportion of part-time employees on payroll reduces the management load when part-time and full-time employees are perfect substitutes. On the other hand, if part-time workers in large firms specialize in performing distinct tasks, then the management load might even increase to cope with the coordination of the added tasks. The estimated coefficient in Table 6 is economically small and statistically insignificant and not leaning towards either explanation. One major reason for this finding is the strong correlation between industries and their use of part-time workers, so that most of the part-time effect is already absorbed by the industry effect.

5.3 Policy Analysis

In the wake of the Global Financial Crisis of 2007, the Australian parliament passed the *Nation's Building and Jobs Plan* Act to preempt a recession through providing a sizable stimulus for the nation's economy. Part of the act is dedicated to growth and job creation by offering small businesses investment incentives. Specifically, small businesses can receive up to 50% tax credit for their "business investment in new tangible depreciating assets and new expenditure on existing assets". For this purpose, the ATO defines a small business broadly as the one with less than \$2 million in total revenues for the previous income year and expecting to make less than \$2 million over the current income year. Larger businesses, on the other hand, can apply to receive a 30% tax credit or a 10% tax credit under certain eligibility conditions. Are transitioning firms receiving full support by this policy as they evolve into a large business?

Before any conclusions, one ought to characterize the typical characteristics of a transitioning firm. A fuzzy specification of a transitioning firm takes into account the fact that a firm in transition should lie on the (fuzzy) intersection of the two classes of small and large businesses, and also be affected by the transitional element. In terms of fuzzy set operations, the membership of firm–year j in the transitioning group can be defined as

$$\bar{\mu}_j = \left(\mu_j^{(S)} \wedge \mu_j^{(L)}\right) \vee \mu_j^{(T)},$$

where S, L, and T stand for small, large, and transitioning, respectively. In forming the membership grades, the upper tail of the membership function for transitional element (T) is intentionally dropped; it clearly does not pertain to size evolution. Employment for a typical transitioning firm can then be found using the center of gravity procedure (Zimmermann, 1996):

$$\overline{EMP} = \frac{\int_{0.4}^{200} e\bar{\mu}\big(\log(e)\big)de}{\int_{0.4}^{200} \bar{\mu}\big(\log(e)\big)de}. \tag{18}$$

The lower bound of the integral clearly has to be nonzero, hence, is set to the smallest firm (with on part-time employee). The upper bound is set to the largest firm observed in the data. Other characteristics for the typical transitioning firm are found using a kernel averaging:

$$\overline{Sales} = \frac{\sum_{j=1}^{n} w_j K(EMP, \overline{EMP}) Sales}{\sum_{j=1}^{n} w_j K(EMP, \overline{EMP})}, \tag{19}$$

in which j indexes firm-years. In this application, the kernel, $K(.,.)$, is a Gaussian function in the log of employment with mean $\log(\overline{EMP})$ and standard deviation 0.05. Other characteristics, such as the number of managers or locations, can be found in the same way by replacing Sales with that variable. A typical small and large firm is also defined in the same manner but by replacing $\bar{\mu}$ with $\mu^{(S)}$ and $\mu^{(L)}$, respectively.

Table 7 lists the typical characteristics of transitioning as well as small and large firms computed by the method above. Job creation and destruction are defined in the same way as in Davis et al. (1996). Specifically, job creation is the change in the number of personnel (full-time, part-time, manager/owner) from June of one year to the next for growing businesses averaged over all businesses in the same group. Job destruction is the change but for contracting businesses.

According to the table, a typical transitioning firm in Australia makes about $2.6 million in sales (2011 dollars), which is particularly above the ATO's income cap for being fully supported by the policy.Therefore, the tax benefit is cut short too soon, before businesses reach too far into the transition. The average transitioning firm also employs about 11 full-time employees, one of which is a manager. Most of these firms operate from one location and close to one-quarter of their workforce is hired on part-time basis. The net job creation by transitioning firms is about 1.4 jobs per business, twice as much as the 0.7 net jobs created by each small business on average but much below the net job creation by larger firms. Reader should also be warned that, owing to the censorship of data by the ABS, the characteristics of large firms are under-estimations.

With this preliminary, the implications of the ATO's 2011 tax code are now better un-

Variable	Small	Transition	Large
Total Employment	3.80	11.0	58.1
Sales ($000)	788.6	2,644.2	22,214.5
Managers	0.14	1.08	4.94
Locations	1.08	1.28	2.41
Fraction of Part-times	0.37	0.24	0.13
Job Creation	1.04	2.24	14.6
Job destruction	0.31	0.84	2.45
Age	9.9	12.6	15.2

Table 7: The typical characteristics of the transitioning firm as well as small and large firms. Sales are in 2011 dollars. The sample is 25,553 firm-years from the BLS.

derstood. The businesses being supported by the tax policy are dominantly small businesses in Australia, for which the tax credit is apparently intended. Businesses that venture into transition with the intention to turn into a large business quickly lose their tax privilege and are practically left to their own devices for the rest of the way to evolution. As a result, the achievements of the the aforementioned tax code are insofar protecting small businesses and shielding them from unfavorable market conditions.

But, as Table 7 shows, job creation begins to take an accelerating pace once firms move through the transition phase and peaks among larger firms. With growth and job creation as a central policy, a support line for the transitioning firms seems necessary. In one instance, Bakhtiari (2011) shows that many firms lose momentum passing through the transition phase and downsize again; the availability of tax credit could have helped these firms to keep the momentum and grow into a large business. As a result, there will be a hoarding of businesses that mature without growing and do not contribute much to the process of job creation (Haltiwanger et al., 2010). With the average age of a small business estimated at 10 years in Table 7, the number of mature small businesses seems to be substantial, and there is a huge job creation potential if these firms ventured into transition with the support of government. A simple fix in this line would be to raise the eligibility cap to $3 million, so that businesses are being supported in their early to mid stages of evolution (assuming firms are comfortable to be on their own in the advanced stages of evolution).

Yet, a more sophisticated and effective approach would be to offer firms incentives to actually step into transition. Such tax system would be composed of three phases:

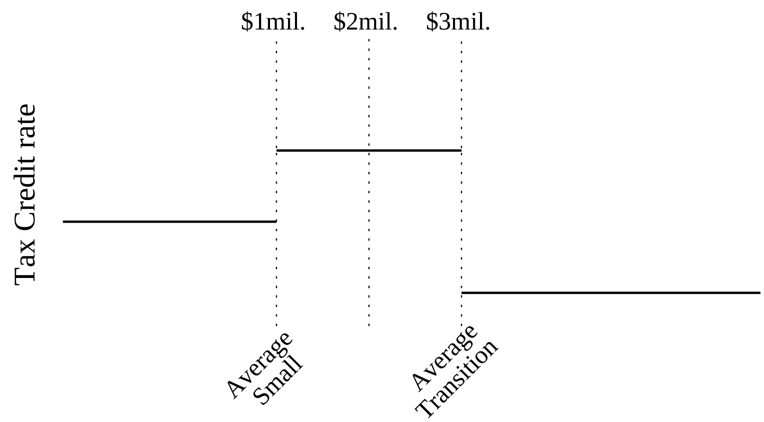


Figure 5: A proposed business tax code targeted at growth and job creation.

- Very small businesses receive some tax credit to protect them and shield them against macro and micro shocks until they can realize their full performance potentials.
- Larger small firms and firms in the early to mid stages of transition receive even larger tax credits so that they can dedicate extra resources into size evolution.
- Large firms and firms in the late stages of transition are taxed at a regular rate.

Figure 5 is s depiction of this policy. Offering bigger incentive to transitioning firms in particular stops businesses from deliberately curbing their growth so that they can constantly receive special treatment. Instead, it offers them a larger prize if they really create jobs.

Here the benchmark is set to Australia, owing to the availability of very detailed data on the composition of employees. However, one might be able to say something about other countries if one assumes that businesses are by and large managed the same way in Australia and most other industrialized countries. In the United States, the IRS considers businesses less than 25 employees as small. This threshold seems to be hitting the right cord as it covers the average transitioning firms and many more firms in their advanced stages of evolution. A threshold of 50 employees is too high, and the results of Section 7 will further confirm that many large businesses, by the management definition presented here, will be included. The same tri-stage tax policy is also applicable in all these case, but of course adjusting for the level of revenue and currency.

6 QLFM versus Mixture Model

The analysis of the previous section depends on how classes small, large and transitioning are defined. A robustness test would be to estimate the same classes using an independent alternate method. Mixture Models (McLachlan & Peel, 2000) work to the same effect by estimating size dependent probabilities that a business belongs to any class of interest. Principally, it assumes that firms are randomly picking a business model upon observing their size and other relevant performance measures with some prior probability model. To make results comparable, three types (Small, Transitioning and Large firms) are included. The prior probabilities of a firm picking each type follow

$$\begin{aligned}
\theta_S(EMP) &= \frac{e^{\left(d_{S0}+d_{S1}\log(EMP)\right)}}{1+e^{\left(d_{S0}+d_{S1}\log(EMP)\right)}+e^{\left(d_{T0}+d_{T1}\log(EMP)\right)}},\\
\theta_T(EMP) &= \frac{e^{\left(d_{T0}+d_{T1}\log(EMP)\right)}}{1+e^{\left(d_{S0}+d_{S1}\log(EMP)\right)}+e^{\left(d_{T0}+d_{T1}\log(EMP)\right)}},\\
\theta_L(EMP) &= 1-\theta_S(EMP)-\theta_T(EMP).
\end{aligned} \tag{20}$$

By picking a certain business type, the number of managers in a firm is then determined from

$$\begin{aligned}
MAN_j = a_0^{(i)} + a_1^{(i)}\log(EMP_j) + b_1\log(NLOC_j) + b_2 PART_j + \sum_{l=1}^{10} c_l ANZSIC_l + \epsilon_j^{(i)},\\
f^{(i)}(\epsilon^{(i)}) = \frac{1}{\sqrt{2\pi\sigma^{(i)}}} e^{-\frac{1}{2}(\epsilon_j^{(i)}/\sigma^{(i)})^2}, \quad i=S,T,L.
\end{aligned} \tag{21}$$

The mixture model described above can be estimated by maximizing the following log-likelihood function

$$\begin{aligned}
\log(L) = \sum_{j=1}^{n}\log\Big(\theta_S(EMP_j)f^{(S)}(\epsilon_j^{(S)}) + \theta_T(EMP_j)f^{(T)}(\epsilon_j^{(T)})\\
+ (1-\theta_S(EMP_j)-\theta_T(EMP)f^{(L)}(\epsilon_j^{(L)})\Big)
\end{aligned} \tag{22}$$

For the application, however, the type of business model each firm picks is unobserved, but the posterior probability can be easily deducted by applying the Bayes rule. Then the

Variable	$MAN^{(S)}$	$MAN^{(T)}$	$MAN^{(L)}$	θ_S	θ_T
Const.	-0.001	0.608	-3.548***	10.992***	6.948***
	(1.034)	(0.608)	(5.127)	(2.024)	(2.030)
$\log(EMP)$	-0.000	0.396***	2.636***	-3.742***	-2.092***
	(0.770)	(0.124)	(1.205)	(0.656)	(0.644)
$\log(NLOC)$		0.003			
		0.155			
$PART$		0.000			
		2.035			
σ	0.1	0.611***	3.527***		
	(1.000)	(0.049)	(0.756)		
Average Sum of Squared-Errors			1.25		
Run Time			2m51.9s		

Table 8: The estimated mixture model with three normal components. Numbers in parentheses are standard errors. *** indicates significance at 1% level. The sample is 25,553 firm-years from the BLS. Sample weights are applied.

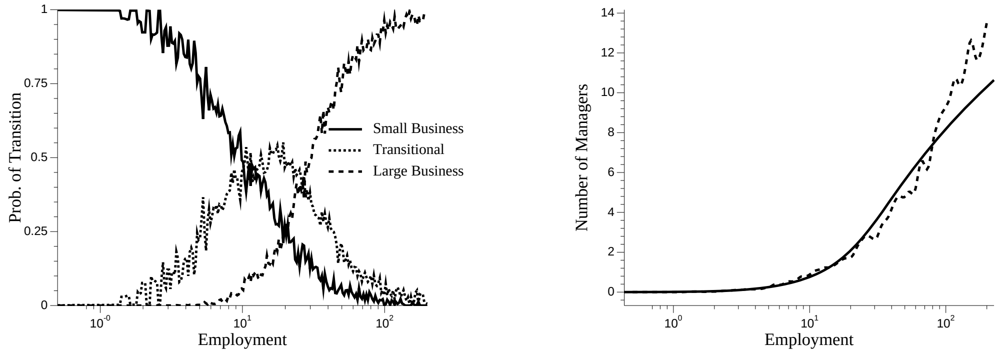
probability that a firm picked type $i = S, T, L$ given its observed characteristics is:

$$\begin{aligned} P_j^{(S)} &= \frac{\theta_S(EMP_j) f^{(S)}(\epsilon_j^{(S)})}{\theta_S(EMP_j) f^{(S)}(\epsilon_j^{(S)}) + \theta_T(EMP_j) f^{(T)}(\epsilon_j^{(T)}) + \theta_L(EMP_j) f^{(L)}(\epsilon_j^{(L)})} \\ P_j^{(T)} &= \frac{\theta_T(EMP_j) f^{(T)}(\epsilon_j^{(T)})}{\theta_S(EMP_j) f^{(S)}(\epsilon_j^{(S)}) + \theta_T(EMP_j) f^{(T)}(\epsilon_j^{(T)}) + \theta_L(EMP_j) f^{(L)}(\epsilon_j^{(L)})} \\ P_j^{(L)} &= 1 - P_j^{(S)} - P_j^{(T)}. \end{aligned} \tag{23}$$

The maximum-likelihood problem is estimated using the same numerical optimization procedure described in Section 3.3, and the results are reported in Table 8. Figure 6(a) shows the average of each posterior probability as a function of size, which bears stark resemblance to those estimated in Section 5 using the QLFM. In particular, the transitional probability, $P^{(T)}$, is doing a very similar function to what transitional membership is doing in the QLFM.

It is proper at this point to point out the differences between the QLFM and mixture model. First, there is a conceptual difference: mixture model assumes that each business has selected one particular business model, and we just do not observe which one. The QLFM, in contrast, works with the possibility that a business model governing a certain business is a combination of various models.

Moreover, in the QLFM, the membership functions act as filters and result in very smooth



(a) (b)

Figure 6: (a) The average probabilities using a non-parametric fit and (b) the quality of fit with respect to the non-parametric relation. The sample used is 25,553 firm–years from the BLS.

prediction of class memberships. In comparison, mixture models predict noisy probabilities: mixture model can assign two firms with the same size and characteristics very different probabilities in each class. The level of noise increases in transition areas.

Finally, the identification of mixture models is generally a difficult job, especially when the number of classes and independent variables increase. the complexity arises from the objective function which is a highly nonlinear and complicated log likelihood function. The QLFM instead relies on estimating a collection of linear models and then combining them nonlinearly using the membership functions.

7 A Threshold Model of Firm Size

Fuzzy membership offers flexibility in descriptions but sacrifices some precision when crisp decisions are needed; this is in fact the very intention of fuzzy logic. However, where crisp thresholds are needed, the QLFM can still be adapted to estimate threshold models. In this context, it helps to make very sharp statements about where transition begins and where it ends. It is also crucial to remind the reader that the QLFM cannot be used to estimate true threshold models; pushing m to infinity causes the gradients of the membership functions to go to infinity at the switch point and zero everywhere else (the same for the Hessian matrix), making convergence impossible. However, it is feasible to estimate near-threshold models

$FX^{(i)}$	$v^{(i)}$	$a_0^{(i)}$	$a_1^{(i)}$	b_1	b_2
Small	2.10	-0.104***	0.182***		
		(0.033)	(0.016)		
Transition	19.80	-2.350***	1.390***	0.556***	0.011
		(0.033)	(0.016)	(0.046)	(0.020)
Large	46.12	-15.900***	5.312***		
		(0.834)	(0.211)		

Table 9: The identified relation for each fuzzy class. Numbers in parentheses are robust standard errors. *** indicates significance at 1% level. The sample is 25,553 firm-years from the BLS.

with the QLFM using relatively large values of m, making the fuzzy regions between classes as narrow as possible.

In the QLFM of the previous section, the evolution of firms from small to large is governed by the combination of three distinct elements: small business model, large business model, and a transitional force. Alternatively, firms in transition can be granted their own category, which is useful in visualizing and analyzing the properties of firms in the process of size evolution. The trends in Table 5, in fact, show that choosing $c = 3$ and pushing m to values near 20 can generate fits comparable to, not if as good as, the fuzzy model of the previous section. These fits present a clear division of size range into "small", "transitioning", and "large" businesses.

Using $m = 20$ and three fuzzy classes, the QLFM is reapplied to the data, and Figure 7 illustrates the identified membership functions and the resulting quality of fit with this specification. Table 9 lists the estimated relations in each segment. The predicted fuzzy relation starts almost flatly at small sizes, slightly slopes upwards as the average number of managers in a firm is passing through one, and then slopes steeply upwards among larger firms. The membership functions are such that firms in each segment of the model are practically governed by only one model of management. It is more straightforward here to talk about the characteristics of large businesses; for instance, one unit increase in the log of employment requires the assistance of five additional managers (Table 9). In the previous model of fuzziness, it was not possible to make such sharp prediction without taking into account the interaction of all fuzzy classes.

Given the very narrowness of the fuzzy areas, it is also possible to pinpoint where the

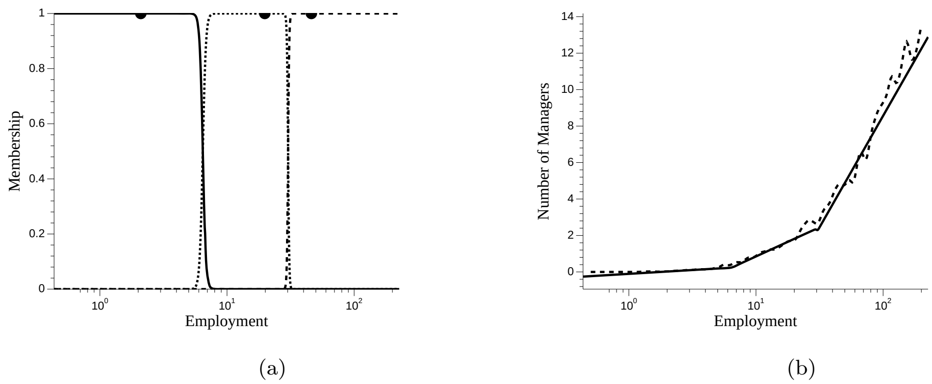


Figure 7: (a) The identified fuzzy membership functions and (b) the quality of fit with respect to the non-parametric relation with $c = 3$ ad $m = 20$. The bullets are the identified centers. The sample used is 25,553 firm–years from the BLS.

Type	Total Employment	Sales ($000)	Number of Managers	Fraction of Part-times	Number of Locations
Small→Transition	6.45	1,355.0	0.42	0.33	1.16
Transition→Large	30.2	9,075.1	2.64	0.20	1.64

Table 10: Typical characteristics of the firm crossing into and out of transition. Sales are in 2011 dollars. The sample is 25,553 firm-years from the BLS.

change from small to transitioning happens, and pinpoint where a transitioning business changes into a large business. Table 10 lists the relevant statistics using (19) and the relevant threshold points. Per these results, a firm starts transitioning with about six employees and evolves into a large business as soon as it hires 30 employees. In terms of annual sales, firms begin their transitioning surpassing the $1.4 million mark. Sales in large firms are expected to exceed $9 million.

In line with the findings of Section 5.3 the very early stages of transition are covered by the $2 million income cap set by the ATO, so that the supported firms are those that just started transitioning and have not yet reached the advanced stages of their evolution. The number of managers and locations clearly increase as firms slip into transition, progress and evolve into large size. The fraction of part-time workforce, however, drops in the process. The correlations in Table 4 make this last result expected.

8 Conclusion

The QLFM has a wide range of potential applications by offering a very conceptual yet useful classification of any economic performance measure. In the application presented, the QLFM shows some superiority towards its closest rival, mixture modeling, by estimating a filtered and "clean" classification system. The QLFM also converges faster in the applications, while it has the flexibility to estimate both smooth transition and threshold models by a mere change of parameter.

In this paper, the method is specifically applied to the classification of firm size. The versatility of the method is demonstrated by estimating both fuzzy and near-threshold models, each one shedding light on some aspects of firm's evolution. The fuzzy model, in particular, emphasizes the presence of three types of forces, with various degrees of relevance, affecting firms in transition: a small business element, a large business element, and a transitional force. The near-threshold model, then, pinpoints the location and the extent of this transition interval. The identification results suggest that the current $2 million income cap set by the ATO falls short of supporting growth especially among businesses that are still passing through their transitioning phase. Instead, a policy of growth and job creation is emphasized by suggesting a three-part system of tax credit, in which transitioning firms that have not reach the advanced stages of evolution receive the largest tax concessions. In this way, firms are lured into the transition area and forced to expand. Given the vastness of small mature businesses, there is huge potentials for job creation once such movement begins.

9 Acknowledgement

In writing this paper, I benefited from discussions with Arthur Ramer, Graham Elliot, Rachida Ouysse, and Olena Stavrunova, which are gratefully acknowledged.

A Proof of Theorem 1

First, some convergence properties are shown by the following lemma:

Lemma 1 *If $\nabla \tilde{J} \to 0$ and H is non-singular, then $v \to v^*$ and $A \to A^*$ for some v^* and*

A^*, where v^* satisfies the feasibility condition (9). In other words, every local solution is an accumulation point of the recursive series.

Proof: Let $t \in \mathbb{N}$ index the iterations, where $\mathbb{N}$ is the set of natural numbers, and let $\{v_t, \tilde{J}_t, \nabla \tilde{J}_t, H_t, \alpha_t\}$ be the state at each iteration. Since $||\nabla \tilde{J}_t|| \to 0$, there exists a subset of indexes $T \subset \mathbb{N}$ over which $||\nabla \tilde{J}_t||$ is monotonically decreasing. Also, notice that

$$||H^{-1}|| = \max_{u \neq 0} \frac{||H^{-1}u||}{||u||} \geq \frac{||H^{-1}\nabla \tilde{J}||}{||\nabla \tilde{J}||}.$$

Then, from (14), we can write

$$0 \leq ||v_{t+1} - v_t|| = \alpha_t ||H_t^{-1} \nabla \tilde{J}_t|| \leq \bar{\alpha} ||H_t^{-1}|| \; ||\nabla \tilde{J}_t|| \to 0, \quad t \in T.$$

As a result, $\{v_t\}_{t \in T}$ is a Cauchy sequence and has an accumulation point v^*. Condition (15) guarantees that $\tilde{J}$ is reduced in every iteration. Since $\lambda > 0$, all the iterations v_t, hence v^*, are kept within the interior of the feasible set. A^* is the solution to (13) with v^* used as centroids. Note that the boundedness of α is crucial to this proof. Also the positive-definiteness of H is not required for this part of the proof. Q.E.D

Proof of Theorem: First, note that the feasible set for the augmented problem is an open set and non-compact, so some extra care is needed for the proof. Since all $\underline{v}$'s and $\bar{v}$'s are finite, it is straightforward to show that

$$|\tilde{J}_t| \geq -2\lambda \log \left(\frac{\bar{v} - \underline{v}}{2} \right).$$

Therefore, with monotone decrease in $\tilde{J}_t$, it must be that $\tilde{J}_t$ is a convergent sequence with an accumulation point J^*. In other words, $|\tilde{J}_{t+1} - \tilde{J}_t| \to 0$ for t large enough. Now, using (15), we can write

$$0 < \xi \alpha (\nabla \tilde{J}_t)' H_t^{-1} (\nabla \tilde{J}_t) < |\tilde{J}_{t+1} - \tilde{J}_t| \to 0. \tag{24}$$

Similar to Bakhtiari & Tits (2003, Lemma 7), let $\inf ||\nabla \tilde{J}_t|| > \sigma_1 > 0$ for all $t > t_0 > 1$ and a contradiction will follow. First, it will be proved that $\inf ||H_t^{-1} \nabla \tilde{J}_t|| > \sigma_2$ for all $t > t_0$ and for some $\sigma_2 > 0$. H_t is positive-definite and symmetric, and so is H_t^{-1}. Hence, the eigenvectors of H^{-1}, $u_k,\ k = 1, \ldots, r \times c$, are an orthogonal basis for $\mathbb{R}^{r \times c}$ and there exist real numbers c_k so that

$$\nabla \tilde{J} = \sum_{k=1}^{r \times c} c_k u_k,$$

Let $\gamma_k > 0$ be the corresponding eigenvalues, then

$$\begin{aligned}
\inf ||H^{-1} \nabla \tilde{J}|| = \inf ||H^{-1} \sum_{k=1}^{r \times c} c_k u_k|| &= \inf || \sum_{k=1}^{r \times c} \gamma_k c_k u_k|| \\
&\geq (\min_k \gamma_k) \inf || \sum_{k=1}^{rc} c_k u_k|| = (\min_k \gamma_k) \inf ||\nabla \tilde{J}|| \\
&> (\min_k \gamma_k) \sigma_1 > 0, \quad \forall t > t_0.
\end{aligned}$$

Call $\sigma_2 = (\min_k \gamma_k)\sigma_1$. To proceed with the rest of the proof, note that, since $-H_t^{-1} \nabla \tilde{J}_t$ is a direction of descent, there exists an $\underline{\alpha} > 0$ such that (15) is satisfied at every iteration for t large enough (Panier et al., 1988, Lemma 3.9). With $\alpha_t \geq \underline{\alpha}$ and given the positive definiteness of H_t for t large enough, (24) demands that $||\nabla \tilde{J}_t|| \to 0$, which is a contradiction to $\inf ||\nabla \tilde{J}_t|| > \sigma_1 > 0$ for t large enough. Therefore, $||\nabla \tilde{J}_t||$ eventually converges to zero and, from Lemma 1, there is an accumulation point $\{A^*, v^*\}$ for the series $\{A_t, v_t\}$ which is the local solution. Q.E.D

B Fuzzy c-mean Algorithm (Bezdek, 1981)

Choose a random starting point for each $v^{(i)}$.

1. Form membership function $\mu^(i)(x)$ using $v^{(i)}$.

2. Let
$$\left(v^{(i)}\right)^+ = \frac{1}{\sum_{i=1}^{c} \left(\mu^{(i)}\left(x(j)\right)\right)^m} \sum_{i=1}^{c} \left(\mu^{(i)}\left(x(j)\right)\right)^m x(j), \quad i = 1, \ldots, c.$$

3. Set $v^{(i)} = \left(v^{(i)}\right)^+$.

Repeat steps 1 to 3 until some stopping criteria is met.

C Gradient and Hessian of $\tilde{J}$

The gradient of $\tilde{J}$ is computed analytically. First the derivative of membership functions to the change in centers need to be computed, which are:

$$\frac{\partial \mu^{(i)}(x)}{\partial v_j^{(i)}} = m \frac{x_j - v_j^{(i)}}{||x - v^{(i)}||} \left(\mu^{(i)}(x) - \left(\mu^{(i)}(x) \right)^2 \right),$$

$$\frac{\partial \mu^{(i')}(x)}{\partial v_j^{(i)}} = -m \frac{x_j - v_j^{(i)}}{||x - v^{(i)}||} \mu^{(i)}(x) \mu^{(i')}(x).$$

Let

$$dX_{ij} = \begin{bmatrix} \frac{\partial \mu^{(1)}\left(x(1)\right)}{\partial v_j^{(i)}} x(1) & \frac{\partial \mu^{(2)}\left(x(1)\right)}{\partial v_j^{(i)}} x(1) & \dots & \frac{\partial \mu^{(c)}\left(x(1)\right)}{\partial v_j^{(i)}} x(1) & \mathbf{0} \\ \frac{\partial \mu^{(1)}\left(x(2)\right)}{\partial v_j^{(i)}} x(2) & \frac{\partial \mu^{(2)}\left(x(2)\right)}{\partial v_j^{(i)}} x(2) & \dots & \frac{\partial \mu^{(c)}\left(x(2)\right)}{\partial v_j^{(i)}} x(2) & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \frac{\partial \mu^{(1)}\left(x(n)\right)}{\partial v_j^{(i)}} x(n) & \frac{\partial \mu^{(2)}\left(x(n)\right)}{\partial v_j^{(i)}} x(n) & \dots & \frac{\partial \mu^{(c)}\left(x(n)\right)}{\partial v_j^{(i)}} x(n) & \mathbf{0} \end{bmatrix}.$$

Then

$$[\nabla J]_j^{(i)} = -2(Y - XA)' W dX_{ij} A,$$

and $\nabla \tilde{J}$ follows as in (12).

Closed form expressions for the Hessian matrix prove much more complicated. Instead, an estimate of the inverse Hessian matrix is constructed by BFGS updates. Call the inverse Hessian C and let

$$s = -\alpha C \nabla \tilde{J}, \text{ and } g = \nabla \tilde{J}^{+} - \nabla \tilde{J}.$$

Note that s is the change made to the location of centers in the current iteration, and g is the change in the gradient from the current iteration to the next. The new estimate of inverse Hessian, C^{+}, is constructed according to (Nocedal & Wright , 1999):

$$C^{+} = \left(I - \frac{sg'}{g's} \right) C \left(I - \frac{gs'}{g's} \right) + \frac{ss'}{g's} + 10^{-5} \mathbb{I}_{r \times c}.$$

To keep C positive definite throughout the iterations, it is necessary to make sure that the curvature condition ($g's > 0$) is always satisfied. In case, it is not satisfied in a particular iteration, C is carried over to the next iteration with no change. The addition of a small multiple of the identity matrix is a precautionary measure and takes effect when the estimate of the inverse Hessian approaches singularity, and it keeps the smallest eigenvalue larger than or equal to 10^{-5}.

References

Acs, Zoltan J. and David B. Audretsch (1988) "Innovation in Large and Small Firms: An Empirical Analysis," *American Economic Review*, 78(4), 678–690.

Alchian, Armen A., and Harold Demsetz (1972) "Production, Information Cost, and Economic Organization," *American Economic Review*, 62(5), 777-795.

Armijo, Larry (1966) "Minimization of functions having Lipschitz continuous first partial derivatives," *Pacific J. of Mathematics*, 16(1), 1–3.

Australian Bureau of Statistics (2009) "Counts of Australian Businesses," *Catalogue No. 8165.0*

Australian Taxation Office (2011) *Small Business and General Business Tax Break.*

Bakhtiari, Sasan and Andre L. Tits (2003) "A Simple Primal-Dual Feasible Interior-Point Method for Nonlinear Programming with Monotone Descent", *Journal of Computational Optimization and Application*, 25(1/2/3/), 17–38.

Bakhtiari, Sasan (2011) "Size Evolution and Outsourcing: Theory and Evidence from Australian Manufacturing," *UNSW Working Paper.*

Barnes, Matthew and Jonathan Haskel (2002) "Job Creation, Job destruction and the Contribution of Small Businesses: Evidence for UK Manufacturing," *University of London Working Paper*, No.461.

Bezdek, Jim C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*, Springer, New York.

Broersma, Lourens and Pieter Gautier (1997) “ Job Creation and Job Destruction by Small Firms: An Empirical Investigation for the Dutch Manufacturing Sector,” *Small Business Economics*, 9(3), 211–224.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh (1996) *Job Creation and Destruction*, MIT Press, Cambridge, Massachussetts.

Davies, David L. and Donald W. Bouldin (1979) “A Cluster Seperation Measure,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227.

Filev, Dimitar P. (1991) “Fuzzy Modeling of Complex Systems,” *International Journal of Approximate Reasoning*, 5, 281–290.

Friedman, Jerome H. (1991) “Multivariate Adaptive Regression Splines,” *Annals of Statistics*, 19(1), 1–67.

Haltiwanger, John C., Ron S. Jarmin, and Javier Miranda (2010) “Who Creates Jobs? Small vs. Large vs. Young,” *NBER Working Paper*, No. 16300.

Her Majesty Revenue and Customs (2005) “Working Towards a New Relationship: A Consultation on Priorities for Reducing the Administrative Burden of the Tax System on Small Business,” UK.

Internal Revenue Services (2011), *Form 8941 Instructions*, USA.

McLachlan, Geoffrey J., and David Peel (2000) *Finite Mixture Models*, Wiley, New York, USA.

Neumark, David, Brandon Wall, and Junfu Zhang (2011) “Do Small Businesses Create More Jobs? New Evidence for the United States from the National Establishment Time Series,” *Review of Economics and Statistics*, 93(1), 16–29.

Nocedal, Jorge and Stephen J. Wright (1999) *Numerical Optimization*, Springer Verlag, New York, USA.

Panier, Eliane R., Andre L. Tits, and Jose N. Herskovits (1988) “A QP-free, Globally Convergent, Locally Super-linearly Convergent Algorithm for Inequality Constrained Optimization,” *SIAM Journal of Control and Optimization*, 26(4), 788–811.

Storey, David J. (1989) "Firm Performance and Size: Explanations from the Small Firm Sectors," *Small Business Economics*, 1, 175–180.

Storey, David J. (1994) *Understanding the Small Business Sector,* Thompson learning, London, UK.

Will, Lou, and Hugh Wilson (2001) "Tricks and Traps of the Business Longitudinal Survey," *Productivity Commission Staff Paper.*

Williamson, Oliver (1967) "Hierarchical Control and Optimum Firm Size," *Journal of Political Economy*, 75(2), 123–138.

Wright, M.H. (1998) "Ill-conditioning and Computational Error in Interior Methods for Nonlinear Programming," *SIAM J. of Optimization*, 9(1), 84–111.

Yager, Ronald R., and Dimitar P. Filev (1993) "Unified Structure and Parameter Identification of Fuzzy Models," *IEEE Transactions on System, Man, and Cybernetics*, 23(4), 1198–1205.

Zadeh, Lotfi (1965) "Fuzzy Sets," *Information and Control*, 8, 338–353.

Zimmermann, H.J. (1996) *Fuzzy Set Theory and Its Applications*, 3rd Edition, Kluwer Academic Publications, Boston/Dordrecht/London.