Application of Chow test to improve analysis of farmer participation in markets in Kenya

David Jakinda Otieno[1*, 2], John Omiti[3], Timothy Nyanamba[4] and Ellen McCullough[5]

[1]Department of Agricultural Economics and Food Marketing, Newcastle University, NE1 7RU, UK, **corresponding author* (d.j.otieno@newcastle.ac.uk)

[2]Department of Agricultural Economics, University of Nairobi, Kenya

[3]Kenya Institute for Public Policy Research and Analysis (KIPPRA), Nairobi, Kenya

[4]Department of Agricultural Economics and Agribusiness Management, Egerton University, Kenya

[5]Agricultural Development Division, Bill and Melinda Gates Foundation, USA

A contributed paper prepared for presentation at the 27th Conference of the International Association of Agricultural Economists (IAAE), 16 – 22 August 2009, Beijing, China

Application of Chow test to improve analysis of farmer participation in markets in Kenya

1. Introduction and aims of the study

High population growth rates and emergence of urban settlements provide opportunities as well as challenges to economic development in many countries. While the supply of labour and markets are potential positive outcomes of these processes, the underlying pressure on scarce resources is often intense. In low-income agriculture-dependent countries such as Kenya, inadequate food supply and lack of other basic social amenities characterize a large share of rural and urban population. The productive capacity and commercial orientation of agriculture and food-related sectors need to be improved in order to reduce famine, malnutrition and poverty. This would entail enhancing farmers' access and participation in both input and out markets.

Horticulture (especially vegetables) is one of the most important sectors in Kenya, where smallholder farmers account for nearly 70% of the output (McCulloch and Ota, 2002). About 23% of Kenya's export revenue is derived from horticultural exports (CBS, 2006; Minot and Ngigi, 2003). Cultivation of vegetable crops (mainly cabbages, tomatoes, kales – *sukuma wiki*, onions and a variety of indigenous vegetables such as amaranthus) forms a crucial source of livelihood for many households in rural and peri-urban areas of Kenya (Omiti *et al.*, 2004).

Promoting investments in agricultural commercialization, more so in developing marketing channels are critical for poverty reduction (Geda *et al.*, 2001). The potential benefits of higher product prices and lower input prices due to commercialization are effectively transmitted to poor households when markets function fairly (IFAD, 2001). In Kenya, recent research show evidence that prioritizing infrastructure development for vegetable production and marketing are necessary for improvement of most livelihoods (Omiti *et al.*, 2006).

Recent transformations in agri-food systems (particularly the rise of supermarkets and technological advances in developing countries' agriculture during the last decade) offer opportunities for smallholder farmers (McCullough *et al.*, 2008). However, these prospects might be countered by population pressure, on-going global economic downturn and the adverse effects of climate change, if alternative policies and strategies are not urgently instituted to reverse the decline in real purchasing power of many households (Food Ethics Council, 2008). In order to support the process of sustained economic growth, there is need for a more refined and targeted analysis of pertinent issues that constrain farm-to-market distribution of food. The analytical role of agricultural economists must therefore expand to comprehensively capture site-specific dynamics of the agri-food systems.

Previous studies on market participation (for example, Alene *et al*., 2008; Chianu *et al*., 2008; Makhura *et al*., 2002) have been based on single or multiple sites. However, the decision to pool data or perform separate analysis is often subjective. In these studies, authors provide elaborate discussions to differentiate sites in terms of geographic features, climatic conditions and socio-economic profiles. Although the findings from such studies might offer useful insights on necessary policies, they lack rigorous objective criteria to support the choice between pooled versus disaggregated analyses. It is important to anchor market analysis on solid statistical criteria in order to give more credence to the resulting site-specific or nation-wide strategies. This would enable implementation of policy interventions that reliably address salient challenges which may vary across sub-regions within a country. As a standard practice, data from multiple sites should be tested to confirm similarities or differences, and to guide the process of data organization (pooling or separation) for analysis.

This study contributes to knowledge on farm-level analysis of market participation through application of the *Chow's seminal* test (Chow, 1960) to examine differences between data from two sites (rural and urban). A truncated regression model is applied in the analysis.

The specific objectives of the study are:

i. To assess difference in the level of market participation between rural and peri-urban farmers;
ii. To estimate factors that influence the share of vegetable marketed by farmers.

2. Methodological approach

2.1 Study sites

This paper reports results from a study done in one rural and one peri-urban area in Kenya; Kisii and Kiambu districts, respectively. These sites were chosen through consultation of experts at district and provincial agriculture offices, and other stakeholders from sixteen districts considered to be representative of Kenya's agricultural production and marketing conditions.

Kiambu district has a total land area of 1458.3 km^2, 97% of which is arable. About 90% of the arable land is under smallholdings while the rest is under large farms (Republic of Kenya, 2001a). Altitude ranges from 1500m to 2591m above sea level, while the average temperature is 26^{o}C. The average annual rainfall is 1239.6mm occurring in a bimodal pattern; long rains in April – May and short rains from October to November. The average population density was 526 persons per km^2 in 1999 (CBS, 2003). Kiambu District in Central Province was selected mainly because of its proximity to Nairobi, where there is potentially huge lucrative urban market for maize meal, dairy and horticultural products. Generally, food production systems in Kiambu are relatively more commercialized; considering its comparative advantage in most physical infrastructure (roads, water, electricity, *etc*.) compared to other parts of the country.

Kisii district on the other hand, has a total land area of 1200 km^2. The altitude ranges from 1000 to 1800m above sea level, with a mean temperature of 22^{o}C. There are two rainfall seasons; long rains in April – June and short rains in September – November, recording an average annual rainfall of 1500mm. About 78% of the land is arable; 58% of which is cropped (Republic of Kenya, 2001b). The average population density was approximately 647 persons per km^2 in 1999. Kisii District, about 400 km from Nairobi in south-western Kenya, is characterized by modest level of commercialization and relatively modest state of infrastructure.

2.2 Data and sampling

The study is based on primary data from a household survey. Sampling was done in a few villages within Kiambu municipality (peri-urban) and those that are outside Kisii municipality (rural and relatively remote). A purposive sample of seventy seven (77) vegetable (Kales) producers, who were selling different proportions of their output to specific channels were interviewed (37 in the rural area and 40 in the peri-urban area). The rural sample consisted of farmers selling vegetables mostly in open-air retail markets, while the peri-urban sample had farmers selling to wholesale markets. During the survey, households were selected randomly. Individuals who normally make decisions on farm management (household heads or one other member) were interviewed on a face-to-face approach. Data was gathered on percentage of Kales output sold, household socio-economic and farm characteristics. A pre-tested structured questionnaire was used to collect the data.

2.3 Analytical framework

i. The Chow test for difference in data sets

Equality of error variances in two linear regression equations is the main restriction assumed in the Chow test (Chow, 1960). This can be illustrated by use of two models for the rural and peri-urban sub-samples (Equation 1 and 2):

$$Y_g = X_g \beta_g + \varepsilon_g = X_g b_g + e_g \tag{1}$$

$$Y_j = X_j \beta_j + \varepsilon_j = X_j b_j + e_j \tag{2}$$

where X_i, (i = g, j for rural and peri-urban sub-samples respectively) are non-singular matrices of explanatory variables, β_i are column vectors of the K regression coefficients and Y_i are column vectors for the dependent variable. It is assumed that the stochastic terms ε_i's are normally distributed with zero mean and variance covariance matrix $\sigma^2 I$, whereby I represent an identity matrix, b and e are estimated vectors.

The Chow test is preferred due to less computational difficulty compared to alternative approaches suggested in literature such as co-integration tests (Campos *et al*., 1996), bootstrap procedure (Diebold and Chen, 1996), Bayesian techniques (Kozumi and Hasegawa, 2000) and comparison of slopes alone (Wilcox, 1997). Application of the Chow test requires that the number of observations in both sub-samples should be nearly the same. In situations where there is a significant difference in the number of observations between sub-samples and greater error variability in the two data sets, a transformation of the data is necessary to assure homoscedasticity before the test can be applied (Ghilagaber, 2004).

The main hypothesis in the Chow test is that the coefficients are equal for both sub-samples (Equation 3):

$$H_0 : \beta_g - \beta_j = 0 \tag{3}$$

Three linear regressions were fitted to operationalise the Chow test; one equation for the restricted model (pooled data) and separate regressions for the unrestricted models (rural and peri-urban data). The test statistic was formally stated as follows:

$$F^* = (RSS_w - (RSS_g + RSS_j))/(RSS_g + RSS_j) * (T - 2K)/K \tag{4}$$

where F* is the test statistic

RSS_w = residual sum of squares for the whole sample (restricted model)

RSS_g = residual sum of squares for the rural sub-sample

RSS_j = residual sum of squares for the peri-urban sub-sample

T = total number of observations in the whole sample

K = number of regressors (including the intercept term) in each unrestricted sub- sample regression

$2K$ = number of regressors in both unrestricted sub sample regressions (whole sample).

In the Chow test, if there is no significant statistical difference between two sub-samples (i.e., if $\sigma_g^2 = \sigma_j^2$), then the regression test statistic in Equation (4) follows an $F(K, T\text{-}2K)$ distribution. However, if the test statistic (F*) is greater than the respective *F-statistic* at 5% level of significance (as in this study), the null hypothesis should be rejected. Consequently the relevant conclusion is that the sub-samples are significantly different (Table 1). This is the statistical evidence which justifies the decision to estimate separate models for the sub-samples and even make comparisons with results of the whole sample analysis.

Table 1: Chow Test outcome

RSS_w	***RSS_g***	***RSS_j***	***F****	***F(K, T-K) at 5% significance level***	**Decision**
15251.69	8032.20	770.33	4.18	1.99	Separate rural and peri-urban data

Source: Survey data (2007)

ii. Descriptive analysis

Significant differences between means and frequencies for important variables that were considered to explain the intensity of market participation among vegetable farmers were tested using simple descriptive measures (Moore, 2006).

a) Means

The difference in means of variables is tested by the hypothesis in Equation 5.

$$H_0 : m_g - m_j = 0 \tag{5}$$

where m_g is the mean for rural sub-sample while m_j represents the mean for peri-urban sub-sample. The test statistic is given by:

$$z = m_x / \sigma_x \tag{6}$$

whereby m_x is the difference between the means of variables in the rural and peri-urban sub-samples ($m_x = m_g - m_j$) and σ_x is the joint standard deviation of both sub-samples given by Equation 7:

$$\sigma_x = \sqrt{s_g + s_j} \tag{7}$$

in which,

$$s_g = \sigma_g^{\ 2} / n_g \tag{8}$$

and $s_j = \sigma_j^{\ 2} / n_j$ (9)

Here, n_g and n_j denote the rural and peri-urban sub-sample sizes respectively.

b) Percentage frequencies

In order to establish differences in percentage frequencies, the test statistic z was calculated as:

$$z = (p_g - p_j) / \sqrt{f_g + f_j} \tag{10}$$

$$f_g = pq / n_g \tag{11}$$

$$f_j = pq / n_j \tag{12}$$

where p_g and p_j are percentages for variables in the rural and peri-urban sub-samples respectively, p is the percentage frequency in the whole sample and

$$q = 1 - p \tag{13}$$

iii. Truncated regression analysis

A truncated regression model was used to analyze factors that influence the percentage of vegetables sold by farmers. The truncated model follows normal distribution with homoscedastic error component (Greene, 2003).

$$Y_i^* = \beta_i X_i + \mu_i \tag{14}$$

where Y_i^* is the percentage of output that is sold by the individual, β_i is the vector of unknown parameters to be estimated, X_i is the set of independent variables and μ_i is the stochastic term. An observation of zero value for Y_i^* occurs when a household does not sell any output, while Y_i^* = 100 if a household sells all output. In the analysis, the dependent variable is truncated at zero (i.e., households who do not sell output were omitted at the sampling stage). Standard selectivity models such as Heckman's two-stage procedure where estimation of amount sold is conditional on the probability of market participation are therefore not applicable in this particular study. Some of the factors hypothesized to influence vegetable sales are summarised in Table 2.

Table 2: Variables hypothesized to influence vegetable sales

Variable	**Description**	**Measurement**	**Expected sign**
Age	Age of the household head	Number of Years	+
Gender	Gender of the household head	0 = female 1 = male	±
Education	Education level of the household head	0 = not completed secondary education 1 = completed secondary education	+
Household size	Number of people in the household	Number	±
Non-farm income	Proportion of non-farm income in total monthly household income	Ratio	±
Output	Total quantity of vegetable produced per season	Kilograms (Kg)	+
Distance	Average distance from farm to main point of sale	Kilometres (Km)	-
Market information	Market information source	0 = informal 1 = formal	±
Unit price	Average price per Kg	Kenya Shillings (Kshs)	+

Source: Survey data (2007)

3. Results and Discussion

Peri-urban farmers sell more vegetable output compared to their rural counterparts (Table 3). Thus, the intensity of market participation for vegetable is higher in peri-urban areas than in the rural areas. In this respect, the study corroborates assertions in literature that a rise in urban population and changes in life styles of emerging middle income households contribute to increased demand for more nutritious and convenience foods especially fresh vegetables.

Table 3: Means and standard deviations for main variables

Variable	**Rural (n=37)**		**Peri-urban (n=40)**		**Whole sample (n=77)**		**Test statistic** *z*
	Mean	σ	Mean	σ	Mean	σ	
Output sold (%)	62.91	17.42	94.95	5.7	79.56	20.49	10.68***
Age of household head (years)	45.27	16.35	43.33	13.37	44.26	14.81	0.57
Household size (number)	6.30	2.87	5.90	3.67	6.09	3.29	0.53
Per capita land (acres)	0.70	0.52	0.70	0.83	0.70	0.69	0.00
Nonfarm	0.16	0.28	0.12	0.23	0.14	0.25	0.68
Output (Kg)	3232.43	3252.87	1869.75	2961.23	2524.55	3159.36	1.75
Distance (Km)	8.68	7.12	2.82	2.83	5.63	6.07	3.71**
Unit price (Kshs)	14.24	3.52	19.98	8.13	17.22	6.93	4.07***

***p<0.01, **p<0.05

Source: Survey data (2007)

Rural farmers travel longer distances to the nearest market outlets, and sell their vegetables at lower prices compared to those located in peri-urban areas. The distribution of other important variables such as the average age of the household head, household size, per capita land, proportion of nonfarm income in total income, and the vegetable total output are not significantly different between farmers in both sites.

On average, more than 60% of farmers sampled in the study had male household heads (Table 4). However, most peri-urban household heads were male; 93% compared to 63% in the rural areas. The study did not find any significant differences in the distribution of levels of education, security of land tenure (possession of title deeds) and main source of market information between the rural and peri-urban farmers.

Table 4: Percentage frequencies for vegetable data

Variable		**Rural (n=37)**	**Peri-urban (n=40)**	**Whole sample (n=77)**	**Test statistic *z***
Gender	Male	62.20	92.50	77.90	3.20**
	Female	37.80	7.50	22.10	
Education	Completed secondary	51.40	57.50	54.50	0.54
	No secondary	48.60	42.50	45.50	
Security of land tenure	Has title deed	62.20	45.00	53.20	1.51
	No title deed	37.80	55.00	46.80	
Market information source	Formal	40.50	35.00	37.70	0.50
	Informal	59.50	65.00	62.30	

**p<0.05

Source: Survey data (2007)

The unit price significantly motivates farmers to increase the percentage of vegetable sold in both rural and peri-urban areas (Table 5). The total amount of vegetable produced and being a male head of a household also significantly increase the percentage sold, while distance reduces percentage marketed in rural areas and for the whole sample.

Table 5: Determinants of amount of vegetable sold

	Rural (n=37)		**Peri-urban (n=40)**		**Whole sample (n=77)**	
Variable	**β**	**t-ratio**	**β**	**t-ratio**	**β**	**t-ratio**
Constant	34.62	1.70	90.46	23.18***	63.39	8.13***
Age	0.23	1.14	0.09	1.28	0.14	1.01
Gender	7.34	2.42**	-0.21	-0.07	13.07	3.24**
Education	-2.96	-0.40	3.53	2.18**	-0.01	-0.44
Household size	-0.98	-3.91***	-0.23	-0.90	-1.00	-1.64
Non-farm income	4.86	0.48	-9.15	-2.80**	-5.06	-0.76
Output	0.16	2.15**	0.03	1.05	0.18	2.60**
Distance	-0.49	-2.42**	-0.52	-1.87	-1.44	-4.85***
Market information	-8.83	-1.21	1.77	1.97**	-7.32	-1.98**
Unit price	0.04	1.96**	0.13	1.99**	1.14	4.48***
	Log likelihood ratio = -152.03 Pseudo R^2=19.22		Log likelihood ratio = -115.92 Pseudo-R^2=21.02		Log likelihood ratio = -312.87 Pseudo-R^2=45.80	

***p<0.01, **p <0.05

Source: Survey data (2007)

It was noted that farmers incur huge losses due to perishability of vegetables and transportation costs associated with long distance from farms to markets. Furthermore, the household size also contributes to a significant reduction in the percentage of vegetable sold by rural farmers. Larger households imply higher consumption needs and low labor supply for production (if a greater proportion of the household consists of children). Although self-sufficiency may be improved from consumption of own-farm produce, it could in some cases constrain ability to sell more (due to limited marketable surplus) even when lucrative markets are available.

For peri-urban farmers, the intensity of market participation is significantly increased by the household head's education level and access to formal market information channels. Nonfarm income on the other hand, significantly reduces amount of vegetable sold. Informal market information sources contribute to significant increments in percentage of vegetable marketed by the whole sample.

The benefit of the Chow test in this study is that it has enabled the disaggregated analysis to capture the effects of three important factors (non-farm income, household size and education), which would otherwise have been ignored if only a pooled sample model were used. These findings show the need for certain site-specific strategies, as well as nation-wide interventions to facilitate increased production and marketing of vegetables.

Conclusions and Policy Implications

It is important to support continued transformation of agriculture from subsistence to a more commercially-oriented sector. In the remote rural areas, development priority might focus on enhancing retail markets (build more outlets and equip them with requisite facilities) closer to farms in order to reduce transaction costs associated with transportation. Deliberate efforts should be made to engage majority of unemployed rural labour in long term infrastructure projects such as construction and maintenance of roads, bridges and rural industries. This would help to lower over-dependence on farm output, and thereby contribute to increased marketable agricultural surplus.

There is need to sustain high levels of food production and trade in the peri-urban areas due to intense rural-urban migration. Efficient production skills are necessary, considering scarcity of land and other inputs, as enterprise competitiveness changes. In addition, innovative approaches are desired in the provision of specific market information. For example, mobile phones and internet communication channels could be utilised to promote linkages between vegetable suppliers and consumers in residential areas, or supermarket attendants and farmers (e.g., instant stock updates and supply requests to the farmer whenever a unit of vegetable is purchased from the shelves).

Policy inferences from future research might be made more applicable to pertinent challenges bedevilling agriculture-dependent livelihoods by refining analytical approaches to be cognizant of statistical variations or similarities in parameters across different sites.

Acknowledgements

The authors wish to express special appreciation to the Food and Agriculture Organization of the United Nations, and the Netherlands Government Partnership Project (FNPP) for funding this study. Comments from various colleagues at the Kenya Institute for Public Policy Research and Analysis (KIPPRA) on earlier drafts are acknowledged. Any errors and omissions remain ours.

References

Alene, A. D., Manyong, V. M., Omanya, G., Mignouma, H. D., Bokanga, M. and Odhiambo, G., 2008. Smallholder market participation under transactions costs: Maize supply and fertilizer demand in Kenya. *Food Policy,* 33: 318-328.

Campos, J., Ericsson, N. R. and Hendry, D. F., 1996. Cointegration tests in the presence of structural breaks. *Journal of Econometrics*, 70:187-220.

CBS, 2003. Geographic Dimensions of Well-being in Kenya: Who and Where are the Poor? From Districts to Locations, Vol. I. Central Bureau of Statistics (CBS), Ministry of Planning and National Development. Regal Press Kenya Ltd. Nairobi.

CBS, 2006. Economic Survey. Central Bureau of Statistics (CBS), Ministry of Planning and National Development, Government Printer, Nairobi.

Chow, G. C., 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica,* Vol. 28, 3 (July 1960).

Diebold, F. X. and Chen, C., 1996. Testing structural stability with endogenous breakpoint – A size comparison of analytic and bootstrap procedures. *Journal of Econometrics*, 70:221-241.

Food Ethics Council, 2008. The food crisis: Scarcity or injustice? The magazine of the Food Ethics Council, Summer 2008, Vol. 3, Issue 2. www.foodethicscouncil.org. accessed online, November 01.

Geda, A., de Jong, N., Mwabu, G., and Kimenyi, M., 2001. Determinants of poverty in Kenya: a household level analysis. Institute of Social Studies (ISS), Netherlands. www.eldis.org/static/DOC9379

Ghilagaber, G., 2004. Another look at Chow's test for the equality of two heteroscedastic regression models. *Quality and Quantity, 38:81-93*, Kluwer Academic Publishers, The Netherlands.

Greene, W. H., 2003. Econometric Analysis. Macmillan, New York.

IFAD, 2001. *The Challenge of Ending Rural Poverty*. Rural Poverty Report. Oxford University Press, Oxford.

Kozumi, H. and Hasegawa, H., 2000. A Bayesian analysis of structural chances with an application to the displacement effect. *The Manchester School of Economic and Social Studies*, 68:476-490.

McCulloch, N. and Ota, M., 2002. Export horticulture and poverty in Kenya. Working Paper No. 174. Institute of Development Studies, University of Sussex, Brighton.

McCullough, E. B., Pingali, P. and Stamoulis, K. G., 2008. The Transformation of Agri-Food Systems: Globalization, Supply Chains and Smallholder Farmers. *The Food and Agriculture Organization (FAO) of the United Nations and Earthscan Press, London*, pp 381.

Minot, N. and Ngigi, M., 2003. Are horticultural exports a replicable success story? Evidence from Kenya and Côte d'Ivoire. Paper presented at the InWEnt, IFPRI, NEPAD, CTA Conference on "Successes in African Agriculture". December 1-3, Pretoria.

Moore, D. S., 2006. The Basic Practice of Statistics, 4th edition, 728pp. W H Freeman and Co publishers, ISBN-13:978 071677 4631.

Omiti, J., McCullough, E., Otieno, D., Madelon, M., Nyanamba, T. and Murage, A., 2006. Participatory Prioritization of issues in smallholder agricultural commercialization in Kenya. Discussion Paper No. 64/2006, Kenya Institute for Public Policy Research and Analysis (KIPPRA), Nairobi.

Omiti, J., Omolo, J. O. and Manyengo, J. U., 2004. Policy Constraints in Vegetable Marketing in Kenya. Discussion Paper No. 061/2004, Institute of Policy Analysis and Research (IPAR), Nairobi.

Republic of Kenya, 2001a. Kiambu District Development Plan 2002-2008. Ministry of Finance and Planning, Government Printer, Nairobi.

Republic of Kenya, 2001b. Kisii Central District Development Plan 2002-2008. Ministry of Finance and Planning, Government Printer, Nairobi.

Wilcox, R. R., 1997. Comparing the slopes of two independent regression lines when there is complete heteroscedasticity. *British Journal of Mathematical and Statistical Psychology*, 50:309-317.

Makhura, M.-N., Kirsten, J. and Delgado, C., 2002. Transaction Costs and Smallholder Participation in the Maize market in the Northern Province of South Africa, pp.463-467 in: *Proceedings of the 7th Eastern and Southern Africa Regional Maize Conference and Symposium on Low-Nitrogen and Drought Tolerance in Maiz*e. Workshop theme "Integrated Approaches to Higher Maize Productivity in the New Millennium", 11 – 15th February 2001, Nairobi.