Does Gibrat's law hold amongst dairy farmers in Northern Ireland?

Kostov, Philip, Patton, Myles, Moss, Joan and McErlean, Seamus
Queens University Belfast

02 October 2006

Does Gibrat’s law hold amongst dairy farmers in Northern Ireland?

Philip Kostov*, Myles Patton**, Joan Moss* and Seamus McErlean***

* Department of Agricultural and Food Economics, Queens University Belfast

** Agri-Food and Biosciences Institute

*** Department of Agricultural and Rural Development in Northern Ireland

Does Gibrat's law hold amongst dairy farmers in Northern Ireland?

Abstract

This paper tests whether the Law of Proportionate Effects (Gibrat, 1931), which states that farms grow at a rate that is independent of their size, holds for the dairy farms in Northern Ireland. Previous studies have tended to concentrate on testing whether the law holds for all farms. The methodology used in this study permits investigation of whether the law holds for some farms or all farms according to their size. The approach used avoids the subjective splitting of samples, which tends to bias results. Additionally we control for the possible sample selection bias. The findings show that the Gibrat law does hold except in the case of small farms. This is in accordance with previous findings that Gibrat's law tends to hold when only larger farms are considered, but tends to fail when smaller farms are included in the analysis. Implications and further extensions, as well as some alternatives to the proposed methodology are discussed.

Keywords:

Gibrat's law; quantile regression, Integrated conditional moments test

JEL codes: C12, C14, O49.

1. Introduction

Farm structural change and the inter-related issue of farm growth continue to attract the interest of agricultural economists, academics and policy makers, because of the wide-ranging implications for agricultural output, efficiency and inevitably the economic welfare of rural communities. Given the sweeping changes that have occurred in the recent round of CAP Reform it is important that the factors affecting structural change in the farm sector are understood. This study will examine some of these factors using data from the period 1997 to 2003. During this period, farm incomes reached very low levels (see Figure 1) and this has certainly put pressure on the structure of the farm sector.

Two inter-related components of farm structural change are entry/exit to the farm sector and growth/decline in continuing farms (Weiss, 1999). Historically, there has been a tendency for academics to consider these two components separately. Studies that have investigated the factors influencing farm exit include Kimhi and Bollman (1999) and Glauben *et al*. (2003). A large number of studies have examined the factors affecting farm growth (e.g. Upton and Haworth, 1987; Shapiro *et al.*, 1987; Clarke *et al.*, 1992; and, Bremmer *et al.*, 2002). However, it has been argued that examining the growth of continuing farms only, whilst ignoring exiting farms in the analysis, runs the risk of biasing results (Weiss, 1999). Consequently, several more recent studies have considered both farm survival (the opposite of exiting) and farm growth (see Weiss, 1998 and 1999; and, Key and Roberts, 2003).

In a seminal paper that describes the Law of Proportionate Effects, Gibrat (1931) provides the starting point for most previous studies of farm growth. He proposed that the growth rate of firms is independent of their initial size at the beginning of the period examined. Gibrat's law proposes that growth is a stochastic process resulting from the random operation of many independent factors. This stochastic process easily generates theoretical farm size distributions (log-normal) that are skewed and similar in shape to the farm size distributions that are observed empirically. Furthermore, this stochastic process also means the variance of the distribution increases over-time mirroring observed increases in empirical concentration measures. It is unsurprising, therefore, that Gibrat's Law, which is not inconsistent with the assumption of constant returns to scale, continues to provide the basic foundation for most studies of farm growth.

Gibrat's law, however, has been subject to some criticism and theoretical justifications for its rejection can be found in the literature, e.g. returns to scale among smaller farms (Weiss, 1999). A number of studies have identified a range of systematic factors that influence farm growth and these should be explicitly considered in any model of farm growth, rather than being subsumed within the random stochastic process implied by Gibrat's law. In light of these recent considerations this paper test Gibrat's law using some determinants of farm growth and survival using farm census data for dairy farms in Northern Ireland. The analysis permits examination of Gibrat's law and considers the influence of such factors as profitability, farm type and farmer associated characteristics on farm growth and survival. An outline of the paper is as follows. A review of previous studies is presented in the next section. The data set and the methodology used in the analysis are outlined in Section 3. Model results are presented in Section 4. These results are discussed in Section 5 and some conclusions are drawn in Section 6.

2. Literature review

There is a wide and extensive literature investigating the growth of firms (for a review see Sutton, 1997; and, Lotti et al., 2003). In comparison the number of studies focusing on farm growth is more limited. The approach used in most studies of farm growth has been to test Gibrat's law. Many of these studies appear to have reached different conclusions. Studies by Weiss (1998, 1999) and Shapiro *et al*. (1987), based on farm census data, rejected

Gibrat's law of proportionate effects for farm growth. These studies found that small farms tend to grow faster than larger farms. However, Upton and Haworth (1987) and Bremmer *et al*. (2002) using FBS data collected in Great Britain and FADN data collected in the Netherlands, respectively, found no evidence to reject Gibrat's law. The data sets used in these two studies exclude very small farms (i.e. farms <8 European Size Units). This may have affected the results obtained because small rapidly growing farms may have been excluded from the analysis. Clark *et al.* (1992) also find no evidence to reject Gibrat's law based on analysis using aggregated data.

An important aspect of the study of farm (size) growth is the definition of farm size. Previous studies have used a variety of different measures of farm size (changes in farm size indicate growth/decline). Measures of farm size proposed in the literature include: acreage farmed, livestock numbers (cow equivalents), total capital value, net worth, gross sales, total gross margins and net income (Allanson, 1992; Clark *et al*. 1992; and, Shapiro *et al*. 1987). Output value measures such as gross farm sales and input value measures such as net worth may be unsatisfactory due to the impact of inflation and changes in relative prices (Weiss, 1998). Physical input measures, such as acres under cultivation and number of livestock, are also problematic since farms are characterised by a non-linear production technology and changes in farm size typically involve changes in factor proportions and changes in technology. Weiss (1998), however, argues that the disadavantages of physical input measures are less that those associated with the value of inputs or outputs and as a result the former should be preferred.

A review of empirical studies indicates that most find evidence to conclude that a range of variables other than size influence farm growth (e.g. Weiss, 1999; and, Bremmer *et al*., 2002). These other explanatory variables that have been identified in the literature can be divided into two sub-groups, namely, farmer-associated characteristics and farm specific factors. Weiss (1999) identified the following farmer-associated characteristics: off-farm employment, age, gender, level of education, family size, age profile of the family, succession information and attitude to risk. Farm (or firm) specific variables that have been suggested as factors influencing growth include: size, solvency, profitability, productivity, farm income, structure, financial performance, input costs, output mix, farm type, mechanisation and location (Bremmer *et al*. 2002; Hardwick and Adams, 2002; Weiss, 1999).

3. Problem characterisation

As with the vast majority of studies of farm growth, Gibrat's law of proportionate effect is used as a starting point for the analysis carried out in this paper. The law states that firm (or farm) growth is determined by random factors, independent of size. This may be tested using the following formula:

$$ln(S_{i,t}) = \beta_1 + \beta_2 ln(S_{i,t-1}) + u_t; \tag{1}$$

where $S_{i,t}$ denotes the size of the individual holding in time t and u_t is a random disturbance term independent of current or past values of the dependent variable. If $\beta_2 = 1$, then growth rate and initial size are independent and this means that Gibrat's Law is not rejected. If $\beta_2 < 1$, small farms tend to grow faster than larger farms – i.e. the effects of randomness are offset by negative correlation between growth and size. If $\beta_2 > 1$, larger farms tend to grow faster than smaller farms. The above is easily illustrated if one subtracts $ln(S_{i,t-1})$ from both sides of equation (1) above. Then the left hand side logarithmic difference is an approximation of the growth. The right hand side will them be either a random walk, when $\beta_2 = 1$, or a dependent process otherwise.

Equation (1) can be generalised augmenting it by farmer associated characteristics and farm specific variables, e.g. profitability, denoted by the *k* explanatory variables X_k below as follows:

$$\ln(S_{i,t}) = \beta_1 + \beta_2 \ln(S_{i,t-1}) + \sum_{j=3}^{k+3} \beta_j X_{j-2} + u_t \qquad (2)$$

There are two problems with using linear regression representations such as equation (2). The first is the assumed linear effect of the additional explanatory variables X_k. Weiss (1999) for example applied non-linear functional form for these and detected significant non-linearities. Specifying an ad-hoc non-linear functional form however is not a viable strategy, since it may impact on the final results in an unpredictable way and often there is little or no information on the way these additional variables may impact on farm growth.

This consideration aside, even in the simple model (1), there is an underlying assumption that the Gibrat's law holds (or is violated) globally. It is difficult to ascertain whether for example small farms obey this law as opposed to large farms. It is in principle possible to split the sample into smaller subsamples and locally estimate the relationships. This would however involve some subjective criteria about how to do the latter partitioning further casting doubt on the final results. If we want to test whether Gibrat's law holds for some farms and not for others, the linear regression framework is too restrictive. Such a test can nevertheless be designed using quantile regression methods, implemented in this paper. Alternatives and extensions to the adopted approach are also discussed.

In order to measure farm growth, farm size must be compared between two specific points in time. However, measures of farm growth are meaningful only for surviving farms. Farms exiting between the points in time over which growth is measured are normally excluded from the sample (as non-surviving farms). However, there is a greater probability that slower growing small farms will be non-survivors compared with slower growing larger farms (Weiss. 1999). Thus, if non-surviving (exiting) farms are excluded from the sample, the estimates of β may be biased downward, which may result in incorrect rejection of Gibrat's law, giving the impression that smaller farms tend to grow faster than larger farms (Hardwick and Adams, 2002; Lotti *et al*. 2003; Shapiro *et al*. 1987; Sutton, 1997; and, Weiss, 1999). Ignoring exiting farms in the analysis is known as the problem of sample selection bias. Various options are available to account for selection bias and these are briefly discussed in the methodology section.

4. Data

The data set used in this study is based on the 1997 and 2003 farm census for Northern Ireland and a structural survey of farms in Northern Ireland that was conducted in 1997. The farm census provided information for individual farms on farm type, acreage farmed and stock numbers (total standard gross margin for each farm can also be inferred from this data). The 1997 structural survey provided additional information on a range of farmer associated characteristics such as gender, age, management status and time spent working on the farm for a subset of the farms included in the farm census (31 % of dairy farms). The individual farm information from the structural survey was matched to the information from the 1997 farm census.

Matching these data sets yielded a total of 1648 dairy farms in 1997. Of these farms, 112 had exited farming by 2003. Of the remaining 1536 farms, 1290 remained in dairy, while the other 246 moved to cattle and sheep. In this study we are specifically interested in farms which remain in dairying and thus the latter farms are treated as farms which exited the dairy sector. Thus, in total 358 exited the dairy sector between 1997 and 2003 (farms which exited farming altogether plus farms which switched from dairy to cattle and sheep).

The measure of farm size used in this study is the livestock numbers measured in cow equivalents per farm. This measure (unlike e.g. land area) is directly proportionate to the final output of dairy farm. Using the dairy farm sector allows us to avoid complications associated with farm entry and thus simplify the sample selection problem.

The following explanatory variables are employed: (logarithm of the) initial (i.e. in 1997) size - LNCE97; an indicator variable denoting whether the farm holder is also a manager of the farm – MSHOLD, indicator of other gainful activities – HAGA3 and a variable showing the age of the farmer – HAGE1. Note that we use a limited set of conditioning variables, since our purpose is to test Gibrat's Law, rather than provide a comprehensive model of farm growth, which would have involved additional behavioural assumptions and theoretical model.

5. Methodology

In the least-squares regression framework the conditional mean function, i.e. the function that describes how the mean of *y* changes with the covariates *x*, is almost all we need to know about the relationship between *y* and *x*. The crucial aspect about this view is that the error is assumed to have exactly the same distribution irrespectively of the values taken by the components of the vector *x*. This can be viewed as a pure 'location shift' model since it assumes that *x* affects only the location of the conditional distribution of *y*, not its scale, or any other aspect of its distributional shape. If this is the case, we can be fully satisfied with an estimated model of the conditional mean function.

The above described location shift model is however rather restrictive. Covariates may influence the conditional distribution of the response in many other ways: expanding its dispersion (as in traditional models of heteroscedasticity), stretching one tail of the distribution, compressing the other tail (as in volatility models), and even inducing multimodality. Explicit investigation of these effects can provide a more nuanced view of the stochastic relationship between variables, and therefore a more informative empirical analysis. The quantile regression is a method that allows us to do so.

Given a random variable *Y* and its distribution function *F*, we denote by

$$Q(\tau) = \inf(y \mid F(y) \geq \tau) \tag{3}$$

the τ th quantile of Y. The sample analogue q of $Q(\tau)$ is called the τ th sample quantile. It may be formulated as the solution of the following optimisation problem, given a random sample (y_n), $n=1;\ldots;N$:

$$\min\left\{ \sum_{\{n|y_n \geq q\}} \tau |y_n - q| + \sum_{\{n|y_n < q\}} (1-\tau) |y_n - q| \right\} \tag{4}$$

There exist a number of alternative quantile regression estimators. Here we will only describe the linear programming type of estimator, since there are asymptotic theory results for it (Koenker and Bassett, 1978). Just as we can define the sample mean as the solution to the problem of minimizing a sum of squared residuals, we can define the median (which is the 50% quantile, i.e. τ =0.5) as the solution to the problem of minimizing a sum of absolute residuals (which follows directly from (4) above).

For any $0 < \tau < 1$, we denote $\rho_\tau(u) = u\left(\tau + I_{[u<0]}\right)$, where $I[.]$ is the indicator function. Following Koenker and Bassett (1978) $\rho_\tau(u)$ is usually referred to as a *check* function. The problem may then be formulated as follows:

$$\min \sum_{n=1}^{N} \rho_\tau(y_n - q) \tag{5}$$

which yields a natural generalization to the regression context.

$$\min \sum_{n=1}^{N} \rho_\tau(y_n - \xi(X, \beta)) \tag{6}$$

where $\xi(X, \beta)$ is some parametric function of the covariates. When this is a linear function, the above minimisation procedure is actually a linear programming problem. Then it may be estimated using some form of simplex algorithm. Koenker and d'Orey's (1987, 1993) adaptation of the Barrodale and Roberts (1974) median regression algorithm to general quantile regression is particularly influential. The Barrodale and Roberts approach belongs to the class of exterior point algorithms for solving linear programming problems. Alternatively, Portnoy and Koenker (1997) have shown that a combination of interior point methods and effective problem preprocessing is very well suited for large-scale quantile regression problems. This is the approach used in this paper (it is often referred to as Frisch- Newton method), although the former (Barrodale-Roberts) method yields similar results, which are available from the authors upon request.

It would be beneficial at this point to clarify a fundamental difference between the quantile regression and the mean regression methods. Could we achieve the same result by simply segmenting the response variable into subsets according to its unconditional distribution and then doing least squares fitting on these subsets? Clearly, this form of truncation on the dependent variable would yield disastrous results in the present example. In general, such strategies are doomed to failure for all the reasons so carefully laid out in Heckman (1979). It is thus worth emphasizing that even for the extreme quantiles all the sample observations are actively used in the process of quantile regression fitting.

It is of course possible to construct local quantile regression estimates using some sort of segmentation (see Knight et al., 2002). Some preliminary results about the conditions where local quantile regression is useful are outlined in Costinot et al. (2000). However we will not pursue this option for reasons given further below.

There are several useful properties of the quantile regression approach. Above we have described the quantile regression for a given quantile. If however one takes the whole range of quantiles, a picture of the overall distribution emerges. Note that in the latter case we obtain a variable coefficients model. In contrast to most variable coefficient methods which usually assume coefficients independence however, in the quantile regression setting, the coefficients are functionally dependent. In the light of the farm growth problem, this is evidently a desirable property, The determinant of the farm growth for slightly different sizes of farms are related in a quantile regression context, while paradoxically they will be assumed independent in most other variable coefficients models. Even the simplest linear quantile regression we adopt here produces a rather flexible non-linear model. Note furthermore that the non-linearities are explicitly formulated with regard to the dependent variables, i.e. with regard to the farm size, which is exactly what ids necessary for testing Gibrat's law. Nevertheless, the quantile regression has to be viewed as a workable approximation to a possibly more general non-linear model. The availability of pointwise convergence results for the quantile regression estimates facilitates the analysis and inference compared to other non and semi-parametric methods.

The potential problem of bias due to sample attrition is known in the literature as a sample selection problem. Its initial description is due to Heckman (1979) who devised a two step procedure for controlling it. The Heckman procedure consists of estimating at step one a survival model. This is typically a probit (although a logit can be used alternatively) equation on the probability of farm survival from the complete sample (including surviving and non-surviving farms). This equation is subsequently used to obtain an additional variable, where the values represent the inverse Mill's Ratio for each observation. In step-two, the additional variable is introduced as a correcting factor into the least squares regression that is based upon a sample that excludes non-surviving farms. The probit model, used in the first step typically has the same explanatory variables as the main equation, though this is not mandatory and variables that are only relevant to the farm survival may be included, as well as some of the variables included in the main equation may be dropped. The Heckman procedure assumes joint normality of the error terms in the two equations. The latter distributional assumption, which can also be employed to construct a more efficient Full Information Maximum Likelihood (FIML) estimator that jointly estimates both equations, however can have serious implications on the robustness of the final results when it is violated. Therefore various alternative estimators have been suggested to circumvent the problem of inadequate distributional assumptions. These can be broadly described as semi-parametric model selection methods.

The problem of sample selection in mean regression model can be broadly defined as problem of the distributional assumptions, which can be controlled for. This is basically done by various methods to relax the parametric specifications employed in the seminal work of Heckman (1979).

The sample selection for quantile regression however remains a challenging and still under-researched problem. Buchinsky (1998, 2001) provided some important contributions to this issue. Unfortunately the method Buchinsky (1998,2001) used in the selection step, namely the Ishimura’s (1993) semiparametric least squares requires that the selection equation includes at least one covariate that is not included in the main equation. This condition is difficult to ensure with the available data set. Therefore a different strategy is followed in this paper. On the first step we estimate an ordinary probit selection equation similarly to Heckman (1979) and from there derive the bias correcting factor (i.e. the inverse Mill’s ratio). In the second step a linear quantile regression is performed instead of the mean regression by additionally including the derived correcting factor. The resulting model is tested for model correctness, which also validates the sample selection step.

The last piece of the jigsaw is therefore to identify appropriate model validity test, applicable to the quantile regression. What is needed is a test on validity of the functional form. Up to our knowledge there are only two appropriate candidates for this. The first is the Zheng’ (1998) approach based on weighted kernel regression estimation and the other one is an extension to the Bierens and Ploberger’s (1997) Integrated Conditional Moment (ICM) test for the quantile regression case due to Bierens and Ginther (2001). They discuss explicitly only the median case, but the necessary modifications for the general quantile regression case are provided in an appendix. We have chosen the latter due to some desirable properties, such as boundness of the test statistic, good local power and relative conservatism of the test statistic. We briefly describe the ICM test below:

Let us have the following expectation model:

$$E\left(y_j \mid x_j\right) = g\left(x_j, \beta\right) \tag{7}$$

then the ICM statistic

$$\hat{T}_{ICM} = \frac{\int |\hat{z}(\xi)|^2 \, d\mu(\xi)}{\int \hat{\Gamma}(\xi,\xi) \, d\mu(\xi)} \tag{8}$$

with $\hat{z}(\xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \hat{u}_j \omega\left(\xi' \Phi\left(x_j\right)\right)$, $\hat{u}_j = y_j - g\left(x_j, \hat{\beta}\right)$ being weighted non-linear residuals, $\Phi(.)$ - bounded one to one mapping and $\omega(.)$ - appropriate weight function, can be used to test the null hypothesis that the probability of the expectation given above in (7) is one, against the alternative of being the less than one. In simple words, this amounts to testing that the expectational model (7) is the right one. Practical implementation of the ICM test involves choice for the weight function and the bounded mapping as well as some computational issue surrounding the computation of the two integrals in (8). For brevity of the exposition we will not discuss these here. Full details on the procedure, which roughly speaking follows Bierens and Ginther (2001)[1], are available from the authors.

There is another point to note. Since we are computing a two-step estimator, the confidence intervals obtained during the quantile regression estimated at the second step will not be valid even asymptotically. The reasons for this are similar to the instrumental variables case and the endogeneity problems. In the sample –selection case analysed here, in general the quantile regression estimated at the second step is actually:

$$Q(\tau) = g\left(X, \beta_\tau\right) + k_\tau \left[f(.)/\left(1 - F(.)\right)\right] + \varepsilon \tag{9}$$

The additional regressor $f(.)/\left(1-F(.)\right)$ can be interpreted as a 'non-selection hazard', that corrects for the effect of the sample selection bias. The *pdf* $f(.)$ and the *cdf* $F(.)$ are obtained from the auxiliary model estimated at step one. Note that (9) is the expression when the auxiliary model models non-selection. If the auxiliary model is the one of selection the

[1] They do not provide sufficient information on the exact quantile regression estimation algorithm they use or the method of obtaining standard errors, thus full replication was not possible.

natural candidate for the non-selection hazard would be $f(.)/F(.)$. In the case of symmetric distributional assumptions used in the auxiliary model (such as the Gaussian used in the probit model), modelling selection or non-selection would be equivalent. This would not however be the case for semiparametric estimators, and thus it is preferable to use the non-lecetion model in the first step for such cases. The “inverse Mill’s ratio” is the representation of the above non-selection hazard obtained from the normal distribution (or the logistic, if a logit model was estimated instead of probit). If one estimates a non-parametric model at step one the empirical *pdf* and *cdf*, obtained from it could be used.

The Heckman type of procedure ignores the correlation between the residuals in the first and second step models. Since the first-step residuals are used to compute the non-selection hazard, the latter would therefore be correlated with the residuals in the second-step equation. With a probit equation estimated in step one, the standard errors in the following linear regression can be adjusted (see Heckman, 1979) obtain asymptotically valid standard errors. An alternative (applicable with any pair of parametric models for the two steps) is to estimate jointly both equations by Full Information Maximum Likelihood, which by accounting for the correlation between the error terms should yield consistent standard errors. Note however that this is not applicable in this case because the quantile regression is a semi-parametric method, which does not make any distributional assumptions. To obtain valid standard errors , which are necessary for testing the Gibrat’s law, we adopt a different approach. We consider both equations as a whole and bootstrap the whole two-stage estimator (instead of only the QR estimates). An adapted quantile regression implementation of the XY (also known as pair, or case) bootstrap is used for this purpose.

6. Results.

The first-step estimates are presented in Table 1. Note that since we use probit model, and thus the choice of auxiliary criterion is not instrumental, we model selection (i.e. we use 1 for farms that remain in dairying)[2]. This leads to a more natural interpretation of the estimated coefficients, if this is necessary. We use the same variables as in the subsequent quantile regression model. The farm size is important determinant of the farm exits, in that larger farms are more likely to remain in dairying. Owing to this if the selection process was not controlled for, one would have obtained biased estimates in the main equation

Insert Table 1.

The discussion in the methodology section concentrated on estimation and testing of a quantile regression model at a given quantile. It is clear that the adopted framework for testing the Gibrat's law involves estimating multiple quantile regression models. It is in principle possible to estimate the whole quantile process (i.e. estimating a quantile regression for every observation, in this case 1290 models). To simplify the process however only a subset of quantile regression models is estimated. This subset consist of all percentiles excluding the lower and the upper 9%. In other words the 81 regression models for the 0.10, 0.07, ,0.89, and the 0.90 quantiles were estimated. The reason we exclude the extreme quantiles is that the conventional quantile regression estimates for these are unreliable. Asymptotic theory and estimation methods for extreme quantiles are developed in Chernozhukov (2000a,b) and Chernozhukov and Umantsev (2001). The main interest of the current paper lies in the overall distribution of the estimated on the logged size in the initial period (1997) and thus we will ignore the extreme quantiles.

[2] Correspondingly the additional variable used in the subsequent quantile regressions is $f(x)/F(x)$.

Plotting the corresponding estimates for the same parameter across the quantile range provides a useful graphical device to informally ascertain the scale invariance hypothesis. Strict formal tests on this are available and results from such tests can be provided by the authors upon request. The conventional approach to such tests (Koenker and Xiao, 2000) uses the Khmaladze (1981) transformation and introduces additional computational burden. Although we omit it here for brevity and simplicity, rigorous modelling practices would require one to implement such tests. Graphical representation of the overall quantile regression is presented on Figure 2.

Insert Figure 2.

The main point of interest here however, is the way the estimate of the logged lagged size varies over the quantile range. For this reason we will not comment on the overall results and will focus our attention to this particular coefficient. The estimates for the coefficient of the initial farm size from the estimated set of quantile regressions are presented on Figure 3, together with the associated 95% confidence intervals. Where the horizontal line drawn at the value of 1 falls within the range defined by these confidence intervals we may say with 95% confidence level, that the Gibrat law holds. These results suggest the following. The Gibrat's law speaking holds except for the small (up to the 0.16 quantile) dairy farms. These smaller farms grow slower than the rest of the sector. Interestingly the coefficient estimate declines for the largest farms, but this decline is not statistically significant. If we also include the extreme quantiles, the largest farms do show slower growth than the rest (results available from the authors). Nevertheless such a result is difficult to verify, since the ICM test statistic is unreliable at the extreme quantiles.

Insert Figure 3.

The smaller growth in the segment of smaller dairy farms is in concordance with farm growth results from Census and farm business surveys data, since, the latter are generally based on larger farms, and thus will tend to support the Gibrat's law as opposed to the latter where the peculiarity of the smaller farms would lead to its rejection. The use of dairy sector data is advantageous in that it reduces the possibility for heterogeneity problems due to different production technologies and farm types being pooled together. McErlean et al. (2004) argue that this heterogeneity needs to be dealt with by explicitly modelling the different farm types instead of using dummies for them. Note that the possible effects of such heterogeneity will manifest themselves in terms of heteroscedasticity problems. The distribution of the ICM test we use to test the model validity however is not affected by neglected heteroscedasticity (Bierens and Ploberger, 1997).

Due to the considerable computational burden of estimating the ICM test statistic (details on the exact procedure available upon request) we only estimate it for the quantiles from 0.1 through to 0.9 with 0.1 steps). The ICM test results are presented in Table 2

Insert Table 2

.

The reason for using several values for c is as follows. The ICM statistic is a ratio of two probability measures estimated over a hypercube whose dimensions are $2c$. ((i.e. in the intervals $[-c, c]$). In principle asymptotically any choice for c is equivalent. In principle however, this choice may have dramatic effects on the small sample properties of the test. In general too small or too large values will reduce the power of the test. (see Bierens and Ginther, 2001 for a more detailed discussion on this in the quantile regression case).

Therefore a range of such values was used to estimate the ICM test. All test statistics estimated fail to reject the null of validity of the estimated quantile regression. The results for c=0.1 similarly to Bierens and Ginther (2001) are probably spuriously low. Nevertheless the range of values for the hypercube dimension is rather extensive (as a comparison Bierens and Ginther (2001) only use values of 1, 5 and 10) and everywhere the ICM test statistic is well below the critical values. This provides conclusive evidence in support of the estimated quantile regression model and its conclusions.

7. Conclusions and further research agenda

Previous studies show that Gibrat's law tends to hold when only larger farms are considered, but tends to fail when smaller farms are included in the analysis. This study is based on a data set that covers the full range of dairy farm sizes in Northern Ireland. The analysis takes account of possible bias due to exiting farms. Our results indicate that the farm growth does not depend on initial size, except for the smaller farms. Small Northern Ireland dairy farms relying on family labour, probably experience resources shortage and have insufficient funds to expand under milk quota restrictions. On the other hand the largest farms seem to grow slower than the rest, hinting a possible a saturation effect, but the latter is inconclusive. The relatively small average dairy farm size in Northern Ireland may explain such an effect.

The use of rather homogeneous data set consisting of only dairy farms have prevented some complications such as possible heterogeneity, but the general approach outlined in the paper is readily applicable to more complex data sets. In such cases the simple probit sample selection step may not be appropriate and alternative semi-parametric formulations may be used instead. The linear quantile regression proved to be sufficient to describe the growth process in the NI dairy sector. In some other cases however the linear assumption may be inadequate. Then nonlinear and non-parametric versions of the quantile regression could be employed instead

References

Allanson, P. 1992. Farm size structure in England and Wales 1939-89. Journal of Agricultural Economics 43(2), 137-148.

Barrodale, I., and Roberts, F. 1974. Solution of an overdetermined system of equations in the L1 norm. Communications of the ACM 17, 319-320.

Bierens H. J. and Ploberger, W. 1997. Asymptotic theory of integrated conditional moment tests. Econometrica 65, 1129-1151.

Bierens, H. J. and Ginther, D. K. 2001. Integrated conditional moment testing of quantile regression models. Empirical Economics 26: 307-324.

Bremmer, J., Oude Lansink, A.G.J.M., Olson, K.D., Baltussen, W.H.M., and Huire, R.B.M. 2002. Analysis of farm development in Dutch agriculture and horticulture. Paper presented at 13th Congress of the International Farm Management Association (IFMA) in National Sports and Conference Centre Papendal, Wageningen, The Netherlands, July 7-12, 2002.

Buchinsky, M. 1998. The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. Journal of Applied Econometrics 13, 1—30.

Buchinsky, M. 2001. Quantile regression with sample selection: estimating women's return to education in the U.S.. Empirical Economics, 26(1), 87-113.

Chernozhukov, V. 2000a. Conditional extremes and near-extremes, MIT Dept. of Economics Working Paper No. 01-21.

Chernozhukov, V. 2000b. Conditional extremes and near-extremes: concepts, asymptotic theory and economic applications, unpublished PhD Dissertation, Stanford University.

Chernozhukov, V. and Umantsev, L. 2001. Conditional value-at-risk: aspects of modeling and estimation. Empirical Economics 26(1), 271 - 292

Clarke, J.S., Fulton, M. and Brown, D.J. 1992. Gibrat's law and farm growth in Canada. Canadian Journal of Agricultural Economics 40, 55-70.

Davies, S. and Lyons, B. (1982). Seller concentration: the technological explanation and demand uncertainty. Economic Journal 92, 903-919.

Gibrat, R. 1931. Les Inegalites Economiques. Librairie du Recueil Sirey. Paris.

Glauben, T., Tietje, H. and Weiss, C. 2003. Agriculture on the move: exploring regional differences in farm exit rates. Working Paper EWP 0308. Dept. of Food Economics and Consumption Studies, University of Kiel.

Hardwick, P. and Adams, M. 2002. Firm size and growth in the UK life insurance industry. The Journal on Risk and Insurance 69(4), 577-593

Heckman, J. J. 1979. Sample selection bias as a specification error. Econometrica 47(2), 153-161.

Key, N. and Roberts, M.J. 2003. Government payments and structural change in agriculture. Paper presented at the Annual Meeting of the AAEA, Montreal, Quebec, July 27-30, 2003.

Kimhi, A. and Bollman, R. 1999. Family farm dynamics in Canada and Israel: the case of farm exits. Agricultural Economics 21, 69-79.

Knight, K, Bassett, G. W. and Tam, M. 2002. Quantile models and estimators for data analysis, Metrika 55, 17-26.

Koenker, R. and Basset, G. 1978. Regression quantiles. Econometrica 46, 33-50.

Koenker, R. and Xiao, Z. 2000. Inference on the quantile regression process. Econometrica 81, 1583–1612.

Khmaladze, E. 1981. Martingale approach in the theory of goodness-of-fit tests. Theory of Probability and its Applications 26, 240–257.

Lotti, F., Santarelli, E. and Vivarelli, M. 2003. Does Gibrat's Law hold among young, small firms?. Journal of Evolutionary Economics 13(3), 213-235.

McErlean, S., Moss, J. Patton, M. Kostov, P. and Jack, C. 2004. Investigating the determinants of farm growth and survival, Paper presented at the AES annual conference, Imperial College, London, April.

Shapiro, D., Bollman, R.D. and Ehrensaft, P. 1987. Farm size and growth in Canada. American Journal of Agricultural Economics 69, 477-483.

Sutton, J. 1997. Gibrat’s Legacy. Journal of Economic Literature 35, 40-59.

Upton, M. and Haworth, S. 1987. The growth of farms. European Review of Agricultural Economics 14, 351-366.

Weiss, C.R. 1998. Size, growth, and survival in the upper Austrian farm sector. Small Business Economics 10, 305-312.

Weiss, C.R. 1999. Farm growth and survival: econometric evidence for individual farms in Upper Austria. American Journal of Agricultural Economics 81, 103-116.

Zheng J. X. (1998). A consistent nonparametric test of parametric regression models under conditioning quantile restrictions. Econometric Theory 14, 123-138.

Figure 1. Net farm income of Northern Ireland dairy farms

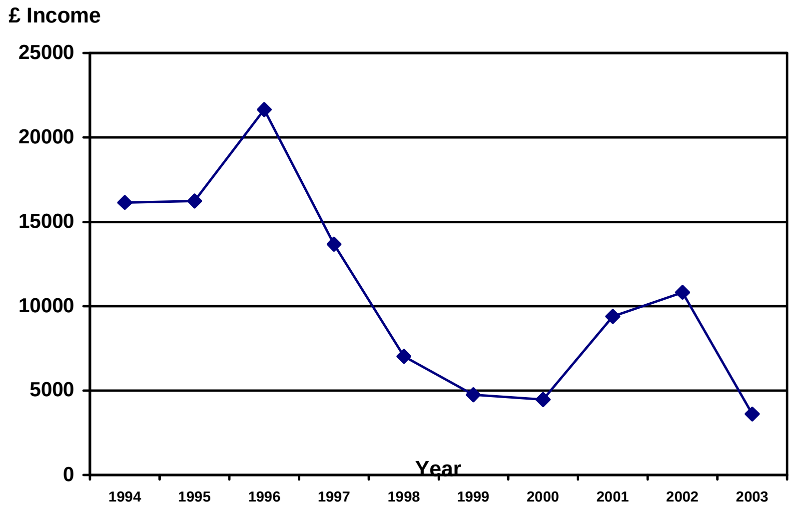

Figure 2. Graphical representation of the overall quantile regression results

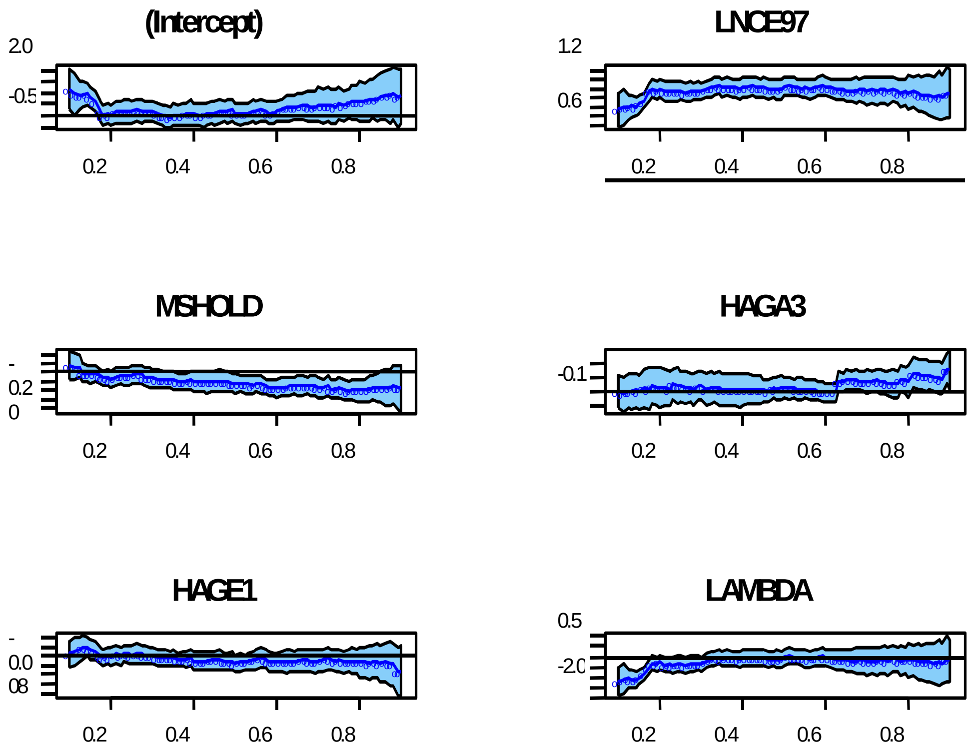

Figure 3. Quantile regression estimates on the lagged size coefficient

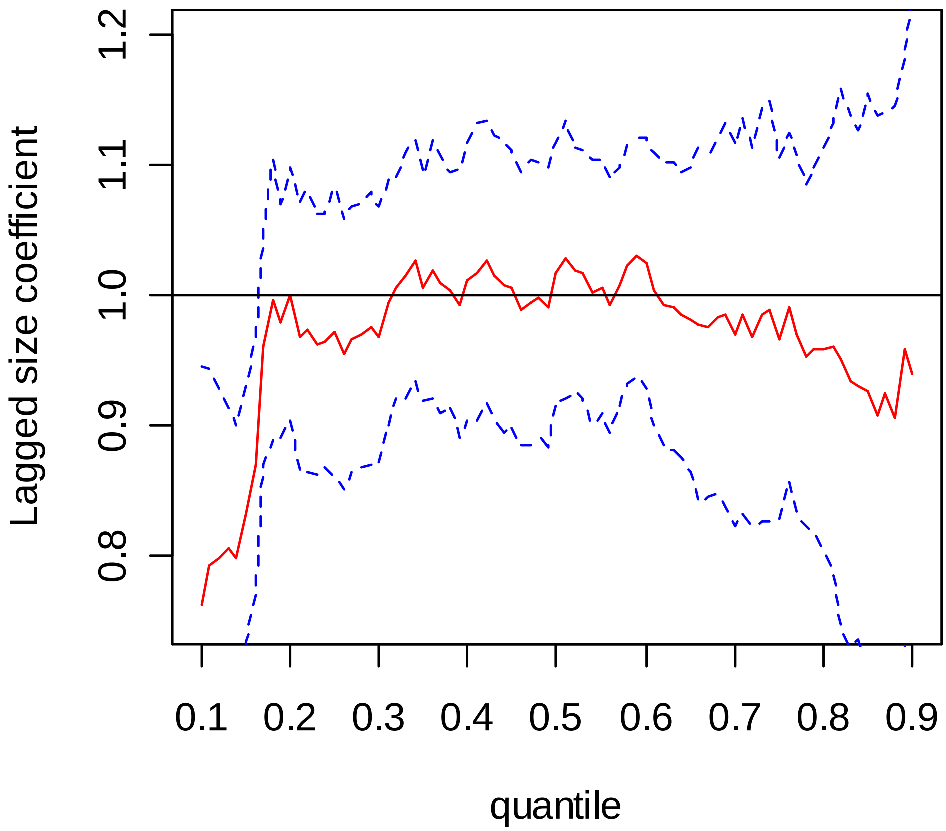

Table 1.First-step equation estimates

Variable	Coefficient	Standard Error
Constant	0.197	0.092
LNCE97	0.180	0.013
MSHOLD	-0.043	0.023
HAGA3	-0.009	0.050
HAGE1	-0.003	0.001

Table 2. ICM test results

Quantile	c=1	c=3	c=5	c=10	c=15	c=20
0.1	0.050	0.247	0.363	0.674	0.698	0.665
0.2	0.110	0.531	0.628	0.746	0.791	0.817
0.3	0.119	0.500	0.521	0.718	0.827	0.810
0.4	0.364	1.683	1.580	1.485	1.336	1.150
0.5	0.338	1.466	1.409	1.655	1.348	1.148
0.6	0.398	1.571	1.231	1.229	1.115	1.025
0.7	0.389	1.076	0.882	0.712	0.755	0.834
0.8	0.179	0.758	0.639	0.797	0.815	0.959
0.9	0.068	0.264	0.399	0.674	0.734	0.733

Critical values (Bierens and Ploberger, 1997):

10%	3.23
5%	4.26