Journal of Rural Development 28 (Summer 2005): 49～63

CHANGES IN FARM SIZE DISTRIBUTION IN KOREA: EVIDENCE FROM FARM-LEVEL PANEL DATA 1998-2002

KANG, HYE-JUNG*

Key words: Farm consolidation, farm size distribution, nonparametric kernel density function

ABSTRACT

This paper investigates how the farm size distribution has changed recently in the Korean agriculture with an aim to document the degree and form of farm consolidation. For the purpose, this study measures and compares Korea farm size distributions, using farm-level panel data collected from 1998 to 2002. To measure accurate and smooth farm size distributions, nonparametric kernel densities, which ensure robustness of the results against possible misspecification, are estimated. The representative sample shows that significant farm consolidation occurred between 1998 and 2002.

Ⅰ. Introduction

In Korean agriculture with many small-sized family farms, the expansion of individual farm land holdings via the exit of marginal farmers--a process generally labeled "farm consolidation" --has been known as the best way to improve productivity and competitiveness (e.g., Kim 1997; Lee 1998).

It is in this context that farm consolidation is receiving serious attention as a policy objective. The Korean government

* Research Associate, Korea Rural Economic Institute.

has sought to expand land holdings per farm through the elimination of marginal farms, and ultimately improve agricultural productivity and competitiveness. Attempting to enlarge individual land holdings of rice-specialized farms, the Korea Agricultural and Rural Infrastructure Corporation (KARICO), a government agency, invested 2.4 trillion won over the period from 1988 through 2002 under "the Farm Size Growth Project." During this period, the number of participating farms bigger than three hectares tripled. Also, the average paddy area per participating farm increased from approximately 2.04 hectares to 3.73 hectares. The KARICO targeted younger, more efficient farmers, and 90 percent of those who received funds were under age 60 (KARICO 2003). The project was executed through several operations, especially the purchasing, leasing, and combining of farmland. Government policies aiming to achieve farm consolidation have also promoted expansion through various measures, such as direct payments to old farmers, who are willing to relinquish farm management.

Although the wide-spreading efforts were made to promote farm consolidation, there has been little empirical experience supporting the claim that farm consolidation actually occurred in the Korean agriculture. This paper is thus designed to have a better understanding of farm size changes in recent years and to look for evidence of farm consolidation. This study made use of the most comprehensive Korean farm level data available for the period from 1998 to 2002. Specifically, the paper attempts to estimate a nonparametric density function of farm size for accurate measurement of farm size distribution. Analysts of farm size distributions often define a small number of size classes and use summary statistics tables and histograms to examine the size distribution. However, the arbitrary choice of number and size of the classes may affect the observed distributional characteristics.

In contrast to these methods, nonparametric kernel densities directly show the size without the need to draw arbitrary size class distributions. Furthermore, two tests on the estimated densities were conducted to support significant farm consolidation patterns.

This paper attempts to provide some careful and systematic evidence on how farm size distribution has changed in Korea in recent years. Given this evidence, those who do have particular policy agendas may need to fine a new rationale for policies related to farm size and structure. This paper is not designed to analyze farm size policy or to measure the policy or other causes underlying farm size distributions. Rather, the purpose of this paper is to find the best available evidence on farm consolidation through a careful examination of change in farm size distribution.

There is some literature on examination of change in farm size distribution in Korea and other countries. For example, through investigation of summary statistics tables, Lee (1998) noted that small farms less 0.5 hectares and large farms above 2.0 hectares have increased, while middle-sized farms in the range from 0.5ha to 2.0ha have decreased since the 1990s in Korea. Likewise, the concentration on both extremities in the distribution of farmland operated has occurred in Japan and the Netherlands since the 1970s and 1980s, respectively. Kwon and Kim (2001) examined changes in cultivated acreage and owned acreage of Korean rice farms from 1993 to 1998, using nonparametric density function techniques. Their results reveal that the relative frequencies of large farms increased over times in terms of cultivated acreage and owned acreage. Wolf and Sumner (2001) analyzed 1989 and 1993 U.S. dairy farm size distributions, using nonparametric density estimation techniques and found that U.S dairy farm size distributions are not bimodal.

II. Data Description

This study relies primarily upon the farm-level data, compiled by the Korean Ministry of Agriculture and Forestry (MOAF) through a national farm survey for the period from 1998 through 2002. The MOAF survey classifies and reports statistics for approximately 2,900 randomly selected farm households, spanning nine provinces. A distinctive feature of this study is that it relies upon farm level data, as opposed to more common aggregate regional or country level data. Since the regional or country level data have the disadvantage of broad, large-scale observations, they may contain aggregation bias.

It is difficult to define farm size when the data set consists of a heterogeneous set of multiple-enterprise farms. Farm size can be measured in a variety of ways: the amount of farmland operated, the quantity of output, value of sales, or added values on the farm (Sumner and Wolf 2002). Since this study focuses on farm consolidation, land operated per farm was used as the measure of farm size. Farmland operated (cultivated land) is defined as the sum of both owned and rented land in paddy, upland, and orchard areas. In this study, the farmland operated refers to land used for crop production and does not include land for livestock, such as pasture land. Paddy refers to land primarily used for flood-irrigated rice. Upland refers to land used for vegetables, grains, and specialty crops cultivation.

Table 1 provides summary statistics for land operated in 1998 and 2002. Average land operated per farm increased by 4

TABLE 1. Summary Statistics for Land Operated, 1998 and 2002

		Land operated (hectare)		
		Cultivated land	Land owned	Land rented
1998	Mean (per farm)	1.47	0.83	0.64
	Std. Dev. (per farm)	1.22	0.73	0.97
	Sample total	3,803	2,147	1,656
	Sample size	2,585		
	Korea total	1,910,081		
2002	Mean (per farm)	1.53 (4%)	0.81 (-2%)	0.72 (13%)
	Std. Dev. (per farm)	1.39	0.77	1.10
	Sample total	3,963	2,099	1,864
	Sample size	2,585		
	Korea total	1,876,142		

Note: Values in parentheses denote percentage growth rates from 1998 and 2002 respectively.

percent between two periods. Also note that operation of the land owned decreased slightly (2 percent), while the land rented per farm increased by 13 percent. Although owned land, which is in operation, represents more than half of the total cultivated land, the share of rented land is large. Increasing shares of rented land are also observed in other countries, such as the United States and Japan.

Summary statistics as in Tables 1 provide information regarding average land operated per farm, but do not reveal which farm size groups have decreased or increased. Table 2 therefore presents the growth patterns of farm size by discrete percentiles of the farm size distribution. At each percentile, growth is represented by changes in the average farm size. Choices of both the number and size of the percentiles are arbitrary. The table also presents the percentage growth rates of an average-sized farm in 2002 relative to 1998 at each percentile. For instance, the average-sized farm in the 5th percentile shows land operated decreased by 23 percent. The average-sized farm in the 95th percentile, however, shows an increase in operated land by ten percent. The growth rate for an average-sized farm in the 99th percentile is the highest among all percentiles (19 percent). Conversely, note that the growth rate for an average-sized farm in the 50th percentile is close to zero. In summary, the distribution of sizes shifted away from smaller towards larger farms, while a segment of middle-sized farms remained stagnant over the two periods.

The values for skewness and kurtosis are also presented in Table 2. A larger skewness value indicates more skewness and less symmetry. From 1998 to 2002, the size distribution of farmland operated becomes less skewed. Positive kurtosis indicates a peaked distribution and negative kurtosis indicates a flatter distribution.[1] In this case the kurtosis values are positive, indicating

[1] Kurtosis is a measure of whether the data have a peaked or flat distribution relative to a normal distribution. That is, a distribution with a high kurtosis tends to exhibit a distinct peak near the mean, rapid decline, and heavy tail. A distribution with low kurtosis usually has a flat top near the mean, rather than a sharp peak.

that farm size distribution is relatively more peaked than a normal distribution.

Table 3 presents farm size growth patterns by discrete percentile ranges, based on the 1998 farm size distribution. Average farm size above the 25^{th} percentile increased from 1998 to 2002, especially that of above 75th percentile (6.2 percent). Average farm size below the 25^{th} percentile decreases by 7.7 percent. These findings confirm the results in Table 2.

TABLE 2. 1998 and 2002 Growth Patterns of Land Operated Per Farm

Year	Skewness	Kurtosis	Hectare per farm								
			Percentiles of 1998 farm size distribution								
			1^{st}	5^{th}	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}	95^{th}	99^{th}
1998	2.53	15.22	0.10	0.22	0.35	0.67	1.19	1.91	2.89	3.73	5.67
			Percentiles of 2002 farm size distribution								
			1^{st}	5^{th}	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}	95^{th}	99^{th}
2002	2.50	13.59	0.08	0.17	0.28	0.62	1.18	2.02	3.09	4.09	6.73
			Growth rate from 1998 to 2002 (%)								
			-20.0	-22.7	-20.0	-7.5	-0.8	5.8	6.9	9.7	18.7

Note: Based on discrete percentiles of the farm size distributions for 1998 and 2002.

TABLE 3. Growth Patterns of Land Operated Per Farm Based in 1998

	Percentiles of 1998 farm size distribution			
	Below 25^{th}	$25\text{-}50^{th}$	$50\text{-}75^{th}$	$75\text{-}100^{th}$
Average farm size in 1998 (ha)	0.393 (0.177)	0.910 (0.147)	1.509 (0.206)	3.072 (1.339)
Average farm size in 2002 (ha)	0.363 (0.181)	0.920 (0.147)	1.515 (0.200)	3.261 (1.543)
Average growth rate(1998-2002)	-7.71%	1.14%	0.42%	6.17%

Note: Based on discrete percentiles of the farm size distribution for 1998. Values in parentheses refer to standard deviations.

The farm size distribution, observed through summary statistics table and discrete percentiles, such as Table 1, 2, and 3, does not show smooth connection among distributional characteristics.

III. Empirical Implementation

1. Measuring Farm Size Distributions

There are many ways to measure farm size distributions including histograms, parametric, and nonparametric densities.[2] Histograms are the most common method of finding out a distribution.

However, histograms still require a strong assumption of arbitrary size class distinctions. In order to overcome this restriction and examine more accurate measure, this paper estimates nonparametric kernel densities of farm size to describe farm size distributions, examining without prejudice against the number of size classes or the size class boundaries. To help understand the nonparametric kernel density estimation techniques to be discussed later, the examination of histograms for farm size distributions will be introduced first.

1.1. Histogram Approach

Histograms are typically expressed as a bar graph where the height of the bar represents the share of observations contained in that window (or width). Histograms assume the probability density, which represents the fraction of observations falling in a given window, is constant within each window. The histogram density estimator is defined as below: (Pagan and Ullah 1999)

$$(1) \qquad \hat{f}_H(x) = (n\lambda)^{-1}\sum_{i=1}^{n} I(x - \frac{\lambda}{2} \le x_i \le x + \frac{\lambda}{2}) = (n\lambda)^{-1}\sum_{i=1}^{n} I(-\frac{1}{2} \le \varphi_i \le \frac{1}{2}) \ ,$$

---

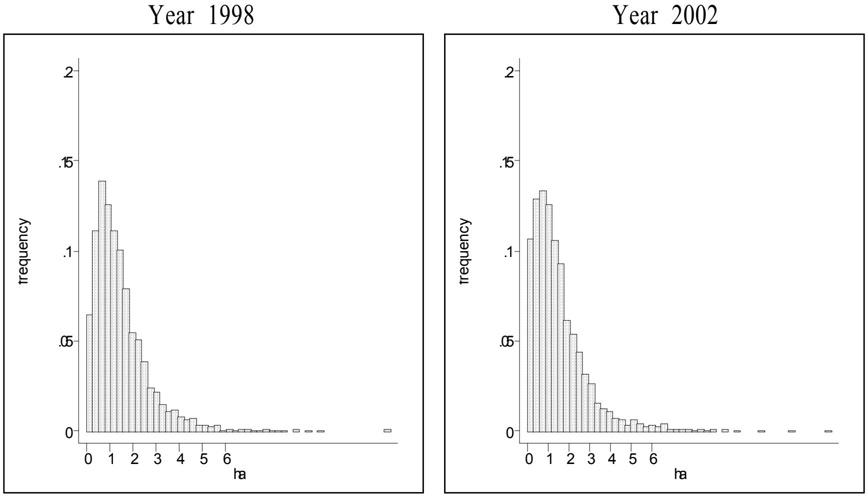
[2] Parametric distributions that have been used to represent firm size include the lognormal, Yule exponential, gamma, and beta distributions (Wolf and Sumner 2001).

$I(A) = 1$, if A is true, zero otherwise,

where $\hat{f}_H(x)$ is the density estimator of, $f_H(x)$, λ is the width of the interval, and $\varphi_i = (x_i - x)/\lambda$. The width of the interval λ controls the amount by which the data is smoothed to produce the estimate (1). However, although the histogram is directly estimated without assuming its form, it has the drawbacks that the density estimate by histogram is a discrete function and has jumps at the points $x_i \pm h/2$ with zero derivative elsewhere. Due to these disadvantages of the histogram approach, the nonparametric kernel estimator should be estimated to produce smooth estimates of $f(x)$.

Figure 1 displays the 1998 and 2002 farm size histograms with fifty windows. The histograms show that the proportions of farms below 0.5 hectares and above 2.0 hectares increased, while the proportion of farms in the range of 0.5 hectares to 2.0 hectares decreased slightly over the two periods. These findings are consistent with those of Tables, but in this case, the window and arbitrary choice of origin determine the shape of the histogram density estimate. Since this estimate does not have a smooth shape, it also cannot be used to examine the modality of distribution.

FIGURE 1. 1998 and 2002 Farm Size Histograms with Fifty Windows



1.2. Nonparametric Kernel Density Function

The distributions obtained by nonparametric kernel densities do not require arbitrary size class distinctions, as histograms do. Also, the estimators "let the data speak" and ensure robustness of the results against possible misspecification. Flexibility in the estimates facilitates identification of salient distribution features.

The general form of a kernel density estimator for univariate observations, $x = x_1, x_2, \ldots, x_n$, is defined as,

$$(2) \qquad \hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) = \frac{1}{nh}\sum_{i=1}^{n} K(\varphi_i), \qquad \int_{-\infty}^{\infty} K(\varphi)d\varphi = 1 ,$$

where K is a real and positive kernel function with window-width (smoothing parameter or band width) h. When using the kernel estimator described in (2), choices need to be made for the kernel function K and the window width h. The choice of kernel function is usually considered to be a minor issue, with any kernel being close to optimal for large samples. In contrast, the selection of the window width h is crucial.

This study employs the Epanechnikov kernel function and follows the Silverman criterion (1986) to obtain the optimal window-width for kernel density estimates.[3] Given the optimal window width h, the Epanechnikov kernel generally minimizes the MISE (Mean Integrated Square Error). The MISE measures the discrepancy between the density estimator $\hat{f}(x)$ and the true density $f(x)$ as,

$$MISE(h) = \int_x \{E(\hat{f}(x)) - f(x)\}^2 dx + \int_x Var(\hat{f}(x))dx ,$$

which is the sum of the integrated square bias and the integrated variance.

---

3 See Pagan and Ullah (1999, 9-23).

The Epanechnikov kernel function is defined as:

(3) $$K(\varphi)=\frac{3}{4\sqrt{5}}(1-\frac{\varphi^2}{5}),\ \text{if}\ |\varphi|\le\sqrt{5}\ ,\quad K(\varphi)=0\ \text{otherwise.}$$

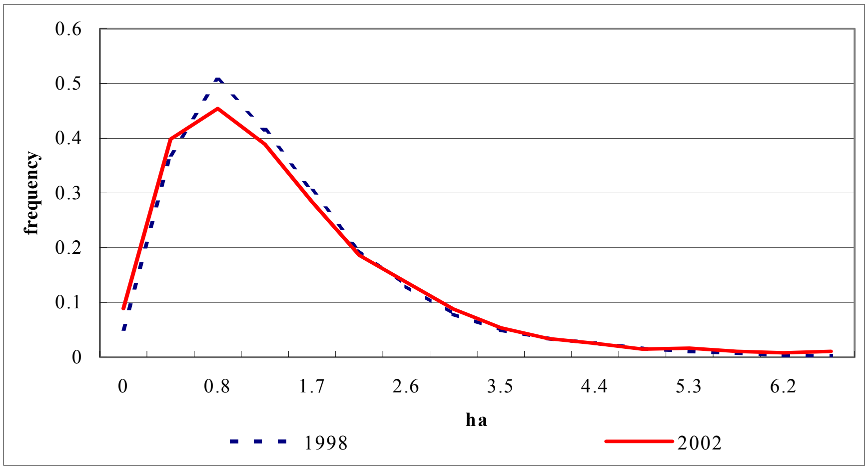
With a specific kernel function, the value of h determines the degree of averaging the estimate of the density function. Window-width h is thus important in determining the shape of the density. A very large h is expected to result in an over-smoothed density, possibly leading to bias in the kernel density estimate. An especially small h may give a noisy and wiggly density estimate, with increased variance but less bias. Silverman (1986) suggests using the window-width, minimizing the MISE. That is:

(4) $$h = 0.9An^{-1/5}\ ,$$

where A=min (standard deviation of the x variable, interquartile range of the x variable /1.34).

Using the kernel function from equation (3) and Silverman's optimal window-widths from equation (4), Figures 2 illustrates both 1998 and 2002 kernel densities for farmland operated.

FIGURE 2. Kernel Densities for Farmland Operated



Note: Horizontal ranges in all figures are truncated for more clear shapes of farm size distribution.

The density is skewed towards the left, indicating the presence of a large proportion of small farms and a smaller proportion of large farms. The density appears unimodal, with a maximum point. From 1998 to 2002, the proportions of smaller farms below 0.5 hectares and larger farms above 2.0 hectares increased slightly, while the proportion of those between 0.5 hectares to 2.0 hectares decreased. These changes in farm size distribution over the two periods are consistent with the findings of the summary statistics tables and histograms (Figure 1) discussed previously.

2. Tests for Modality and Comparisons between Densities

2.1. Test for modality of the farm size densities

As previously illustrated by the histograms and the nonparametric kernel densities, farm size distribution seems to appear an increasing portion of small farms and large farms and a slightly declining number of farms with moderate size from 0.5 to 2.0 hectares in 1998 and 2002. To help determine the true shapes of farms size distributions, the number of modes in a single nonparametric density estimate should be statistically tested. The nonparametric kernel density estimator allows the farm size data to determine the number of modes rather than assuming a single mode like histograms. Silverman (1981) developed a method to test the null hypothesis that a density f has k modes against the alternative that f has more than k modes, where k is a non-negative integer. The test statistic in this case is the critical window width, defined by

$$h_{crit}(k) = \inf\{h \mid \hat{f} \text{ has at most } k \text{ modes}\}. \tag{5}$$

For $h < h_{crit}(k)$, the estimated density has at least $k+1$ modes. The value of $h_{crit}(k)$ is computed through a binary search process, and its significance level can be assessed by the smoothed bootstrap procedure (Efron 1979).

Table 4 presents the critical window-widths and significance levels for a test of the null hypothesis that the underlying density has at most one mode against the alternative that it has more than one mode. The *p*-values are computed by simulating 1,000 replications from the critical density. As Silverman (1981) suggests, this study looks only for the window-width, where a second mode appears to test the null hypothesis to find out whether the farm size distributions are bimodal or unimodal. The *p*-values are the smallest size at which the null hypothesis would be rejected and reflect the strength of evidence against the null hypothesis of one mode (Wolf and Sumner 2001; Henderson et al. 2002). The results reject the bimodal hypothesis in favor of the null hypothesis that farm size distributions are unimodal in both 1998 and 2002 at the five percent and ten percent level of significance, respectively.

TABLE 5. Critical Window-widths and Significance Levels for Unimodal Farm Size Distributions

Farm Size Distribution	Window-Width	P-value
Year 1998	0.18	0.05
Year 2002	0.30	0.10

2.2. *Testing for Equality of Farm Size Densities*

The kernel densities illustrated in Figure 2 show that the shapes of farm size distributions have changed slightly from 1998 to 2002. To obtain more rigorous statistical evidence supporting this visual determination, a formal statistical test should be conducted.

For the comparison of two unknown densities of the different time periods, this study tested $H_0 : f_1(x) = f_2(x)$ vs. $H_a : f_1(x) \neq f_2(x)$. $\{x_{1i}\}$ and $\{x_{2i}\}$ were taken as two equally sized samples of size $n(n_1 = n_2 = n)$ from the two density functions f_1 and f_2 respectively. Testing problems can be resolved by considering the measure of global distance between the two densities $f_1(x)$ and $f_2(x)$. The relevant test statistic (Li 1996) is defined as,

$$
\begin{aligned}
I &= \int (\hat{f}_1(x) - \hat{f}_2(x))^2 dx = \int \hat{f}_1^{\,2}(x)dx + \int \hat{f}_2^{\,2}(x)dx - 2\int \hat{f}_1(x)\hat{f}_2(x)dx \\
&= \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j=1(i\neq j)}^{n}[K(\frac{x_{1i}-x_{1j}}{h}) + K(\frac{x_{2i}-x_{2j}}{h}) - K(\frac{x_{2i}-x_{1i}}{h}) - K(\frac{x_{1i}-x_{2i}}{h})] \ , \\
&+ \frac{1}{n^2 h}\sum_{i=1}^{n}[2K(0) - 2K(\frac{x_{1i}-x_{2i}}{h})] \\
&= I_1 + I_2
\end{aligned} \tag{6}
$$

where $\hat{f}_1(x)$ and $\hat{f}_2(x)$ are the kernel density estimators of $f_1(x)$ and $f_2(x)$.

As Li points out, when a central limit theorem is applied to this statistic, under the null hypothesis, the statistic has an asymptotically standard normal distribution as follows:

$$
I = \frac{nh^{1/2} I_1}{\hat{\sigma}} \xrightarrow{d} N(0,1) \ , \tag{7}
$$

where $\hat{\sigma}^2 = \frac{2}{n^2 h}\sum_{i=1}^{n}\sum_{i=1}^{n}[K(\frac{x_{1i}-x_{1j}}{h}) + K(\frac{x_{2i}-x_{2j}}{h}) + 2K(\frac{x_{1i}-x_{2i}}{h})]\int K^2(\varphi)d\varphi$.

Table 6 presents the null hypothesis testing results that two unknown kernel densities are identical. For size densities for 1998 and 2002, the null hypothesis is rejected at the ten percent significance level, implying that the two density estimates are statistically different for the two periods.

TABLE 6. The Li Test Statistics for Differences in Farm Size Distributions

Year comparison	T-test
1998 vs. 2002	1.70[a]

Note: The null hypotheses are that two density functions are identical.
The 90 percent critical value for the standard normal distribution is 1.645.
[a] H_0 is rejected at the 10 percent level.

IV. Summary and Conclusions

This paper investigated how farm size distribution (measured by area farmed) recently changed in Korea with an aim to document the degree and form of farm consolidation. For the purpose, this study measured and compared farm size distributions in Korea, using farm-level panel data collected from 1998 to 2002. For measuring accurate and smooth farm size distributions, nonparametric kernel densities, which ensure robustness of the results against possible misspecification, were estimated.

As a conclusion, it was found from the representative sample that significant farm consolidation occurred between 1998 and 2002. The proportions of smaller farms less 0.5 hectares and larger farms above 2.0 hectares increased, while the proportion of those between 0.5 to 2.0 hectares decreased. The study finding suggests that some farms (0.5 to 2.0 hectares) leased out or sold their land, while their acreage remained in agriculture, thus making some farms leasing or purchasing the acreage for greater land holdings. Farms which leased out their land, have smaller landholdings than before. The test result of the equality of two unknown kernel densities confirmed that the two density estimates are statistically different between the two periods.

Even though the results show a slight change in farm size distribution over a short sample period investigated, this study is meaningful in that it describes the shape of Korean farm size distributions accurately using the Kernel densities, in addition to summary statistics tables and histograms. This study provides a guidance for understanding recent changes in farm size distribution in the Korean agriculture, which is an evidence of the effort to expand individual farm landholdings in Korean agriculture. Without a careful examination of the best available evidence on farm consolidation, there will be no progress in the studies about farm consolidation. This study also helps evaluate policies performed for farm consolidation in Korea. A better sense of the true size distribution will lead to developing better direction and strategy of the Korean agricultural transformation. Thus, although there is little policy interest in this issue, accurate measuring of farm size distribution in Korea is required.

In addition, changes in farm size distribution over a long period and comparisons of changes in farm size distribution across country are important and interesting topics. Building on this research, further studies for these topics will be conducted, where the factors affecting farm consolidation will be identified, and the impact of farm consolidation on productivity will be examined.

REFERENCES

Efron, B. 1979. "Bootstrap methods-Another look at the Jack-Knife," *Annals of Statistics* 7: 1-26.

Kim, H.S. 1997. "Evaluation and Policy Tasks of Large-Scale Farming." *Korean Rural Economics* 20.

Kwon, O.S. and M.H. Kim. 2001. "Acreage Distribution in Korean Rice Farming: A Kernel Density Analysis." *Korean Journal of Agricultural Management and Policy* 28: 703-716.

Lee, J.H. 1998. *Transformation of Korean Farm Structure: From Beginning to End.* Research series 21. Korea Rural Economic Institute.

Li, Qi. 1996. "Nonparametric Testing of Closeness between Two Unknown Distribution Functions." *Econometric Reviews* 15(3): 261-274.

Pagan, A. and A. Ullah. 1999. *Nonparametric Econometrics.* Cambridge University Press.

Silverman. 1981. "Using Kernel Density Estimates to Investigate Multimodality," *Journal of the Royal Statistical Society.* Series B 43(1): 97-99.

Silverman. 1986. *Density estimation for statistics and data analysis.* Chapman and Hall: New York.

Sumner, D.A. and C.A. Wolf. 2002. "Diversification, Vertical integration, and the Regional Pattern of Dry Farm Size," *Review of Agricultural Economics* 24: 442-457.

Wolf, C.A. and D.A. Sumner. 2001. "Are Farm Size Distributions Bimodal?: Evidence from Kernel Density Estimates of Dairy Farm Size Distributions." *American Journal of AgriculturalEconomics* 83(1): 77-88.