

EFFICIENCY AND FARM SIZE IN WESTERN CAPE GRAPE PRODUCTION: POOLING SMALL DATASETS

Beatrice Conradie, Graham Cookson And Colin Thirtle*

Abstract

A stochastic frontier is fitted to two panels of wine grape farms (34 in Robertson and 36 in Worcester) for 2003 and 2004 and a cross section of 37 table grape farms in De Doorns for 2004 only. The panel of wine grape farms is the best model as tests show that De Doorns is different. Output is explained by land, labour and machinery and efficiency is affected by labour quality, age and education of the farmer, location, the percentage of non-bearing vines and expenditures on electricity for irrigation. There is evidence of a small degree of increasing returns to scale.

JEL Classification: Q12

1. INTRODUCTION

Battese and Coelli's (1995) inefficiency model, in which the efficiency differences are simultaneously estimated from the stochastic frontier and explained by further variables, also incorporates tests that choose between functional forms and between frontier and traditional models. This model is here applied to five small samples of grape producers in the Western Cape province of South Africa. The data covers two panels of wine grape farms (34 in Robertson and 36 in Worcester) for 2003 and 2004 and 37 table grape farms in De Doorns for 2004 only. The two years were similar, with no unusual weather, and the three regions are located close together, with all farms using irrigation. These data cover outputs, inputs and farm specific characteristics that can be used to explain efficiency at the farm level. The focus of the paper is simple. The majority of production frontier studies in agriculture pool cross section and time series data or use panel techniques to improve results. The recent literature on panel data (Baltagi, 2005) has begun to pay more attention to tests that determine whether data should be pooled. Here, we exploit the fact that despite the small samples, these data are good enough to produce acceptable results without pooling, to see if pooling tests are useful in determining what level of aggregation to use.

The paper proceeds as follows. The next section reviews the salient features of grape production in the three regions, with the aid of summary statistics and describes the data used for estimation. Section three outlines the stochastic frontier model with inefficiency effects and reports on the hypothesis tests for model selection. The fourth section reports the results and is followed by a brief conclusion.

2. SUMMARY STATISTICS AND VARIABLE DEFINITIONS

The descriptive statistics for this sample presented by Conradie (2005) show that wine grapes are different from table grapes in several respects. Wine farms are significantly larger in terms of output and fruit produced, although table grape operations employ more workers. The average wine farm in these four samples produces 1,140 tons of wine grapes and 340 tons of other fruit, consisting mostly of apricots, peaches, plums and pears. The average table grape farm produces 1,008 tons of table grapes and 28 tons of other fruit, including wine grapes and citrus. The cost of production is R2,515 per ton for the sample as a whole, but is significantly higher in the table grape industry (R6,255/ton) than for the wine industry (R1,550/ton). This is not surprising given that table grapes are packaged on the farm while most wine grapes are processed centrally. Table grape costs include packing, cooling and shipping while cellar costs are not included in the wine industry data. The estimated processing costs are roughly R900 per ton, so even net of this figure table grapes are far more expensive to produce than wine grapes. This is balanced out by the price, which is about 3.7 times the cost of wine grapes.

Table 1 reports the summary statistics for these samples, in terms of the variables used in estimation. Thus, following the convention of keeping the inputs in physical terms, the outputs are expressed in terms of tons of wine grape equivalents, with fruit and table grapes converted to wine grapes at average relative prices. The first column shows that the average output per farm is similar for the four wine grape samples, while the table grape farms show more variance and produce almost twice as much, due to the higher price of table grapes. This is despite the average farm size, in terms of area, which is far smaller than Worcester, while Robertson is in between. The greater input intensity of table grapes is again evident in labour use, which despite the smaller farms, is about three times that of the wine farms. Use of machinery (tractors, plus a few diggers and harvesters) is fairly uniform across the samples and machinery was the third and last input in most estimates. However, fuel costs were also recorded and for table grapes these proved a better measure of machinery use. As the Table shows, about twice as much fuel was used in table grape production, which suggests more intensive use of the available machinery. For the panel of wine farms, the best measure of machinery input proved to be the more sophisticated service flow from the capital stock, which was taken to be 10% depreciation on the machinery value, plus the running costs, represented by fuel expenditures.

The next five variables are the farm-specific factors that are used to explain the efficiencies in the second part of the model. The first is the average wage, which varies as some farms employ more skilled labour. The Table shows that wages are higher in Worcester than in Robertson and highest in De Doorns. An analysis of labour differences and wages can be found in Conradie (2005), but here the wage serves to pick up the lack of quality adjustment in the labour variable. Age and education of the farmer are both uniform across the samples, but three farmers have over 17 years of education. They are older and have “semi-retired” into farming.

One difficulty of measuring efficiency for a permanent crop like grape vines is partly captured by the percentage of the hectareage which is too recently planted to be yielding grapes. This varies from zero to two thirds of the farm in one case and as the new vines still use inputs, this must affect efficiency, as inputs are used on vines that are as yet unproductive.

Table 1. Summary Statistics for the Samples

Variable	Output	Land	Labour	Machines	Fuel	Wage	Age	Education	Non-bearing	Electricity	Output/ Land	Output/ Labour	Land/ Labour	Machinery/ Labour
Unit of Measurement	Kgs of Grape Equivalent	Hectares	Full Time persons	Number	Rand	R per month	Years	Years	Percent	Rand				
	Robertson – 34 wine farms 2003													
Mean	1887.08	87.92	33.21	5.74	67.10	908.13	41.91	14.26	0.16	91.62	22.58	57.45	2.75	0.09
SD/Mean	0.73	0.79	0.75	0.58	0.92	0.24	0.25	0.14	0.52	0.99	0.44	0.42	0.48	0.55
Minimum	102.85	13.20	7.56	2.00	10.00	562.19	28.00	10.00	0.00	6.56	7.44	11.97	1.02	0.04
Maximum	5482.50	299.00	129.17	16.00	350.00	1400.70	64.00	20.00	0.41	374.62	52.70	103.60	5.79	0.20
	Robertson – 34 wine farms 2004													
Mean	1982.98	94.80	34.80	5.97	70.23	1031.65	41.91	14.26	0.16	98.43	22.59	56.97	2.77	0.09
SD/Mean	0.74	0.79	0.71	0.58	0.93	0.34	0.25	0.14	0.52	0.96	0.37	0.41	0.49	0.60
Minimum	273.13	12.20	10.12	2.00	16.00	456.08	28.00	10.00	0.00	8.50	9.04	13.47	0.60	0.03
Maximum	5845.00	330.00	113.90	16.00	380.00	1733.33	64.00	20.00	0.45	390.00	44.49	113.57	6.26	0.20
	Worcester – 36 wine farms 2003													
Mean	1675.59	101.30	35.15	6.25	79.21	923.18	41.08	14.68	0.17	80.96	16.17	47.38	2.99	0.08
SD/Mean	0.73	0.62	0.69	0.66	0.66	0.31	0.21	0.11	0.51	0.69	0.28	0.35	0.29	0.71
Minimum	100.00	6.00	4.50	2.00	8.38	267.63	24.00	12.00	0.00	7.59	7.02	17.76	1.20	0.03
Maximum	5727.05	312.00	110.00	24.00	205.40	1605.61	57.00	17.00	0.42	260.00	27.40	87.88	4.80	0.33
	Worcester – 36 wine farms 2004													
Mean	1846.82	104.23	35.87	6.58	83.72	1058.32	41.08	14.68	0.20	95.38	16.96	50.63	3.08	0.08
SD/Mean	0.76	0.63	0.70	0.64	0.73	0.30	0.21	0.11	0.54	0.72	0.25	0.31	0.32	0.72
Minimum	95.00	5.50	5.50	2.00	10.00	268.00	24.00	12.00	0.04	8.00	7.82	17.27	1.00	0.03
Maximum	7124.80	324.50	113.10	25.00	261.30	1933.94	57.00	17.00	0.67	337.00	29.48	85.79	5.25	0.36
	De Doorns – 37 table grape farms 2004													
Mean	3852.56	60.80	119.20	6.70	158.6	1239.92	41.21	14.38	0.16	175.86	69.08	36.25	0.55	0.07
SD/Mean	1.03	1.27	1.34	0.94	1.33	0.33	0.19	0.12	0.75	1.41	0.24	0.35	0.42	0.43
Minimum	965.70	15.00	24.25	2.00	15.00	352.11	27.00	12.00	0.00	4.02	35.34	16.57	0.30	0.05
Maximum	22477.50	452.00	969.76	36.00	1100.00	2839.01	56.00	20.00	0.50	1500.00	106.38	74.16	1.29	0.28

Thus, there is normally an expectation that farms that are heavily involved in the establishment phase may appear to be less efficient. Expenditures on electricity which is used mainly for irrigation systems, are also far higher in table grape production. The next two columns report land and labour productivity. Land productivity is highest in De Doorns and lowest in Worcester, while labour productivity is lowest in De Doorns and highest in Robertson. Finally, the last two columns show factor ratios. The land/labour ratio is far lower in table grapes, while the machinery/labour is far less different across the samples.

3. CHOICE OF MODEL, FUNCTIONAL FORM AND LEVEL OF AGGREGATION

The measurement of firm level technical efficiency has become commonplace with the development of frontier production functions, as the survey by Battese (1992) shows. Fried, Lovell and Schmidt (1993) provide a comprehensive survey of methods and applications. Following Aigner, Lovell and Schmidt (1977) the general form of the production frontier, for a set of panel data is

$$Y_{it} = a + \sum_{j=1}^J \beta_j X_{jt} + e_{it} \quad / \quad e_{it} = v_{it} - V_{it} \quad (1)$$

with $u \sim N(0, \sigma_u^2)$ and $v \sim N(0, \sigma_v^2)$

where Y_{it} is output, the X_{jt} are inputs and the two elements of the error term e_{it} are independent. The v 's are independently and identically distributed random error terms and uncorrelated with the regressors, and the u 's are non-negative random variables associated with the technical inefficiency of the firm.

In Battese and Coelli's (1995) inefficiency effects model, the u 's, in equation (1) are assumed to be non-negative random variables that reflect the technical efficiency of production. They are assumed to be independently distributed and are obtained by truncating the normal distribution, with mean μ , and variance σ^2 , at zero. Thus,

$$u_{it} = z_{it}'\delta + w_{it} \quad (2)$$

where Z_{it} are explanatory variables associated with firm level technical inefficiencies in production, δ are unknown parameters to be estimated and w 's are the errors. The variance, σ_u^2 , is also an unknown parameter to be estimated.

First, the functional form of the stochastic frontier is determined by testing the adequacy of the Cobb-Douglas relative to the less restrictive translog. These frontier models are defined as

$$Y_{it} = f_0 + \sum_{j=1}^J \beta_j X_{jt} + \sum_{j=1}^J \sum_{k=1}^K \beta_{jk} X_{jt} X_{kt} + v_{it} \sim U_{it} \quad (3)$$

where all of the variables are in logarithms and if terms under the double summation are not significantly different from zero, the translog reduces to the Cobb-Douglas. Y_{it} is grape output in physical terms and the independent variables (X_{jt}) are land, labour and machinery. This gives nine independent variables in the translog due to the addition of three squared terms and three cross products. In the inefficiency model, there are five explanatory variables, which are the wage rate, farmer's age and education, the percentage

of the farm area planted with non-bearing vines and electricity expenditure. The remaining two variables are regional dummies for Worcester and De Doorns, to allow for regional variations.

The maximum-likelihood estimates (MLE) of the parameters in the Cobb-Douglas and translog stochastic frontier production function models defined by (3), given the specifications for the technical inefficiency effects defined by (2), were obtained using FRONTIER 4.1 (Coelli, 1994). First, a series of hypothesis tests were conducted to select the level of aggregation, the functional form and to choose between the frontier model and the standard average production function. The results reported in Table 2 are interdependent, in the sense that functional form and frontier test results are used in the pooling tests. For the functional form tests the null hypothesis (H_0) is that $\mathbf{jS}^{\wedge} = 0, i, j = 1, \dots, n$, meaning that the Cobb-Douglas frontier is an adequate representation for these data. Generalised Likelihood Ratio (LR) tests show that the Cobb-Douglas is an adequate representation of the data for all five grape samples, as \mathbf{X} is less than the critical value. However, for the three panels that the pooling tests allow, the translog is preferred in two. The problem here is that the results for the translog do not comply with the theoretical restrictions for any production function. The basic requirement is that the coefficients of the three inputs must all lie between zero and unity, since they are output elasticities. Thus, the Cobb-Douglas results are preferred, despite the tests.

Having selected the Cobb-Douglas functional form, the next section of Table 2 reports the results of tests of the hypothesis that the technical efficiency effects are not simply random errors. The key parameter is $\mathbf{J} = (\mathbf{J}_U^2 / (\mathbf{J}_U^2 + \mathbf{O}^2))$, which is the ratio of the errors in equation (1). So, \mathbf{J} is defined between zero and one, where if $\mathbf{J} = 0$, technical inefficiency is not present, and if $\mathbf{J} = 1$, there is no random noise. The null hypothesis is thus that $\mathbf{J} = 0$, indicating that the mean response function (OLS) is an adequate representation of the data, whereas the closer \mathbf{J} is to unity, the more likely it is that the frontier model is appropriate. If \mathbf{J} is not significantly different from zero, the variance of the inefficiency effects (W_i in equation 2) is zero and the model reduces to a mean response function in which the inefficiency variables enter directly (Battese and Coelli, 1995). This test is unambiguous, with all values close to unity and all t tests indicating that the frontier is the appropriate model. The next column in this section reports \mathbf{X} , the LR test values for the more powerful test with the null hypothesis that $\mathbf{y} = \mathbf{S}_0 = (\mathbf{S}; = 0$, which means that in addition to \mathbf{J} being insignificant, the inefficiency effects are not present in the model. The null hypothesis, H_0 , is soundly rejected in all cases at the 5% level, with degrees of freedom (DOF) equal to the numbers of parameters set to zero.¹

In the last section, LR tests determine the extent to which the five samples can be pooled, or estimated as a panel. The test is that suggested by Battese and Coelli (1988), which compares the LR for the pooled model (H_0) with the sum of the LRs for the sub-samples estimated separately (H_1). Thus, the LR when both Robertson years are pooled is -1.382 , compared with 8.815 (the sum of the two H_0 LRs in the functional form test, below), giving a test statistic (X) of 19.134 . This is compared with the critical χ^2 value at the 5% significance level, with 12 DOF. The DOF is the number of parameters estimated, which is 12 (see Table 3) times by the difference in the number of

¹ As the null hypothesis involves parameter g , which as a ratio of two variances is necessarily positive, the test statistic follows a mixed chi-squared distribution and the critical values are found in Kodde and Palme (1986).

Table 2. Hypothesis Tests

Functional Form Test		Log-Likelihoods		LLR Test Statistic	DOF	χ^2 Critical Value 5%	Outcome
	Parameter Restrictions	H ₀ Cobb-Douglas	H ₁ Translog				
Robertson 2003	H ₀ : All $\beta_k = 0$	0.569	4.870	8.602	6	12.59	
Robertson 2004		7.616	8.211	1.190	6	12.59	Accept H ₀ – CD is adequate
Worcester 2003		5.920	12.210	12.580	6	12.59	Accept H ₀ – CD is adequate
Worcester 2004		16.490	17.051	1.122	6	12.59	Accept H ₀ – CD is adequate
De Doorns 2004		7.254	11.860	9.212	6	12.59	Accept H ₀ – CD is adequate
Robertson both years		-1.382	10.893	24.550	6	12.59	Reject H ₀ – CD is inadequate
Worcester both years		19.494	25.312	11.636	6	12.59	Accept H ₀ – CD is adequate
All 4 wine samples		-4.454	11.880	32.668	6	12.59	29003
Frontier Tests		Gamma	t stat	LLR Test Statistic	DOF	χ^2 Critical Value 5%	Outcome
	H ₀ : $\gamma = \delta_i = 0$						
Robertson 2003		0.999	1911.000	40.893	7	13.401	Reject H ₀ – frontier not OLS
Robertson 2004		1.000	277.393	34.782	7	13.401	Reject H ₀ – frontier not OLS
Worcester 2003		1.000	624.090	31.835	7	13.401	Reject H ₀ – frontier not OLS
Worcester 2004		1.000	16.291	21.500	7	13.401	Reject H ₀ – frontier not OLS
De Doorns 2004		0.895	102.464	141.554	3	7.054	Reject H ₀ – frontier not OLS
Robertson both years		0.952	22.808	51.381	7	13.401	Reject H ₀ – frontier not OLS
Worcester both years		0.958	19.107	19.494	7	13.401	Reject H ₀ – frontier not OLS
All 4 wine samples		1.000	16017.000	48.461	7	13.401	Reject H ₀ – frontier not OLS
Pooling Tests		H ₀ Pooled	H ₁ Separate	LLR Test Statistic	DOF	χ^2 Critical Value 5%	Outcome
Robertson both years	Cobb-Douglas	-1.382	8.185	19.134	1 * 12 = 12	21.03	Accept H ₀ – can pool
Worcester both years	Cobb-Douglas	19.494	22.410	5.832	1 * 12 = 12	21.03	Accept H ₀ – can pool
Both regions 2003	Cobb-Douglas	-9.611	6.489	32.200	1 * 12 = 12	21.03	Reject H ₀ – can't pool
Both regions 2004	Cobb-Douglas	9.550	24.106	29.112	1 * 12 = 12	21.03	Reject H ₀ – can't pool
All 3 regions 2004	Translog	19.045	31.360	24.630	2 * 12 = 24	49.77	Accept H ₀ – can pool
All 3 regions 2004	Cobb-Douglas	4.511	31.360	53.698	2 * 12 = 24	36.42	Reject H ₀ – can't pool
All 4 wine samples	Cobb-Douglas	-4.454	30.595	70.098	3 * 18 = 48	73.00	Accept H ₀ – can pool
All 5 samples	Translog	1.928	37.849	71.842	4 * 18 = 72	92.8	Accept H ₀ – can pool
All 5 samples	Cobb-Douglas	-10.479	37.849	96.656	4 * 12 = 48	67.00	Reject H ₀ – can't pool

Table 3. Stochastic Production Frontier and Inefficiency Model Results

	Robertson 2003	Robertson 2004	Robertson 2003, 2004	Worcester 2003	Worcester 2004	Worcester 2003, 2004	De Doorns 2004	Panel – all 4 wine samples
Stochastic Frontier	Parameter (t stat)	Parameter (t stat)	Parameter (t stat)	Parameter (t stat)	Parameter (t stat)	Parameter (t stat)	Parameter (t stat)	Parameter (t stat)
Intercept	0.531 (29.977)	0.204 (1.826)	0.444 (9.171)	0.099 (1.270)	0.131 (1.340)	2.646 (18.116)	0.731 (0.749)	3.327 (0.236)
Land	0.135 (1.694)	0.653 (9.439)	0.385 (5.875)	0.867 (7.425)	0.758 (9.356)	0.842 (12.058)	0.758 (11.743)	0.758 (12.484)
Labour	0.577 (9.444)	0.384 (3.555)	0.524 (5.466)	0.348 (1.580)	0.284 (3.196)	0.319 (3.589)	0.566 (2.575)	0.321 (4.224)
Machinery	0.100 (2.059)	0.171 (1.605)	0.032 (0.347)	0.056 (0.383)	0.104 (1.677)	0.067 (0.897)	0.174 (2.132)	0.137 (1.676)
Sum	0.812	1.207	0.941	1.271	1.146	1.228	1.498	1.215
Inefficiency Model	Parameter (t stat)	Parameter (t stat)	Parameter (t stat)	Parameter (t stat)	Parameter (t stat)	Parameter (t stat)	Parameter (t stat)	Parameter (t stat)
Intercept	-2.751 (-2.445)	-5.987 (-2.654)	-6.461 (-2.318)	0.043 (0.044)	1.446 (2.055)	0.860 (0.658)	0.659 (0.436)	-1.405 (-0.099)
Wage	-0.667 (-2.512)	-0.247 (-1.554)	-0.510 (-1.648)	-0.222 (-0.818)	-0.206 (-2.368)	-0.376 (-3.389)	-	-0.158 (-2.369)
Electricity	-0.916 (-5.185)	0.129 (1.304)	-0.313 (-2.228)	0.222 (0.780)	-0.015 (-0.216)	0.752 (3.244)	-0.868 (-3.017)	0.441 (1.507)
% Non-bearing	-4.790 (-3.275)	-0.386 (-0.563)	-1.771 (-1.839)	0.142 (0.182)	0.862 (3.033)	-0.633 (-1.939)	-	0.635 (-3.337)
Farmer's age	1.456 (2.524)	0.374 (1.362)	0.781 (2.161)	0.473 (1.497)	0.310 (2.399)	0.404 (1.172)	-	0.165 (4.691)
Education	2.447 (2.780)	2.245 (3.170)	3.289 (3.615)	-0.288 (-0.597)	-0.331 (-1.332)	0.198 (2.799)	-	-0.871 (3.810)
Sigma-square	0.224 (2.807)	0.038 (2.594)	0.163 (2.985)	0.057 (2.028)	0.026 (4.335)	0.057 (3.690)	0.485 (4.565)	0.072 (8.616)
Gamma	1.000 (1911)	1.000 (277)	0.952 (22.808)	1.000 (624)	1.000 (16.291)	0.958 (19.107)	0.895 (102.464)	1.000 (16017)
Worcester dummy	-	-	-	-	-	-	-	0.194 (3.783)
Critical value 5%	1.717	1.717	1.671	1.711	1.711	1.671	-	1.658
Critical value 10%	1.321	1.321	1.296	1.318	1.318	1.296	-	1.289

estimating equations, which is two, minus one. The outcome is close, but the two can be pooled, as can the two years for Worcester, with greater certainty. However, the two wine regions should not be pooled in either year, which is a little surprising since the years were not very different. The three regions can be pooled for 2004 if the function is translog, but this gave unacceptable results and is not pursued further. This was also the case with pooling all five samples, so the only high level of aggregation allowed is all four wine samples, which narrowly qualifies even with a Cobb-Douglas function. These tests explain why only three panels are reported in the next section.

4. STOCHASTIC PRODUCTION FUNCTION AND INEFFICIENCY MODEL RESULTS

(i) *Output Elasticities, Returns to Scale and Farm Size* For all five samples, the Cobb-Douglas function was found to be an adequate representation of the unknown, underlying production function, meaning that the cross products and squared terms did not improve the fit sufficiently to justify inclusion. Table 3 reports the parameter estimates and t statistics for these models, beginning with the output elasticities for the inputs.

For Robertson in 2003, all three elasticities are significant at the 5% level and a 1% increase in labour increases output by 0.577%.² Land is far less important and machinery contributes still less, so that the elasticities sum to only 0.812, which indicates that on average, there is decreasing returns to scale.

In the second year, land becomes the dominant input, labour falls and machinery is still last, but the sum is 1.207, which would suggest increasing returns to scale. With samples of only 34, it is perhaps not surprising that the results are so unstable, so forming a two-year panel to reduce the variability makes perfect sense. This results in more reasonable elasticities for land and labour and a sum that is much closer to constant returns to scale, but at the cost of machinery being insignificant. Aggregation by panel estimation may well be inferior to simply aggregating the two previous results. The pooling test is also odd, in that it allowed pooling despite such different slope coefficients. Were the two sub-samples larger, such aggregation could well be destroying real information rather than improving the estimates.

The Worcester results are less different, with land dominating both years and machinery contributing least, so it is no surprise that pooling was permitted, but in this circumstance it really isn't needed and rather than improving the significance of machinery, it makes this input insignificant. Again, a simple average would perhaps have been preferable. For De Doorns, the three elasticities all have reasonable values and are significant, but the sum of 1.489 is rather too much evidence of increasing returns.

The last results are for a panel comprised of both regions in both years. The programme for the inefficiency model does not handle panels, but equivalent results are obtained by using time and regional dummies. All three inputs have reasonable elasticities and are significant at the 5% level, while there is still evidence of increasing returns to scale and the time dummy proved insignificant.

² The t test critical values at the 5% and 10% levels are shown at the bottom of the Table. The test is one tailed as the elasticities must be positive.

In many papers, where the data refuses to cooperate, this panel could well have been the only results reported, but in this case the small samples gave good results, so the pooling issue could be examined. The pooling tests are somewhat useful: for instance, confirming the impression given by the summary statistics, that table grapes really are different from wine grapes. This conforms with the expectation stated in section two, based on the differences between the two crops and we thank an anonymous referee for pointing out that De Doorns also has a different micro climate, being considerably colder. However, it is not clear how well the tests guide the researcher beyond this point. Not allowing pooling of the two regions in either single year, when the small sample sizes suggest that this form of aggregation may be needed, is unfortunate. Yet when the two years have already been pooled and the sample size has been increased, further pooling is allowed when it is much less needed.

The sum of the output elasticities provides an indication of the predominant scale effect in a sample, but it is an average and can be quite misleading, if farms that are too large and those that are too small balance out. The Frontier programme calculates an efficiency level for each farm, so if the farms are then ranked according to size, some indication of the effect of size on efficiency can be gained. However, just as yield is a partial measure of productivity, returns to scale is output per unit of all inputs, not just land. Data envelopment analysis could be used to calculate scale efficiency, but this paper merely notes that the efficiency levels from the stochastic frontier estimation do appear to be a monotonically increasing function of farm size. The quartile of smallest farms has an average efficiency of 65%, the next quartile 71%, the next 75% and the largest 76%, so the scale effects are relatively unimportant, when judged according to land area. This conforms with the only previous study of scale efficiency in wine grape production, by Townsend, Kirsten and Vink (1998), which found no consistent relationship between yield, labour productivity, or total factor productivity and farm size, which was measured by area or number of workers. In their Robertson and Worcester samples, the small and medium sized farms were often more efficient than the largest farms.

(ii) *Explaining the Inefficiencies* The same variables are reported for each sample in explaining the inefficiencies, to facilitate comparisons, with the exception of De Doorns, where only electricity was significant. For all but De Doorns and Worcester in 2003, higher average wages decrease inefficiency (hence the negative coefficients). This is to be expected, since this serves as quality adjustment for the labour input. Expenditures on electricity reduced inefficiency in De Doorns, which suggest that irrigation is important for table grapes. For Robertson the same effect dominates, but for Worcester pooled and the full pooled sample the sign is positive. This may well reflect the locations of the farms, as those on higher ground will have to spend more on pumping irrigation water.

The same type of problem arises with non-bearing vines, which increase inefficiency, in accordance with the conventional wisdom, only in Worcester in 2004. In the four cases in which non-bearing vines increase efficiency, the causality may run from efficient production to planting new vines. This follows, as any farmer who intends to continue producing has to do some replanting almost every year. If a farm is prospering, it is also likely to be investing, so it seems to be those that are inefficient that are not investing. Thus, the dynamics of the situation reverse the expected static result. Age and education often give odd results in these models and here age reduces efficiency, but so too does education, in Robertson especially. As was noted above, wine farmers with 17 years of

education may have bought vineyards late in life as an attractive retirement lifestyle (the prospect certainly appeals to the authors). The last statistics are the variance parameters $s^2 = s_u^2 + s_v^2$ and $g = s_u^2/s_v^2$ which do not require further discussion.

5. CONCLUSIONS

This paper uses a stochastic frontier and inefficiency model to test the efficiency of grape production in the Western Cape. The data cover two panels of wine grape farms (34 in Robertson and 36 in Worcester) for 2003 and 2004 and 37 table grape farms in De Doorns for 2004 only. Tests show that Cobb-Douglas stochastic production frontiers, with variables to explain the inefficiencies are an appropriate representation of the five individual samples.

The stochastic frontier results indicate that output can be explained by land, labour and machinery and that efficiency can be affected by labour quality, age and education of the farmer, location, the percentage of non-bearing vines and expenditures on electricity for irrigation. Efficiency is also dependent on farm size, so returns to scale are further investigated, showing that there are very modest returns to scale in grape production.

These data are sufficiently good to produce reasonable results without pooling, but most applied economists would consider the possibility of improving the estimates by pooling the samples. However, pooling tests show that in this situation with small samples, when pooling is permissible it may not be helpful and when it is needed it may not be permitted. More effort on determining the true distributions is needed to improve the way small samples are handled and Bayesian methods may be helpful in this respect.

REFERENCES

- AIGNER, D., LOVELL, K. and SCHMIDT, P. (1977). Formulation and Estimation of Stochastic Frontier Models, *Journal of Econometrics*, 6, 21-37.
- BALTAGI, B. H. (2005). *Econometric Analysis of Panel Data*, John Wiley, New York.
- BATTESE, G. and COELLI, T. (1988). Prediction of Firm-Level Technical Efficiencies with a Generalised Frontier Production Function and Panel Data, *Journal of Econometrics*, 38, 387-399.
- BATTESE, G. (1992). Frontier Production Functions and Technical Efficiency: a Survey of Empirical Applications in Agricultural Economics, *Agricultural Economics*, 7, 185-208.
- BATTESE, G. and COELLI, T. (1995). A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data, *Empirical Economics*, 20, 325-332.
- COELLI, T. (1994). *FRONTIER Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation*, Department of Econometrics, University of New England, Armidale, NSW.
- CONRADIE, B. (2005). Wages and wage elasticities for wine and table grapes in South Africa, *Agrekon*, 44:1, 138-156.
- FRIED, H., LOVELL, C. A. K. and SCHMIDT, S. (1993). *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, New York.
- KODDE, D. and PALME, F. (1986). Wald Criteria for Jointly Testing Equality and Inequality Restrictions, *Econometrica*, 54:5, 1243-1248.
- TOWNSEND, R., KIRSTEN, J. and VINK, N. (1998). Farm Size, Productivity and Returns to Scale in Agriculture Revisited: A Case of Wine Producers in South Africa, *Agricultural Economics*, 19, 175-180.