Micro Theory Question

This question asks you to explore various dimensions of one key shock to food markets due to the coronavirus pandemic. Meat processing (pork, beef, chicken) was severely disrupted this spring when outbreaks of the virus among processing plant workers caused the temporary shutdown of several U.S. processing plants, resulting in higher consumer prices (P^R) and lower prices for live cattle (P^F). This increase in the farm-to-retail price spread, P^R- P^F, has led to calls from farm groups and politicians to investigate meat packers for violations of U.S. antitrust laws, i.e., laws to prevent unlawful exercise of market power.

a. Suppose you were asked to investigate the impact of the retail price changes on consumer welfare. To limit our scope, let us focus on a single meat product, pork. How would you propose to compute the welfare losses to U.S. pork consumers from this price increase? Focus on the relevant economic theory here and *not* on empirical measurement issues. In other words, assume you will be able from the available data to know the magnitude of price increase for pork products and how long the price increase persisted.

A good answer here should have you defining and discussing concepts of equivalent variation (EV), compensating variation (CV), and/or consumer surplus (CS). The best measure for this setting would be CV because it is correct theoretically and it is an ex post measure, for a price change that has already occurred. EV is ex ante and, thus, is less appropriate than CV here. CS only approximates a welfare impact due to the income effects that are present in the Marshallian demand function and, thus, CS. However, pork is a small budget share for most people, so you could say that income effects would be unimportant here and work directly with CS.

In addition to identifying the concept that you would propose to use and why you have chosen it, you should talk about how the welfare analysis would be implemented. You would need to have an estimate of the Marshallian demand for pork in the US, from which you could derive the Hicksian demand to compute CV or EV by removing the income effect from the Marshallian curve. You would probably want to look at different time intervals because prices fluctuated during this period. For example, you could compute welfare impacts across different weeks of the pandemic based on the time path of pork prices.

Graphs, though not essential, would be good to illustrate your approach.

b. For many politicians and commentators, it did not make sense that consumer prices were higher after the plant closures, but farm prices were lower. This is what led to suspicion that the meat processors were manipulating the market. Suppose, however, that meat markets are perfectly competitive. Focusing again on a single meat, pork, use one or more graphs, given our assumption of perfect competition, to show how higher P^R and lower P^F could emerge as equilibrium responses to the processing plant closures.

The plant closures disrupted supply of pork to retail. You could depict this as a simple shift of the retail supply function to the left, holding demand constant. An even better answer could recognize that retail demand for pork likely increased during the pandemic due to either *consumer hoarding or increased purchases for food at home given that restaurants and most food service were shutdown during this time, forcing consumers to eat most of their meals at home. The supply shift is essential. The demand shift is nice but inessential to a successful answer.*

You also need to address what is happening to the demand and supply for live hogs to process. The supply coming off farms would not have been affected by the processing plant shutdowns, but demand would have been reduced because the plants that closed due to COVID 19 would not have been procuring hogs. Thus, you should show a leftward shift in demand for live hogs due to the plant closure, holding supply constant, so farm price and quantity fall, matching the price movements that we observed and that led to the aforementioned concern.

c. Obviously, the quantities, Q, of pork sold by pork processors declined as a result of the plant closures. However, I am going to propose the following proposition: *sales revenues earned by pork processors actually* ***increased*** *as a result of the plant closures.* Your job is to discuss this proposition. You might agree or disagree, but the key to a good answer is to explain the economic factors that would go into determining whether my proposition is true or not. Graph(s) may be helpful. Define carefully any notation you introduce.

Total revenue earned by pork processors is simply the wholesale price (P^w) they received per unit (kilogram) sold times the kg. sales volume (Q). Let us write the inverse wholesale demand as $P^w(Q), P^{w\prime} < 0$. Thus, total revenue is the function TR(Q)= $P^w(Q)Q$. Whether TR(Q) is increasing or decreasing in Q depends on whether marginal revenue, MR(Q) = $\frac{dTR(Q)}{d(Q)}$ is positive or negative at current sales volumes. This, though, is just an elasticity of demand question. Ideally you remember that there is a precise mathematical relationship between MR and the price elasticity of demand, $\varepsilon_{Q,P}$ such that MR(Q) > 0 when demand is elastic, MR(Q) = 0 when $\varepsilon_{Q,P} = -1$, and MR(Q) < 0 when demand is inelastic. Thus, assuming a constant demand curve during this period, my validity of my proposition depends upon whether demand is inelastic (in which case my proposition is true) or elastic (in which case my proposition is not true). Note that if you believe retail demand for pork increased due to the pandemic (which likely was true), then it increases the likelihood that my proposition is true because the demand shift will contribute to higher prices.

d. Pork processing in the U.S. is highly concentrated, i.e., a few firms procure most of the live hogs sold in the U.S., and sell most of the fresh pork products in the U.S. Suppose you were building an economic model of the U.S. pork industry and wanted to allow for the possibility that pork processors exercised market power in (i) acquiring live hogs from farmers and (ii) selling processed pork products downstream to food service and retailers. Show how you would model pork processors' behavior to allow for the exercise of market power. A good answer here will involve setting up an objective function for a representative processor and deriving first-order conditions. Hint: You can always choose your units of measurement (i.e. normalize) so that one unit of live hog produces one unit of processed pork and thereby have a single quantity variable for each.

Here I'm looking for an answer right out of 204B. Let purchases of live pork and sales of processed pork for a representative processor by denoted by q. Remember, by choosing units of measurement properly, I am requiring one unit of live hog to produce one unit of processed pork. I have already defined the inverse retail demand function, $P^w(Q)$*. Let the inverse farm supply of hogs to the processing sector be S(Q), where* $S'(Q) > 0$*. The objective function of the processor can thus be expressed as:*

$$max\, q \quad \pi = P^w(Q)q - cq - S(Q)q,$$

where I have introduced the parameter c to account for per-unit processing costs excluding the farm product. The first-order condition to this problem is as follows:

$$\frac{d\pi}{dq} = P^w(Q) + P^{w\prime}(Q)\frac{dQ}{dq} - c - (S(Q) + S'(Q)\frac{dQ}{dq} = 0.$$

The term $\frac{dQ}{dq}$ *measures departures from perfect competition in this market. If* $\frac{dQ}{dq} = 0$*, the FOC reverts to the competitive FOC of equating an exogenous price with the processor's marginal cost.*

Recall that $\frac{dQ}{dq}$ *can be written as* $\frac{dQ}{dq} = \frac{dq}{dq} + \frac{dQ_{\sim}}{dq}$

Smoking and COVID-19.

Suppose that you would like to investigate cumulative effects of long-term smoking on infection fatality rate (IFR) of a new acute respiratory disease, COVID-19. You have data from one country for all individuals who tested positive for the novel coronavirus. For simplicity, assume that every individual, infected with the virus either dies or fully recovers within 14 days after being infected, and your sample only includes people who were tested positive at least 14 days ago. Suppose you observe age of the individuals, A_i, years of smoking, S_i, and whether they recovered or passed away, Y_i. If a person recovers, $Y_i = 0$. Otherwise $Y_i = 1$. Let $D_i = 1$ if $S_i > 0$ and $D_i = 0$ if $S_i = 0$.

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

$$P\{Y_i = 1\} = \beta_0 + \beta_1 D_i + \beta_2 D_i S_i + \gamma_1 A_i + \gamma_2 A_i^2. \quad (1)$$

   (a) Provide an interpretation for coefficients β_1 and β_2.

   Solution: β_1 captures a average (treatment) effect of starting smoking, while β_2 captures marginal effects of additional year of smoking.

   (b) Why is it important for estimation of β_2 to include term $\beta_1 D_i$ in (1)? (Hint: people with known underlying lung conditions may stay away from trying smoking, and thus represent a different population)

   Solution: D_i is positivly correlated with S_i. So omission of D_i may result in a bias in estimates of β_2. There are multiple possible answers to this questions: 1) including extra variables can only reduce chance of model misspecification; 2) people who never tried smoking may have underlying health conditions (D_i is positively correlated with lung conditions), so omitting $\beta_1 D_i$ may result in downward bias in β_2; 3) people who don't smoke make other healthy choices (D_i is negatively correlated with lung conditions), so omitting $\beta_1 D_i$ may result in positive bias in β_2.

   (c) Write down a linear regression model corresponding to (1).

   Solution:
$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 D_i S_i + \gamma_1 A_i + \gamma_2 A_i^2 + U_i,$$
   where $E(U_i|S_i, A_i) = 0$.

   (d) Suppose that the data (Y_i, S_i, A_i) is i.i.d. Is the linear regression model corresponding to (1) homoscedastic or heteroscedastic? Does your answer affect consistency of the OLS estimates of β_1 and β_2?

> **Solution:** Using the formula for the variance of a binary r.v. we get
>
> $$Var(U_i|A_i, S_i) = Var(Y_i|A_i, S_i) = P\{Y_i = 1|S_i, A_i\}(1 - P\{Y_i = 1|S_i, A_i\}),$$
>
> where $P\{Y_i = 1|S_i, A_i\}$ was defined in part 1.a. Clearly, the conditional variance is not a constant; the linear model is heterscedastic by definition. Heterscedasticity is irrelevant for consistency of the OLS estimator.

(e) It is well known, that respiratory diseases tend to have more severe outcomes in older populations. What would be the likely direction of the bias in β_2 if you would omit all the terms with A_i in (1)?

> **Solution:** A_i and S_i are positively correlated: one need to live long enough to smoke long enough. So if we omit A_i and A_i^2, the estimator of β_2 will capture the effects of age and thus will be upward biased.

(f) Discuss drawbacks and limitations of the linear model specification (1) for estimation of the IFR.

> **Solution:** This question can have multiple answers. Some possibilities include: 1) Predicted probability of death may be outside of the interval $[0, 1]$; 2) Marginal effects do not depend of the regressors; 3) Model is likely misspecified, as linear functional form is too artificial for a probability of death.
>
> Also note, that (1) does not take into account sample selection issues. It was implicitly assumed here that we have the data about all infected people in the country and all people got tested for the virus without conditioning on their symptoms.

2. Then you came up with a latent index model

$$Y_i^* = \beta_0 + \beta_1 D_i + \beta_2 D_i S_i + \gamma_1 A_i + \gamma_2 A_i^2 + \varepsilon_i, \tag{2}$$

where Y_i^* is an unobservable index that measures health condition of the lungs and ε_i represents all unobserved factors affecting the lungs condition. Healthy lungs correspond to low values of Y_i^*. The patient dies if Y_i^* surpasses a certain threshold, $\bar{Y}$. Suppose that $\varepsilon_i \sim N(0, 1)$. CDF and PDF of $N(0, 1)$ are denoted $\Phi(z)$ and $\phi(z)$, correspondingly. You do not need to use explicit form 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.

> **Solution:** It is a random utility representation of the Probit model.

(b) Derive IFR, which is conditional probability of death given S_i and A_i.

> **Solution:** For simplicity of notation, lets use vector notation $\theta = (\beta, \gamma)$, $X_i = (1, D_i, S_i, A_i, A_i^2)$.
> $P\{Y_i = 1|S_i, A_i\} = P\{\theta' X_i + \varepsilon_i \geq \bar{Y}|X_i\} = P\{\theta' X_i - \bar{Y} \geq -\varepsilon_i|X_i\} = \Phi(\theta' X_i - \bar{Y})$.
> Here I used the fact that $-\varepsilon_i$ has the same CDF as ε_i.

(c) What is the average effect of trying smoking on IFR? Does it depend on A_i?

> **Solution:** By definition, the average effect of D_i is $E(Y_i|X_i, D_i = 1) - E(Y_i|X_i, D_i = 0) = P\{Y_i = 1|X_i, D_i = 1\} - P\{Y_i = 1|X_i, D_i = 0\}$. From the previous part, it is equal
>
> $$\Phi(\beta_0 + \beta_1 + \beta_2 S_i + \gamma_1 A_i + \gamma_2 A_i^2 - \bar{Y}) - \Phi(\beta_0 + \gamma_1 A_i + \gamma_2 A_i^2 - \bar{Y}).$$

It obviously depends on A_i, since $\Phi(z)$ is a non-linear function.

(d) What is the marginal effect of an additional year of smoking on IFR? Does it depend on A_i?

Solution:
By definition, the marginal effect of S_i is $\frac{\partial E(Y_i|X_i,D_i=1)}{\partial S_i}$ Strictly speaking, the marginal effect is changing discontinuously at $S_i = 0$. Consider first $S_i > 0$. Then the answer is

$$\phi(\beta_0 + \beta_1 + \beta_2 S_i + \gamma_1 A_i + \gamma_2 A_i^2 - \bar{Y})\beta_2.$$

At $S_i = 0$ the marginal effect is the same if $\beta_1 = 0$, otherwise it is infinite (the regression function is discontinuous). Either way, the marginal effect depends on A_i.

Case $S_i > 0$ was sufficient to get full credit for this question.

3. Now suppose that you realized that only symptomatic patients get tested and end up in the sample. We can assume that the symptoms only occur in patients with $Y_i^* \geq \underline{Y}$. Suppose that $\underline{Y}$ is known (one can only identify $\underline{Y}$ if a random sample of infected people is available).

(a) Discuss how $\underline{Y}$ would affect the IFR estimated without taking into account the sample selection.

Solution: An informal argument would suffice here. If we only consider people with symptoms, we observe the upper part of the distribution of Y_i^*. People in this subgroup have higher chances of dying, so it neglecting the sample selection will result in inflated IFR.

(b) Write down the probability of being tested conditional on S_i and A_i and being infected.

Solution:
Using the solution in 2.b, $P\{Y_i^* \geq \underline{Y}|S_i, A_i\} = \Phi(\underline{Y} - \theta' X_i)$

(c) Write down the probability of death conditional on S_i and A_i and being infected and tested.

Solution:
Using the solution in 2.b, the probability of being tested is $P\{Y_i^* \geq \underline{Y}|S_i, A_i\} = \Phi(\theta' X_i - \underline{Y})$ Then the conditional version becomes

$$\frac{P\{Y_i^* \geq \bar{Y}|S_i, A_i\}}{P\{Y_i^* \geq \underline{Y}|S_i, A_i\}} = \frac{\Phi(\theta' X_i - \bar{Y})}{\Phi(\theta' X_i - \underline{Y})}$$

(d) Write down a likelihood function for the version of model (2) that takes into account the sample selection.

Solution: Using the previous part, one can write down the likelihood for a single observation is

$$\left(\frac{P\{Y_i^* \geq \bar{Y}|S_i, A_i\}}{P\{Y_i^* \geq \underline{Y}|S_i, A_i\}}\right)^{1\{Y_i=1\}} \left(\frac{P\{Y_i^* < \bar{Y}|S_i, A_i\}}{P\{Y_i^* \geq \underline{Y}|S_i, A_i\}}\right)^{1\{Y_i=0\}}$$

The full likelihood in explicit form is

$$\prod_{i=1}^{n} \frac{\left(\Phi(\theta' X_i - \bar{Y})\right)^{1\{Y_i=1\}} \left(1 - \Phi(\theta' X_i - \bar{Y})\right)^{1\{Y_i=0\}}}{\Phi(\theta' X_i - \underline{Y})}$$

Note that it looks almost like the Probit likelihood function. The difference is the denominator that corrects for the sample selection. This additional factor is called a propensity score.

QUESTION 3.

Economists are generally obsessed by markets, including where they emerge, how they function, when and why they fail to function well, and how they shape welfare outcomes. Many important market questions hinge on how markets interact across physical distance and whether they are spatially integrated.

1. Adam Smith argued that "The division of labor is limited by the extent of the market." Although he didn't articulate it explicitly or as an algebraic model, there is nonetheless a model behind this claim.
   (a) Describe two assumptions of the implicit model that supports this claim.
   First, this implicit model assumes that there are limits to the extent of markets – typically due to transportation or other transaction costs that enable markets in different locations to have distinct sizes or characteristics. That is, markets are spatially separate. Second, it assumes that specialization in a particular good or service is characterized by economies of scale – e.g., due to subsistence constraints such that only beyond a given level of demand for a specialized good or service can one more than cover one's subsistence demands. Other assumptions include:
      - Individuals who can specialize do so even at the risk of losing interesting diversity of tasks in their work life.
      - Heterogeneity in resources (or learning by doing) that enable specialization to yield efficiency gains.

   (b) Briefly discuss the welfare implications of this claim.
   This claim implies that expanding a market or integrating separate markets enables greater specialization, which – in Adam Smiths broader work – is essential to expanding economic growth and human welfare.
2. The introduction of mobile phones has had important impacts on markets around the world. These impacts have been especially pronounced in places that previously lacked good communication or transportation infrastructure. Muto and Yamano (2009) study the impact of mobile phone coverage on market participation and other outcomes for farmers in Uganda. Consider key features of the conceptual framework they use in their analysis.

3. CONCEPTUAL FRAMEWORK

Suppose that the farm-gate price of farmer i at time t of commodity j is defined as $p_{itj}^{FG} = p_{tj}^{M} - \gamma_j(I_t)\tau_i^2$. p_{tj}^{M} is the price of commodity j in the nearest market to the household i at time t. τ_i is the distance between the market and farmer i. We assume that $\gamma_j(I_t)$ is the sensitivity of the output price of commodity j with regard to the distance to the market and that the farm-gate price has a quadratic functional form with the distance to the market. As the information, I_t, increases by one unit, $\gamma_j(I_t)$ decreases: $\frac{\partial\gamma_j(I_t)}{\partial I_t} < 0$.

(a) Use the subscripts as they appear on the four right-hand side variables in this simple farm-gate price model to explicitly state the assumptions the authors make.

The market price varies by time and commodity, but not by household.
Gamma is commodity-specific and depends on the amount of information available at a given point in time, but this function is time invariant.
Distance between a farmer and the main market affects farm-gate prices for the farmer in a non-linear way (squared distance).

(b) The authors further assume that

$$\left|\frac{\partial \gamma_P(I_t)}{\partial I_t}\right| > \left|\frac{\partial \gamma_N(I_t)}{\partial I_t}\right|$$

where P denotes a perishable commodity and N denotes a non-perishable commodity. Explain in words what this assumption means and provide a brief justification for it using a specific example that involves the introduction of mobile phones, which increases I_t.

This means that better information has a larger effect on the price sensitivity of perishable commodities to distance than on the price sensitivity of nonperishable commodities to distance. This make sense in the case of mobile phones since better information will enable farmers and traders to better coordinate and thereby reduce spoilage and associated transaction costs – and this effect is larger for perishable commodities at greater risk of spoilage than for nonperishable commodities.

(c) Take the 2nd cross-partial derivative of p_{itj}^{FG} with respect to I_t and then τ_i. Interpret this derivative as specifically as possible in the context of this research setting.

$$\frac{\partial^2 p_{itj}^{FG}}{\partial I_t \partial \tau_i} = -2\frac{\partial \gamma_j(I_t)}{\partial I_t}\tau_{it}$$

This is positive and a linear function of distance to the market, which means that increases in information improve the farm-gate price and that this effect is larger for farmers who are further from the market. Moreover, given 2(b) above, this effect will be more pronounced for perishable commodities than for nonperishables.

(d) Farmers throughout Uganda produce maize and bananas, much of which they consume in their own households. If they choose to sell any of their production in the market, they generally rely on traders. These traders buy from many farmers at p_{itj}^{FG} then deliver the goods to larger markets in the hopes of selling for a profit.

Without access to the mobile phone network, most farmers (especially farmers with limited land holdings (i.e., smallholder farmers)) have no reliable means of communicating with traders, who typically visit farmers in an area unannounced and try to fill their truck as quickly as possible with a specific crop produced by local farmers. This creates a coordination problem that delays the delivery to market and reduces the quality of a perishable good like bananas. In this setting, traders can exploit the fact that they have much better information than farmers.

i. True or False: "The costs associated with this coordination problem are not technically considered transaction costs." Justify your answer. False. Transaction costs include all costs of transacting an exchange between two locations, including the costs of bulking production from many smallholder farmers and the associated spoilage or other losses.

ii. True or False: "Asymmetric information between farmers and traders creates costs that are not technically considered transaction costs." Justify your answer. True. Asymmetric information does change the market outcomes, but not all of these costs are associated with transacting an exchange across space.

(e) In a related article, Jensen (2007) ["The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector" *The Quarterly Journal of Economics* 122(3): 879-924] studies the impact of mobile phones on fishermen and fish markets in India. In contrast to the Ugandan farmers in Muto and Yamano (2009) who do not directly market what they produce, these fishermen both produce (i.e., capture) and market fish.

i. Carefully explain how reliance on traders may change the expected economic impacts of mobile phones on Ugandan farmers. Reliance on traders likely reduces the overall impacts of mobile phones on farmers relative to a situation where the farmers directly marketed their production to consumers in the market. While the benefits to consumers of more efficient information flows between farmers and traders and then between traders and consumers may be comparable, the distribution of these efficiency gains between traders and farmers will favor the party with the most bargaining power and the best information.

ii. Does reliance on traders alter the impacts you would expect to see of expanding the mobile network coverage on those who do not have access to their own mobile phone? Be specific. Yes. In the fisherman case, those without mobile phones still benefited from improved market performance at the time they marketed their fish (i.e., they were much less likely to encounter over-supply of fish in a given market). But in the farmer case, benefiting passively from others' mobile phone access is less likely. Only by communicating via or with those with mobile phones are farmers in this situation likely to benefit from others' phones.

3. Muto and Yamano (2009) use a panel dataset based on two large household surveys conducted in 2003 and 2005 in their analysis. They leverage the fact that the mobile phone network expanded from 46% of the population in 2003 to 70% in 2005 to estimate the impact of mobile phones on banana farmers.

   (a) Their identification strategy relies on the expansion of mobile phone coverage to some new villages between these two survey rounds. Specifically, in one part of their analysis they compare market-level outcomes in these new villages to those in villages that had coverage or lacked coverage in both years.

      i. This identification strategy relies on one key assumption. Explain what this assumption is and why it is central to their strategy. The key identification assumption is that the expansion of phone coverage was random or at lease exogenous to other market forces that are correlated with the market-level outcomes of interest.

      ii. Provide an example of a violation of this assumption that would clearly undermine causal identification in their case. If cell towers were built in anticipation of new agrifood processing investments (e.g., a banana wine production facility) in a given area, then the estimated impacts on market-level outcomes may be attributable to the processing capacity absorbing all local production of the perishable good rather than to the cell towers per se.

   (b) The authors also estimate market participation at the farmer-level using specifications that take as dependent variables the probability that a farmer sells banana and the ratio of sales to production (i.e., the share of total banana production that he sells) and use household fixed effects (FE) to account for time invariant household characteristics. They report their results in Table 7 (we focus here only on the results for Pr(selling banana).

Table 7. *Market participation and mobile coverage/possession: banana*

Variables	Pr(selling banana)			Ratio of sales quantity to production		
	FE	FE	FE-IV	FE	FE	FE-IV
Household mobile phone possession dummy[a]	0.203* (2.35)	0.209* (2.44)	0.151 (0.19)	0.118* (2.39)	0.121* (2.45)	−0.062 (0.14)
Community mobile phone coverage dummy	0.055 (1.57)	−0.095 (1.41)	−0.094 (1.11)	0.054** (2.68)	−0.011 (0.29)	0.000 (0.02)
Distance to district center (miles) × mobile coverage		0.008** (2.62)	0.007** (2.53)		0.003* (1.99)	0.003* (1.81)
Year 2005(=1)	0.481** (17.73)	0.502** (17.84)	0.505** (10.85)	0.174** (11.18)	0.183** (11.31)	0.191** (7.04)
F-stat on IVs			2.58			2.58
# of observations	1,161	1,161	1,151	1,161	1,161	1,151

Note: Numbers in parentheses are absolute *t*-values. The distance to the district center is not included in the models because it is fixed over time.
[a] Instrumental variables: The household mobile possession (HHmobile) is instrumented by the four interaction terms between the mobile coverage dummy and the four household characteristics: log of farm equipments value, age of household head, education of male adult, and education of female adult. These IVs together passed the over-identification test at the 1% significance level.
* Indicates significance at the 5% level.
** Indicates significance at the 1% level.

      i. Carefully compare the results in the first and second column of results in Table 7. Write a concise paragraph that interprets and evaluates these two columns of results. Pay particular attention to what we learn in the context of the research question as we move from the first to the second column. "In the second year of the survey, we find that farmers are generally much more likely to sell bananas than in the first (48-50 percentage point increase in this probability). In addition to this overall trend, we find that owning a mobile phone is associated with a significant 20 percentage point increase in the likelihood that a farmer sells bananas to local traders. Because owning a mobile phone is likely to be endogenous, we are careful not to interpret this as a clean causal effect but as an association. Finally, we see that while community-level access to mobile phone networks does not significantly change this probability of selling banana in general, it does seem to increase this probability on more remote communities. The specification on which these results are based controls for time invariant household characteristics by including household fixed effects, which means the phone possession and mobile coverage coefficients are estimated based on switching households that changed status between 2003 and 2005. "

ii. In column FE-IV, the authors instrument for "Household mobile phone possession dummy" with the instruments described in footnote 'a.' Explain how the "*F*-stat on IVs" is computed. Why does it matter? How does the value of this F-stat help you to interpret these FE-IV results? This F-stat on IVs is computed in the 'first stage' estimation that takes household phone possession as the dependent variable and regresses it on all the independent variables in the second stage (reported in Table 7) along with the instrumental variables described in note 'a'. Specifically, it tests the joint null that the IVs have no effect on household phone possession. This is important in IV estimation as it indicates how much of the variation in the endogenous variable is captured by the (arguably) exogenous variation in the IVs. As a rule, when IVs have a first stage F-stat that is below 10, we worry about them being weak instruments (i.e., we worry that they may introduce more noise into the second stage than they isolate clean exogenous variation). Because this is well below this rule of thumb value, we should interpret the FE-IV results cautiously.