On the extent of the market: a Monte Carlo study and an application to the United States egg market

Aklesso Egbendewe-Mondzozo

Department of Agricultural Economics
Texas A&M University
2124 TAMU
College Station, TX 77843-2124
Email: aklesso@tamu.edu

Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, Atlanta, Georgia, January 31-February 3, 2009

Summary

This paper investigates the extent of the market, using a switching regimes model similar to those used in stochastic frontiers estimations. We started by performing a Monte Carlo simulation on our model, seeking to evaluate its performance in terms of correctly estimating the probability of integration of two markets. Our Monte Carlo results under the assumption of half-normal and exponential distribution of the errors, revealed that these two distributions predict almost correctly the probability of integration of two markets. The half-normal error distribution model tends to slightly underestimate the true probability of integration, while the exponential error distribution model tends to slightly overestimate the true probability of integration. We, finally, applied the model to the United States egg market using data from three highly productive states and one less productive state. The model predicts that, the markets pairs considered are integrated. That is, the four markets studied belong to the same economic market in the sense of Marshall. Further, based on our Monte Carlo study, we find that the true probability of integration of two given markets lies in between the half-normal model estimates and the exponential distribution model estimates.

1. Introduction

Since the seminal work of Stigler and Sherwin (1985) on the extent of the market, a body of literature has been developed to analyze the market mostly in terms of its geographical definition. Most of the works in the area have been motivated by the definition of “antitrust markets” compared to the “economic markets”. As defined by Alfred Marshall (1961), “a market for a good is the area within which the price of a good tends to uniformity, allowance being made for transportation.” This definition is related to the economic market where differences in prices of the same commodity observed at different places are due to transaction costs. Therefore, according to the definition of a market, in the same geographic region it is almost impossible that prices of the same commodity display a greater difference than the transaction costs over a long period of time. The reason is that, arbitrage will always occur when the price differences in different geographic regions are more than the transaction costs.

From this definition, we can readily infer that the antitrust markets are just the opposite of the economic markets in the sense that greater difference in prices (where price differences cannot be explained by transaction costs alone) will be observed as non transitory between two geographical markets. Hence, the place that maintains the higher price is a potential antitrust market. We can see from these two market definitions, that an antitrust market cannot be incorporated in the economic market. This point of view was defended by Spiller and Huang (1986). The initial studies in this field, particularly in Stigler and Sherwin (1985), the analysis were oriented towards cointegration of prices. That is, geographical locations that happen to be in the same market will display a parallel trend of their prices once care has been taken to remove the common shocks. Technically, prices collected from two different locations are said to be cointegrated if they are both integrated of the same order, say I(d), and that their linear combination generates an innovation term that is integrated of order I(d-1). The test of cointegration can usually be obtained by the Augmented Dickey-Fuller (ADF) test. This test is popular but can sometimes fail to do a good job when we have a near unit root situation. That is, a near unit root situation makes the test less powerful for this kind of analysis. Furthermore, the cointegration method falls short when it comes to calculating the average transaction costs. Average transaction costs are used to measure differences in prices of two market locations.

To overcome these problems, Spiller and Huang (*op. cit.*) have developed a switching regimes regression model that performs well in terms of distinguishing economic markets from potential antitrust markets. In this paper, we will use a similar model to analyze the U.S market for eggs. Our approach will be a bit different from that used by the above authors in that we will adapt the model to possibly utilize the entire price data observations without having to split them in two. Also, we propose different error distributions to evaluate the robustness of our method. The remainder of this paper is organized as follows: In section 2, we will introduce our theoretical model, in section 3 we will carry out a Monte Carlo study on the model involving Half-normal and Exponential distributions. In section 4, we will briefly present the United States egg market, in the section 5, we will apply our model to that market and finally in section 6 we will conclude.

2. The model

As we mentioned above, the model is in the spirit of the Spiller and Huang (1986). Suppose we are looking at two geographical locations A and B with observed prices p_t^A and p_t^B respectively. These prices move around their means with eventually some random shocks. For simple expositional purpose, suppose that we have the situation $p_t^A > p_t^B$. Let T_t be the transaction costs of moving the commodity from the location B to the location A. We have the following situation:

If $p_t^A - p_t^B < T_t$ [1]

it will be impossible to have arbitrage between the two locations because the seller in location B will have to pay more than the transaction costs to sell at the location A and the assumption of rationality of the producer is violated .

Alternatively, if we have:

$$p_t^A - p_t^B > T_t \quad [2]$$

a possible arbitrage can occur between the two locations and this will lead to an equalization of prices with the transaction costs wedge. In other words, if both locations belong to the same economic markets, we will likely see that only transaction costs will separate the prices in those markets.

That is we have:

$$p_t^A - p_t^B = T_t \quad [3]$$

Now let suppose that the transaction costs are distributed geometrically with mean k such that we have:

$$T_t = k\exp(\nu_t) \quad [4]$$

where ν_t is a random variable normally distributed with mean zero and variance σ_ν^2.

From equation [1], we can define the probability of having the price differences less than the transaction costs as:

$$\Pr ob[p_t^A - p_t^B < T_t] = \Pr ob[p_t^A - p_t^B < k\exp(\nu_t)] = \beta \quad [5]$$

To depart from the Spiller and Huang model, equation [5] can be rewritten by squaring both terms in the bracket before we take the logarithm in order to relax the assumption of $0 < p_t^A - p_t^B$ made in their model. The point we want to make is that the sign of price differences can change arbitrarily to include the case where $p_t^B - p_t^A < T_t$ in equation [5]. The reason is that, the scalar β is the probability that the price difference is less than the transaction costs. This can also be looked at as the number of times where we have $0 < p_t^A - p_t^B$ and $p_t^A - p_t^B < T_t$ or $0 < p_t^B - p_t^A$ and $p_t^B - p_t^A < T_t$. Hence, the parameter β is the relative frequency with which we cannot move commodities between markets because of our price differences gain being less than the transaction costs[1]. That is, a high β implies a less integrated market. The advantage of squaring the terms in equation [5] is that in our estimations we will not need to split our sample in two in order to conform to the assumption of $0 < p_t^A - p_t^B$ as Spiller and Huang did. Furthermore, the new formulation implies that the transaction costs between two geographic markets must be unique whether we move commodities from location A to B or from location B to A. We, therefore, have the following expression:

$$\Pr ob[\log(p_t^A - p_t^B)^2 - 2\log k < 2v_t] = \Pr ob[\frac{1}{2}\log(p_t^A - p_t^B)^2 - \log k < v_t] = \beta \qquad [6]$$

We are now ready to set up our switching regimes model. First, we define a positive random variable $\eta_t > 0$ so that we have the following expressions:

$$\frac{1}{2}\log(p_t^A - p_t^B)^2 = D + \nu_t - \eta_t \quad \text{with probability } \beta \qquad [7]$$

$$\frac{1}{2}\log(p_t^A - p_t^B)^2 = D + \nu_t \quad \text{with probability } 1-\beta \qquad [8]$$

where $D = \log k$.

The equilibrium in equation [7] corresponds to the no arbitrage opportunity where the difference between the two prices is usually less than the transaction costs. The equilibrium in equation [8] corresponds to the arbitrage state where the equality between the two prices prevails most of the time. With the assumption that η_t is half-normally distributed and truncated above zero with variance σ_η^2, we can express the likelihood function by the following equation:

$$L = \prod_{t=1}^{n} [\beta f_t^1 + (1-\beta) f_t^2] \qquad [9]$$

f_t^1 and f_t^2 are the density functions of $\varepsilon_t = \nu_t - \eta_t$ in equation [7] and ν_t in equation [8] respectively. By letting $Y_t = \frac{1}{2}\log(p_t^A - p_t^B)^2$ we have the following equations (see Aigner et al., 1977):

---

[1] Basically, what is required in our model is that the absolute value of the price differences should be less than the transaction costs so that arbitrage does not occur in both directions. This will mean that the markets are isolated from each other and are similar to the autarky markets.

$$f_t^1 = \left(\frac{2}{\sqrt{\sigma_\nu^2 + \sigma_\eta^2}}\right)\phi\left(\frac{Y_t - D}{\sqrt{\sigma_\nu^2 + \sigma_\eta^2}}\right)\left[1 - \Phi\left(\frac{(Y_t - D)\sigma_\eta}{\sigma_\nu\sqrt{\sigma_\nu^2 + \sigma_\eta^2}}\right)\right] \qquad [10]$$

$$f_t^2 = \frac{1}{\sigma_\nu}\phi\left(\frac{Y_t - D_t}{\sigma_\nu}\right) \qquad [11]$$

Under an alternative assumption that η_t follows exponential distribution, we will have an exponential density function $g(\eta_t) = \frac{1}{\theta} * \exp(\frac{\eta_t}{\theta})$. The mean of the exponential distribution is θ and the variance is θ^2. With a little algebra (see Greene, 2003) one can show that we will have the following expression:

$$f_t^1 = \frac{1}{\theta}\left[1 - \Phi\left(\frac{Y_t - D}{\sigma_\nu} + \frac{\sigma_\nu}{\theta}\right)\right]\exp\left[\frac{(Y_t - D)}{\theta} + \frac{1}{2}\frac{\sigma_\nu^2}{\theta^2}\right] \qquad [12]$$

f_t^2 being the same as in equation[11] . $\Phi(\bullet)$ is a standard normal *CDF* and $\phi(\bullet)$ is a standard normal *PDF* .

The maximum likelihood estimation of the logarithm of [9] will lead to the estimates of the parameters of interest $(\sigma_\nu^2, \sigma_\eta^2, D, \beta)$ for the half-normal specification and the parameters $(\theta^2, \sigma_\eta^2, D, \beta)$ for the exponential distribution. Therefore, the probability of integration of any two markets is equivalent to $1-\beta$. That is a lower β corresponds to higher integration between two markets and a higher β corresponds to lower integration.

3. Monte Carlo experiments

3.1 Data generation process

To test our model, we set up an experiment where we have generated price differences data consistently with equations [7] and [8]. The average transaction costs is taken to be $k = 0.04$ such that $\log k = D = -3.22$. We generate one thousand observations of the error terms v_t which are normally distributed with mean zero and variance $\sigma_\nu^2 = 0.20$ so that we have $v_t \sim N(0, 0.20)$. We also generate one thousand observations of the error terms η_t.

In the case where η_t follows half-normal distribution, the variance is taken to be $\sigma_\eta^2 = 1.16$ so that we have $\eta_t \sim HN(0, 1.16)$ and in the case where η_t follows an exponential distribution, the variance is taken to be $\theta^2 = 0.36$ so that we have $\eta_t \sim Exp(0, 0.36)$. The relative frequency of no arbitrage (the probability of no integration) between markets is taken to be $\beta = 0.3$; which means that the true probability of integration in the experimental market pairs is $1-\beta = 0.7$. Once we generated the random samples for the composite error terms v_t and η_t, we calculated the variable Y_t in such a way that it represents the expected value of the equations [7] and

[8]. That is, we have generated $Y_t = \beta(D + \nu_t - \eta_t) + (1-\beta)(D + \nu_t)$. To analyze the results as a function of the sample size, we generate estimation sample sizes of 200, 250 and 500 observations. The results of the experiments for both half-normal and exponential distribution models are given in the table below.

3.2 Monte Carlo experiments results

Table1: **Simulation results for half-normal and exponential distribution models**

	Half-Normal				**Exponential**			
	β =0.30	σ_ν^2 =0.25	σ_η^2 =1.16	D =-3.22	β =0.30	σ_ν^2 =0.25	θ^2 =0.36	D =-3.22
200								
Mean	0.371	0.256	1.136	-3.389	0.256	0.237	0.360	-3.239
Bias	-0.071	-0.006	0.023	0.169	0.043	0.012	-0.001	0.019
MSE	0.073	0.019	0.023	0.180	0.050	0.014	0.002	0.033
250								
Mean	0.370	0.256	1.136	-3.388	0.231	0.241	0.364	-3.245
Bias	-0.070	-0.006	0.023	0.168	0.069	0.005	-0.004	0.025
MSE	0.071	0.018	0.024	0.179	0.281	0.079	0.035	0.281
500								
Mean	0.370	0.257	1.137	-3.387	0.249	0.238	0.361	3.240
Bias	-0.070	-0.007	0.022	0.060	0.050	0.011	-0.001	0.020
MSE	0.071	0.019	0.023	0.071	0.145	0.033	0.016	0.051

The results from table1 above show that both error specifications (half-Normal and Exponential) estimate closely the relative frequency of no arbitrage β for the two experimental markets. The half-Normal specification overestimates the frequency of no arbitrage by 0.07 and the Exponential specification underestimates the frequency of no arbitrage by 0.04. These biases of the estimates are almost zero in the both specifications. In fact, the probabilities of integration $(1-\beta)$ of the two markets in the experiment are estimated as 0.63 by the half-normal model and as 0.751 by the Exponential model. That is, both models perform well in the estimation of the parameters of the models with minor biases. However, we acknowledge that the true probability of integration is bounded below by the half-normal distribution model estimates and bounded above by the exponential distribution model estimates as the true probability of integration is 0.7. In the next section we will apply our model to the U.S egg market.

4. The U.S egg market

Eggs are usually classified as a poultry product by the United State Department of Agriculture (USDA). Their production is concentrated in the eastern and southern of the United States. The egg firms located in the rest of the country are scattered and smaller. This is probably because of the climatic conditions of the center and the northern areas of the United State (see map in the appendix). From this picture, our conjecture is that the prices will tend to be lower in the southern and eastern parts than the rest of the country, as the south and the east have higher egg production. Furthermore, for example, if the transportation of eggs cannot be easily sustained from the south and the east to the northwest (mainly because of the distance), the markets located in the northwest will tend to be different from those located in the south and the east. In other words, these two locations (the east and the south on one hand, and the northwest on other hand) will be less likely to be in the same market. In our empirical study, we will use the price data of Texas and Georgia to represent southern states and the prices data of Pennsylvania and Washington State to represent, respectively, the eastern and northwestern states. The choice of our data is dictated by their availability and our desire to look at the markets in the south and the east versus the markets in the northwest.

5. Data and estimation results

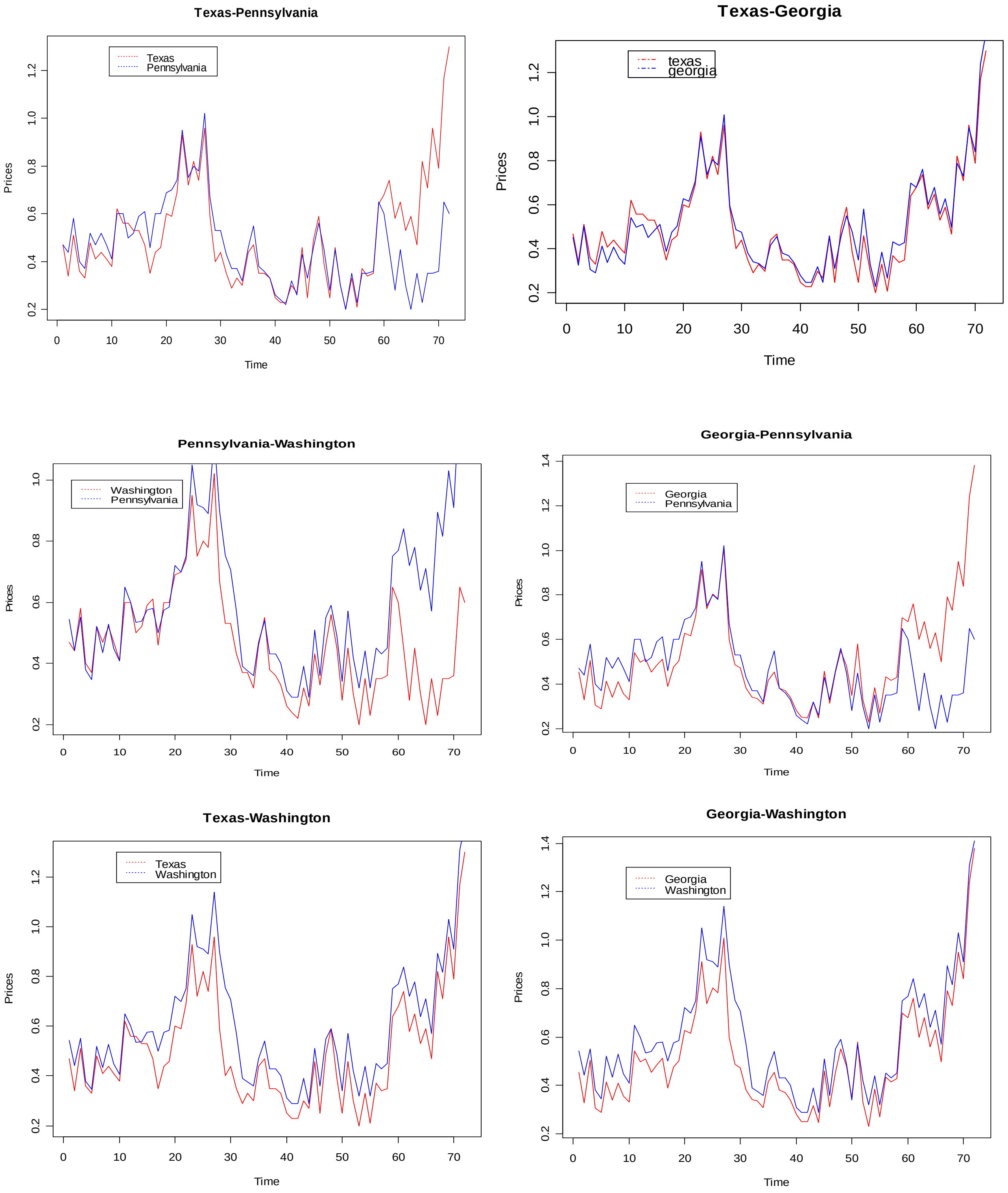
5.1 Data

The data are 72 monthly egg prices of 4 states obtained form the United States Department of Agriculture - National Agricultural Statistics Service (USDA-NASS). We have data for the period of January 2002 to December 2007. The markets of the states analyzed in this study are Texas, Georgia and Pennsylvania for the southern and eastern markets and Washington State for the northwestern market. As mentioned in the section above, our choice of these markets is motivated by the fact that the egg industry is more productive in the southern and eastern states than in the rest of the country. For the northwest states, only the Washington State's data is considered because we do not have any data for states such as Montana and Oregon. We, also, believe that the distance of Washington State from the South and the East is big enough for the possibility of "isolated" markets. Obviously, the transportation issues are very important in determining whether the markets in two different geographic locations are integrated. The summary statistics of the four markets are represented in the table below.

Table 2: **Summary statistics of price data**

States	**Mean**	**Standard deviation**	**Minimum**	**Maximum**
Texas	0.503	0.222	0.200	1.300
Georgia	0.517	0.226	0.230	1.380
Pennsylvania	0.468	0.173	0.200	1.020
Washington	0.599	0.240	0.290	1.410

From these summary statistics we can see that the average egg prices of the eastern and southern states (Texas, Georgia and Pennsylvania) are lower than the average prices of the northwestern state (Washington State). A look at the map in the appendix shows that among all the States, Pennsylvania had the highest production in 2007 (about 6,392 millions of eggs) and therefore, had the lowest price. The next highest productive state in our study was Texas (about 4,994 millions of eggs) followed by Georgia (about 4,792 millions of eggs) and lastly we had Washington State (about 1,520 millions of eggs). Also among the four States, Washington State has achieved the highest minimum price and the highest maximum price. The graphs showing the trend of the pairs of market prices used in the study are presented below.



<u>Figure1</u>: Plot of the level of prices for the six market-pairs**.**

5.2 Estimation results

The maximum likelihood estimation results of the model described in equation [9] is given in the tables below. We first present the results under the half-normal distribution assumption, and second we present the results under the exponential distribution assumption.

Table 3: Half-Normal Distribution Model results

	Texas Pennsylvania	**Texas Georgia**	**Texas Washington**	**Georgia Pennsylvania**	**Georgia Washington**	**Pennsylvania Washington**
β	0.194	0.214	0.399	0.614	0.246	0.466
	(0.21)	(1.320)	(2.220)*	(4.950)*	(1.960)*	(1.850)**
σ_v^2	1.417	0.306	0.190	0.943	0.216	1.223
	(5.74)*	(3.030)*	(3.170)*	(3.210)*	(4.240)*	(3.030)*
σ_η^2	0.037	2.305	0.945	2.814	1.306	2.039
	(0.220)	(1.590)	(2.020)*	(3.070)*	(2.110)*	(1.600)
D	-2.930	-3.226	-2.219	-2.035	-2.468	-2.094
	(-6.120)*	(-21.650)*	(-18.490)*	(-12.880)*	(-27.120)*	(-5.640)*
$EXP(D)$	0.053	0.039	0.109	0.131	0.085	0.123
	(2.088)*	(6.711)*	(8.333)*	(6.329)*	(10.989)*	(2.695)*
$LogL$	-115.691	-82.483	-70.197	-118.655	-69.743	-120.726

Table 4: Exponential Distribution Model results

	Texas Pennsylvania	**Texas Georgia**	**Texas Washington**	**Georgia Pennsylvania**	**Georgia Washington**	**Pennsylvania Washington**
β	0.145	0.207	0.158	0.102	0.122	0.104
	(0.895)	(1.848)**	(1.837)**	(2.090)*	(2.140)*	(2.080)*
σ_v^2	1.325	0.345	0.355	1.502	0.265	2.234
	(3.371)*	(3.255)*	(3.777)*	(7.989)*	(5.333)*	(5.488)*
θ^2	0.934	1.235	1.932	0.084	2.668	0.154
	(1.644)**	(2.061)*	(1.507)	(18.26)*	(0.815)	(7.700)*
D	-2.914	-3.255	-3.27	-2.001	-2.537	-2.245
	(-21.120)*	(-27.120)*	(-29.730)*	(-9.615)*	(-32.530)*	(10.112)*
$EXP(D)$	0.054	0.039	0.038	0.135	0.079	0.106
	(7.246)*	(6.369)*	(9.091)*	(4.348)*	(12.821)*	(3.717)*
$LogL$	-115.691	-85.049	-70.750	-129.990	-70.375	-112.601

The asymptotic z-statistics are in the parentheses. * Significant at 5% level, ** Significant at 10% Level.

The results from the half-normal distribution assumption in Table 3 suggest that all egg markets are highly integrated except Georgia-Pennsylvania markets and Pennsylvania-Washington markets. Indeed the probability of integration is 0.806 for Texas-Pennsylvania markets, 0.786 for Texas-Georgia markets, 0.601 for Texas-Washington markets, 0.386 for Georgia-Pennsylvania markets, 0.754 for Georgia-Washington markets and 0.534 for Pennsylvania-Washington markets. The variances (σ_v^2) of the price random components are statistically significant suggesting that there are many stochastic shocks that affect the prices. Also, the average transaction costs between the markets are statistically different from zero. This is evidence that at least the transportation costs will be non zero between these markets pairs.

The results from the assumption of exponential distribution in Table 4 suggest that all the β's are low implying that the markets pairs are highly integrated with this model as well. These findings are similar to the results of the assumption of half-normal distribution. However, the degrees of integration of markets are higher in the exponential model than in the half normal model consistently with our previous Monte Carlo study. Typically, our empirical results and the Monte Carlo results support the fact that the half-normal model, tends to underestimate the degree of integration while the exponential model tends to overestimate the degree of integration.

In our model, we also assume that if these markets are integrated in the sense that they are in the same economic markets, the average difference in prices of the market will be closely equivalent to the transaction cost found in our model. Table 5 below displays this comparison.

.

Table 5: Model performance in the estimation of the transaction costs

	Texas Pennsylvania	Texas Georgia	Texas Washington	Georgia Pennsylvania	Georgia Washington	Pennsylvania Washington
Actual Data	0.036	0.014	0.095	0.049	0.081	0.130
Half-Normal	0.053	0.039	0.109	0.131	0.085	0.123
Exponential	0.054	0.039	0.039	0.135	0.079	0.106

Overall, both models performed well in terms of predicting the transaction costs. Some of the differences between the actual transaction costs and the models predictions might be related to the relatively small sample size of our data and to the exogenous shocks on the prices that we found very significant.

6. Conclusion

In this paper, we applied a modified version of the 1986 Spiller and Huang paper to analyze the degree of integration of the United States egg market. We found that the markets considered in the model are highly integrated except the Georgia-Pennsylvania markets and the Pennsylvania-Washington markets that are less integrated. The two different error distribution assumptions made did not yield significantly different estimates. Also, the models reveal the presence of significant stochastic shocks on the prices. When we compare the price differences of the market to the transaction cost calculated from the model, we find that the model performs well in terms of the transaction costs estimation. A further investigation of this model will be to apply it to commodity markets with large sample observations. These future works may validate our Monte Carlo results and give us a better estimate of the transaction costs.

References

Aigner, D. Lovell, C. A. K. and Schmidt, P. (1977), "Formulation and estimation of stochastic frontier production models", *Journal of Econometrics*, Vol.6, No.1 (July), pp. 21-37.

Greene, W. (2003), *Econometric analysis,* 5th ed., Prentice Hall, New York.

Marshall, A., (1961), *Principles of economics* 325, Variorum edition.

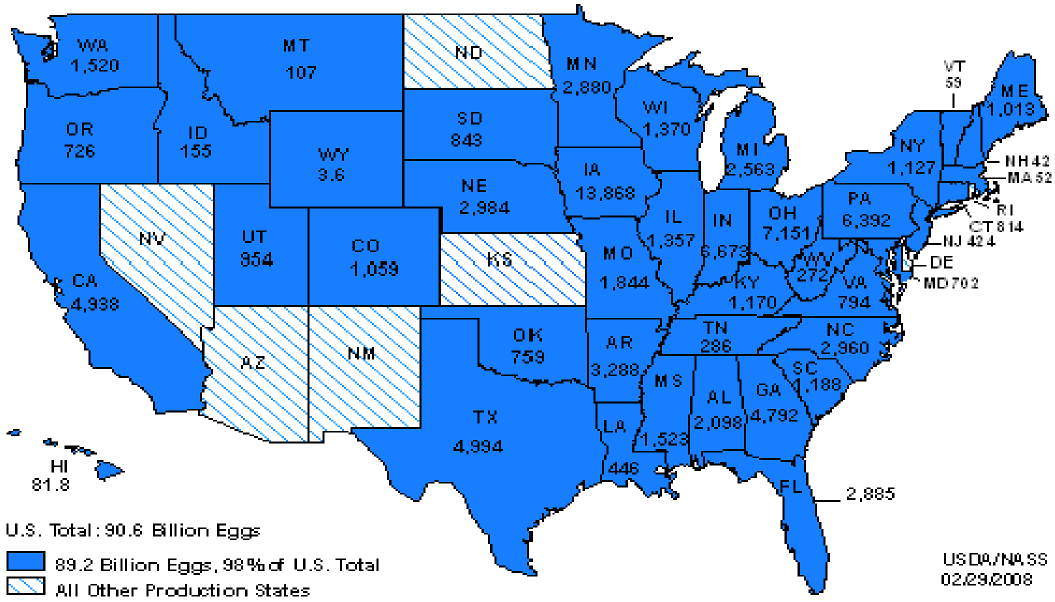
Spiller, P.T. and Huang, C. J. (1986), "On the extent of the Market: Wholesale gasoline in the northern United State". *Journal of industrial economics,* Vol. XXXV, No.2. (December).

R Development Core Team (2008), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Stigler, G.J and Sherwin, R.A. (1985), "The Extent of the market". *Journal of Law and Economics*, Vol.28, No.3. (Oct.), pp.555-585.

Appendix

ANNUAL EGG PRODUCTION BY STATES
NUMBER PRODUCED, MILLION, 2007



USDA/NASS
02/29/2008