M.S. COMPREHENSIVE EXAMINATION

June 2022

You have four hours, after a 20 minute reading period. You do not need to use the whole time period. You need to answer all three questions, which will be weighted equally.
Watch the time carefully. The logic used to answer each question is important, so be sure to specify your reasoning with full sentences. You may support your answers using diagrams or mathematical derivations where appropriate. If you use graphs, make sure that they are large enough. We are expecting precise and concise answers. Also make sure your writing is legible: if we cannot read it, your answer will be assumed wrong.

I. The welfare costs of U.S. dairy policy (5 points)

In this problem, we will demonstrate how simple economic information can be used to compute the welfare costs of farm policy. Specifically, we consider the U.S. federal milk marketing orders (FMMOs), which are market institutions governing the transformation milk at the regional level. To simplify the analysis, we consider a single region where dairy products are produced and consumed, and we ignore trade.

The key prerogative of a FMMO is to implement a price premium for farm milk to be used in the manufacturing of fluid milk relative to other dairy products such as cheese or yogurt. That is, manufacturers purchasing farm milk for the purpose of producing fluid milk must pay a higher price than manufacturers of other dairy products. The FMMO makes sure that there is no resale of farm milk from manufacturers of other dairy products to manufacturers of fluid milk.

[0.1] 1. Briefly explain why there is an incentive for manufacturers of fluid milk to try and purchase farm milk from manufacturers of other dairy products.

Assume that there is a single consumer whose preferences over milk (x_1), other dairy products (x_2), and all other goods (x_0) are representable by the following utility function:

$$U(x_0, x_1, x_2) = x_0 + \beta \ln \left\{ \left[(\beta_1)^{\frac{1}{\sigma}} (x_1)^{\frac{\sigma-1}{\sigma}} + (\beta_2)^{\frac{1}{\sigma}} (x_2)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \right\}$$

where $\beta > 0$, $\beta_1 \geq 0$, $\beta_2 \geq 0$, $\beta_1 + \beta_2 = 1$, and $\sigma > 0$.

We normalize the price of x_0 to one and denote p_i the price of x_i, for $i = 1, 2$.

[0.2] 2. What is the consumer's income elasticity of demand for x_1 and x_2? There is no need to do any derivation.

[0.2] 3. Interpret the parameter σ in words. Again, no need to do any derivation.

[0.2] 4. If we had more time, using the Lagrange method we would show that the Walrasian demand for x_1 is equal to

$$x_1 = \frac{\beta \beta_1 p_1^{-\sigma}}{\beta_1 p_1^{1-\sigma} + \beta_2 p_2^{1-\sigma}} .$$

Instead, I will just ask you to derive the value of the Lagrange multiplier on the budget constraint.

[0.3] 5. Using the result in part 4, argue that the expenditure on fluid milk and other dairy products, $p_1 x_1 + p_2 x_2$, is a constant, and state what that constant is.

[0.4] 6. Now show that x_1 can also be written as

$$x_1 = \beta \beta_1 \left[\mathcal{P}(p_1, p_2) \right]^{\sigma-1} p_1^{-\sigma}$$

for some price index $\mathcal{P}$ that you will make explicit. Verify that $\mathcal{P}$ satisfies the properties of a price index, and name these properties. (Hint: There are three of them.)

To simplify the problem, we assume that the supply of farm milk is infinitely inelastic, and there are no explicit costs to produce it. We denote by M the fixed quantity of farm milk and by m the price of farm milk paid by manufacturers of other dairy products.

Fluid milk is produced from farm milk according to fixed proportions, one unit of farm milk giving one unit of fluid milk. The per-unit processing cost for fluid milk is denoted $c_1 > 0$. All other dairy products are also produced from farm milk using fixed proportions, and by choice of units we can assume that one unit of farm milk gives one unit of other dairy products, with per-unit processing cost $c_2 > 0$.

[0.2] 7. Write the relationship between the quantity of farm milk (M) and the quantities of fluid milk (x_1) and other dairy products (x_2).

Since there are constant returns to scale in the production of fluid milk and other dairy products, in equilibrium dairy product manufacturers make zero profit.

[0.2] 8. Write the relationship between the price of farm milk paid by manufacturers of other dairy products and the price of other dairy products in equilibrium.

Assume that the FMMO mandates that manufacturers of fluid milk pay a relative premium $\tau > 1$ over the price of milk paid by manufacturers of other dairy products. That is, the price paid by fluid milk manufacturers is $\tau m > m$.

[0.2] 9. Write the relationship between the price of farm milk and the price of fluid milk in equilibrium.

We now consider the comparative statics of the equilibrium with respect to the milk price premium, that is, how the equilibrium responds to policy-induced changes in τ. For instance, one may be interested in knowing what would happen if the FMMO were eliminated. We denote with "primes" the values of variables after the change, and with "hats" the *relative changes* in these variables, meant here as the *ratio of the final value to the initial value*. For instance, $\hat{\tau} \equiv \frac{\tau'}{\tau}$ denotes the ratio of the final value of τ to its initial value, $\hat{m} = \frac{m'}{m}$ denotes the relative change in the farm milk price, $\hat{x}_1 \equiv \frac{x_1'}{x_1}$ denotes the relative change in the consumption of fluid milk, and so on.

[0.1] 10. What would be the value of $\hat{\tau}$ corresponding to the elimination of the FMMO?

[0.2] 11. Show that the relative change in x_1 can be written as a function of the relative changes in p_1 and $\mathcal{P}$ as

$$\hat{x}_1 = \hat{\mathcal{P}}^{\sigma-1}\hat{p}_1^{-\sigma} \,.$$

(Hint: Use part 6.)

[0.5] 12. Show that

$$\hat{\mathcal{P}} = \left[b_1\hat{p}_1^{1-\sigma} + b_2\hat{p}_2^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$$

for parameters b_i that you will make explicit and interpret. (Hint: Do not overspend time on this part.)

[0.3] 13. Show that

$$s_1\hat{x}_1 + s_2\hat{x}_2 = 1$$

for parameters s_i that you will make explicit and interpret. (Hint: Use part 7.)

[0.5] 14. Using parts 8 and 9, show that

$$\hat{p}_1 = \psi_1 \hat{m} \hat{\tau} + 1 - \psi_1$$

and

$$\hat{p}_2 = \psi_2 \hat{m} + 1 - \psi_2$$

for parameters ψ_i that you will make explicit and interpret. (Hint: Start with the second equality.)

[0.5] 15. Now summarize the set of equalities that determine the equilibrium in relative changes. Check that there are as many equations as unknowns. State explicitly what the unknowns are. Summarize the calibrating information needed to solve this equilibrium, for a given value of $\hat{\tau}$. (Hint: I am not asking you to solve the equilibrium; there are no new derivations involved.)

To wrap up, we will show how to compute the change in farm profit $\Delta\Pi \equiv \Pi' - \Pi$, as well as the change in consumer surplus, which in this case is also equal to the change in utility $\Delta U \equiv U' - U$. We assume that the consumer's income is unaffected by the change.

[0.2] 16. Show that the change in farm profit relative to the value of milk is

$$\frac{\Delta\Pi}{mM} = \hat{m} - 1 \, .$$

[0.5] 17. Show that

$$\Delta U = \beta \ln \left[b_1 \left(\hat{x}_1\right)^{\frac{\sigma-1}{\sigma}} + b_2 \left(\hat{x}_2\right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \, .$$

(If we had more time, we could also show an even cooler result, namely that $\Delta U = -\beta \ln \hat{\mathcal{P}}$!)

[0.2] 18. Deduce the change in consumer surplus relative to the expenditure on dairy products, that is, the value of $\frac{\Delta U}{p_1 x_1 + p_2 x_2}$. (Hint: Use part 5.)

II. Choice of mode of transportation (5 points)

Suppose that you would like to investigate the choice between modes of commuting to work. Assume that an individual makes a choice between driving a car and taking public transportation to work. The main variables that determine their decision are the gas price G_i (in dollars per gallon), the public transportation fare F_i (in dollars), and the average waiting time for public transportation W_i (in minutes). There are two outcomes, using public transportation ($Y_i = 1$) or driving a car ($Y_i = 0$).

1. The first model specification that you try is the following:

$$\Pr\{Y_i = 1\} = \beta_0 + \beta_1 G_i + \beta_2 F_i + \beta_3 W_i + \beta_4 W_i^2. \tag{1}$$

   (a) Provide an interpretation for coefficients β_1 and β_2. What are the expected signs of these coefficients? Explain.

   (b) What does the sign of β_4 tell us about the readiness to wait for a bus/train an extra minute? Give an example that would imply a specific sign for β_4.

   (c) Write down a linear regression model corresponding to (1). What condition would the corresponding regression residual satisfy?

   (d) Discuss drawbacks and limitations of the linear model specification (1) for estimation of the choice model.

2. Then you come up with the following latent index model:

$$Y_i^* = \beta_0 + \beta_1 G_i + \beta_2 F_i + \beta_3 W_i + \beta_4 W_i^2 + \varepsilon_i, \tag{2}$$

   where Y_i^* is an unobservable index that measures an individual's utility difference between the two options and ε_i represents all unobserved factors affecting their choice. The individual prefers public transportation if $Y_i^* \geq 0$.

   Suppose that $\varepsilon_i \sim \mathcal{N}(0, 1)$. The CDF and PDF of $\mathcal{N}(0, 1)$ are denoted $\Phi(z)$ and $\phi(z)$, respectively. You do not need to use the explicit forms of $\Phi(z)$ and $\phi(z)$ to answer the questions below.

   (a) What is the name of this model? You don't need to elaborate on the properties here.

   (b) Suppose that you know the coefficients in the model. Write down the share of people who prefer driving a car in a city with gas price G^*, public transport fare F^*, and average waiting time W^*.

   (c) What is the marginal effect of a reduction in the waiting time on the share of public transportation users?

   (d) Now suppose that you observe a random sample of N individuals from different cities in a country. Write down the corresponding log-likelihood function.

III. Costs and benefits of land fragmentation (5 points)

In many contexts, farmers simultaneously produce the same crop on multiple plots of land. A given rice farmer may, for example, produce rice on a total of one hectare of land that is spread across four non-contiguous (i.e., non-neighboring) plots. Such land fragmentation has potentially important economic implications. Based on concerns that land fragmentation raises rice production costs, Japan has long tried to incentivize and encourage consolidation as a way to reduce rice prices.

This question leverages an article on "The costs and benefits of land fragmentation of rice farms in Japan" published by Kawasaki (2010) in the *Australian Journal of Agricultural and Resource Economics*.

1. Conceptualizing the costs and benefits of land fragmentation:

[0.3] (a) Using economic reasoning and concepts, explain how fragmentation could increase production costs. List three distinct potential explanations and discuss specifically how they would raise costs.

[0.3] (b) Some claim that fragmentation may not be all bad and could also bring production benefits in the form of greater diversification. What specifically would have to be true in a given context for land fragmentation to generate such diversification benefits?

[0.2] (c) Like farmers elsewhere, Japanese farmers prefer to continue to cultivate some rice land even as they begin to invest in non-agricultural assets and activities as part of a broader livelihood portfolio. How is this specific diversification motive different than potential diversification benefits associated with land fragmentation?

[0.3] (d) Use the example of a rice farmer with one hectare of land spread across four different plots to discuss specifically how transaction costs may discourage consolidation in the absence of explicit incentives or encouragement. (Hint: Complete consolidation would entail this farmer cultivating one contiguous hectare of land, i.e., not fragmented into four different plots.)

2. Data and transition to empirical analysis:

[0.4] (a) Kawasaki (2010) uses panel data from a national survey of Japanese rice farmers. He constructs a Simpson index (SI) defined by this formula as his main measure of land fragmentation

$$1 - \frac{\sum_i A_i^2}{(\sum_i A_i)^2} \, .$$

Based on this formula, provide a brief explanation and interpretation of this key SI measure, including its range. Why is this SI measure better than the number of rice plots cultivated by a given farmer?

[0.4] (b) Table 1 provides some basis descriptive evidence on the extent of land fragmentation in the data Kawasaki (2010) uses. (For your reference, more complete summary statistics are provided at the conclusion of this question as Table 2.) Provide a concise 2–3 sentence description of the primary patterns you see in this table.

Table 1 Mean values of land fragmentation indices by farm size

	Obs.	Area planted (ha)	No. of parcels	Simpson Index (SI)
Under 0.5 ha	3,553	0.35	2.20	0.596
0.5–1 ha	4,491	0.73	3.03	0.753
1–2 ha	4,798	1.43	4.00	0.839
2–4 ha	2,872	2.72	5.05	0.900
4–8 ha	1,592	5.60	6.11	0.943
Over 8 ha	1,477	16.20	9.29	0.966
All classes	18,783	2.77	4.18	0.801

[0.3] (c) According to Kawasaki (2010): "As it is difficult for farmers to choose their plot location or control the degree of fragmentation. . . the [SI] is assumed to be exogenous." (pg. 513) Explain why this is an important claim given the title of this article.

[0.3] (d) As mentioned, the Japanese government has long tried to encourage land consolidation among rice farmers. Does this fact change the credibility of this important claim? If so, how?

3. Cost of land fragmentation:

   (a) Kawasaki (2010) uses a translog functional form to estimate a stochastic frontier for the cost of rice production. As described in the article (pg. 512):

3.1 Model

In this section, attention is focussed on the cost side of fragmentation. The Battese and Coelli (1995) stochastic frontier model is employed, which allows for identifying factors that may explain differences in efficiency levels between farms. Consider the following function,

$$\ln C_{it} = \alpha_0 + \alpha_Y \ln Y_{it} + \frac{1}{2}\alpha_{YY} \ln Y_{it} \ln Y_{it} + \sum_m \alpha_m \ln P_{mit} + \sum_m \alpha_{Ym} \ln Y_{it} \ln P_{mit} + \frac{1}{2}\sum_m \sum_n \alpha_{mn} \ln P_{mit} \ln P_{nit} + \sum_t \alpha_t T_t + U_{it} + V_{it}. \quad (1)$$

Here, we employed well-known translog form (Christensen *et al.* 1973) where C represents the production cost, Y is the output, P is the input price, T is the year dummy capturing technological change, U is a non-negative random variable associated with cost inefficiency, V corresponds to statistical noise distributed with $N(0, \sigma_V^2)$, and αs are parameters to be estimated. It is further assumed that U and V are independently distributed from each other. Subscript i and t denote farm and time (year), respectively, and subscript m or n denote inputs. There are four types of inputs: land, capital, labor, and materials.

[0.2] i. Explain briefly how estimating a stochastic frontier model is different than estimating a standard cost function.

[0.2] ii. Explain how this "well-known translog form" is different than a Cobb-Douglas functional form. What would have to be true about a given production context for the translog form to be preferred to a Cobb-Douglas form?

[0.3] iii. How could you econometrically test whether the translog or the Cobb-Douglas form is better suited for a particular dataset?

[0.4] (b) Kawasaki (2010) situates Table 1 and his stochastic cost frontier results (not shown here) in the context of an ongoing debate about economies of scale on Japanese rice farms (pg. 519):

> Among the agricultural economics literature (see, for example, Castle 1989 and Alvarez and Arias 2003), there is debate as to whether economies of size exist on large size farms or disappear (average cost curve is L-shaped), or whether diseconomies exist (U-shaped). For Japanese rice farms, several authors (e.g., Kako 1983, 1984; Chino 1985) has tackled this issue using aggregated data and found that economies of size disappeared on farms of sizes over 5 ha. However, our results imply that the economies of size do not disappear, even at sizes of 16 ha (average farm size of the largest category. See Table 1). One of the reasons for such a difference is that aggregated data studies did not control for the impact of land fragmentation. As positive correlation exists between the degree of fragmentation and farm size as shown in Table 1, if output (size) increases, fragmentation will be exacerbated and partially offsets economies of size. As aggregated data studies did not use fragmentation variables explicitly, the impact of such an offset effect is not excluded in their estimates. On the other hand, it is excluded in this article, meaning that economies of size are derived when output is assumed to increase without exacerbation of fragmentation.

In your own words, explain the essence of his argument in this paragraph. What exactly is he claiming as his contribution to this debate about economies of scale? How compelling do you think his claim is?

4. Application of Just and Pope production function:

[0.3] (a) In order to estimate the potential diversification benefits of land fragmentation, Kawasaki (2010) uses a Just and Pope production function defined as:

$$\mathbf{y}_{it} = f(\mathbf{x}_{it}) + e^*_{it} = f(\mathbf{x}_{it}) + [h(\mathbf{z}_{it})]^{1/2} e_{it}$$

where $\mathbf{y}_{it}$ is rice yield in production per hectare. Explain how this production function is different than a conventional production function and why it is an appropriate function to use in the context of this analysis.

(b) To estimate this Just and Pope production function, Kawasaki (2010) first estimates the $f(\cdot)$ function, then uses the log of the squared residuals from this regression as the dependent variable to estimate the $h(\cdot)$ function. The estimation results from this second stage are shown below as Table 6. (Note: Kawasaki refers to the omitted category for dummy variables as "standard.")

[0.2] i. Using the QD results, carefully and concisely interpret the estimated coefficient on *Capital*. Provide an explanation for this result that you think makes sense given what you know about this research context.

[0.2] ii. Using the QD results, carefully and concisely interpret the *Year dummies* in 2-3 sentences (1995 is the omitted ("standard") category). What can you conclude about these different years based on these estimated coefficients?

[0.4] (c) Carefully interpret and discuss the SI coefficient in the QD results in Table 6 in the context of Kawasaki's research question.

[0.3] (d) Explain how this estimated SI coefficient could be used to determine the cost of land fragmentation *net of diversification benefits*. Be specific.

Table 6 Estimates of variance function parameters

Production function specifications	Generalized Leontief (GL)	Quadratic (QD)
Land fragmentation indices		
No. of parcels	−0.0107	−0.0125
	[1.8]*	[2.1]**
Simpson Index (SI)	−0.527	−0.542
	[3.3]***	[3.4]***
Labor (log)	0.004	−0.014
	[0.1]	[0.3]
Materials (log)	0.083	0.104
	[1.1]	[1.4]
Capital (log)	−0.0534	−0.0490
	[2.2]**	[2.0]**
Land (log)	−0.026	−0.031
	[0.3]	[0.4]
Family labor ratio	−0.310	−0.355
	[1.4]	[1.6]
Geographical dummies (flat farming area is standard)		
Urban area	−0.084	−0.087
	[1.4]	[1.5]
Hilly or mountainous area	0.102	0.110
	[2.1]**	[2.3]**
Land improvement ratio dummies (under 50% is standard)		
50–80%	0.016	−0.006
	[0.2]	[0.1]
Over 80%	−0.034	−0.035
	[0.6]	[0.6]
Outsourcing dummy	−0.016	−0.020
	[0.3]	[0.4]
Regional dummies (Tohoku region is standard)		
Hokuriku	−0.214	−0.218
	[3.4]***	[3.5]***
Kanto/Tosan	−0.113	−0.116
	[1.8]*	[1.8]*
Tokai	−0.291	−0.300
	[3.2]***	[3.2]***
Kinki	−0.053	−0.091
	[0.6]	[1.1]
Chugoku/Shikoku	0.057	0.061
	[0.8]	[0.8]
Kyushu	0.183	0.172
	[2.4]**	[2.3]**

	GL	QD
	(continued)	
Year dummies		
1996	−0.531	−0.526
	[6.1]***	[6.0]***
1997	−0.365	−0.385
	[4.2]***	[4.5]***
1998	−0.088	−0.109
	[1.0]	[1.2]
1999	−0.358	−0.362
	[4.1]***	[4.1]***
2000	−0.428	−0.438
	[4.8]***	[4.9]***
2001	−0.354	−0.362
	[4.0]***	[4.0]***
2002	−0.363	−0.387
	[3.8]***	[4.0]***
2003	−0.061	−0.079
	[0.6]	[0.8]
2004	0.074	0.076
	[0.7]	[0.7]
2005	−0.648	−0.638
	[5.5]***	[5.4]***
2006	−0.334	−0.373
	[2.8]***	[3.1]***
Constant	2.03	1.91
	[2.9]***	[2.7]***
Observations	13268	13268
R^2	0.015	0.016

Note: Asterisk (*), double asterisk (**) and triple asterisk (***) denote variables significant at 10%, 5% and 1% respectively. Absolute value of t statistics in brackets.

Table 2 Summary statistics

Variable	Unit	Mean	Standard deviation	Minimum value	Maximum value
Production cost	1000 yen	2658	2872	70	34900
Output	1000 kg	9.60	13.08	0.66	166.95
Yield	1000 kg/ha	5.22	0.67	2.56	7.89
Factor price					
Wage	1000 yen/hour	1.53	0.22	0.79	2.42
Material price	–	1.00	0.02	0.96	1.16
Land rent	1000 yen/ha	239	95	24	1370
Capital price	–	0.54	0.58	0.004	4.96
Factor input					
Labor	hour	541	502	10	6742
Material	–	422941	552134	15095	7890924
Land	ha	1.81	2.47	0.13	37.18
Capital	1000 yen	2737	3681	31	66500
Land fragmentation indices					
No. of parcels	–	4.08	3.91	1.00	90.00
Simpson Index (SI)	–	0.79	0.17	0.00	0.99
Family labor ratio	–	0.95	0.09	0.00	1.00
Geographic dummies					
Flat farming area	0 or 1	0.48	0.50	0	1
Urban area	0 or 1	0.18	0.38	0	1
Hilly or mountainous area	0 or 1	0.35	0.48	0	1
Land improvement ratio dummies					
Under 50%	0 or 1	0.18	0.38	0	1
50–80%	0 or 1	0.10	0.30	0	1
Over 80%	0 or 1	0.72	0.45	0	1
Outsourcing dummy	0 or 1	0.64	0.48	0	1

Total observations = 13268.