EVIDENCE AGAINST IMPOSING RESTRICTIONS ON HURDLE MODELS AS A TEST FOR SIMULTANEOUS VERSUS SEQUENTIAL DECISION MAKING

WILLIAM J. BURKE

This is a pre-copyedited, author produced PDF of an article accepted for publication in The American Journal of Agricultural Economics *following peer review. The version of record will be available online at:*

http://ajae.oxfordjournals.org

doi: TBD

Abstract

Agricultural economists frequently employ hurdle models to estimate the determinants of truncated outcomes such as market participation and adoption. A pervasive belief is that restrictions can be placed on hurdle models to test whether the decisions made in the underlying data generating process occurred sequentially or simultaneously. This article argues against the ability to draw this conclusion and further submits there is a negative correlation between failing to reject these restrictions and sample size. Evidence to support both proposals comes from data collected in a natural setting as well as simulated data with a known data generating mechanism.

Keywords: adoption; craggit; hurdle models; market participation; sequential decision; simultaneous decision; tobit

JEL Codes: C01 C15 C18 C34 Q12 Q13

Dr. William J. Burke is an economist at Agricultural and Food Policy Consulting in Baltimore, MD. https://www.afpconsulting-burke.com/

email: William@afpconsluting-burke.com

Acknowledgements:

I would like to thank Danielle Nierenberg for helpful comments on early drafts. I am grateful for the comments from three anonymous reviewers whose questions and suggestions improved the work substantially. I particularly appreciate the guidance of the editor, Dr. Timothy Richards. Any errors or omissions are my own.

EVIDENCE AGAINST IMPOSING RESTRICTIONS ON HURDLE MODELS AS A TEST FOR SIMULTANEOUS VERSUS SEQUENTIAL DECISION MAKING

WILLIAM J. BURKE

Agricultural economists frequently employ hurdle models, which are also called corner solution models or multi-stage models. They are useful when data show a "pile up" at some value, usually at a truncation point, over an otherwise continuous distribution of outcomes. The hurdle model framework allows us to estimate the parameters explaining the likelihood that the outcome is not part of the "pile up" as separate from the parameters explaining the expected value of outcomes that are not part of the "pile up". Examples are in market participation or adoption studies where part of the population of interest does not participate or adopt. For this reason, hurdle models are often applied to what some call "zero inflated" data. We can estimate the likelihood that a given observation has a non-zero outcome as a function of different parameters from those determining the expected quantities they will use, buy, or sell, given that the quantity is not zero.

For any hurdle model there is a special case where the underlying processes are the same in each stage of the model, allowing us to simplify the likelihood function and conserve degrees of freedom by estimating fewer parameters. The most common example is one of the double hurdle models (DHM) proposed by Cragg (1971), for which there is a special case that reduces to its very popular predecessor that was introduced by Tobin (1958). These are sometimes called the Craggit and Tobit models, respectively.

Given sufficient data, the more flexible versions of a DHM can be useful because it is unlikely that the underlying parameters are actually identical for most natural settings or data generating mechanisms (DGMs). In market participation models, for example, the most cited reason for allowing flexibility is the fact that theory suggests fixed costs will have countervailing effects on the likelihood of a non-zero outcome and the expected value of a given non-zero outcome. Lin and Schmidt (1984) discuss other scenarios where countervailing effects might be expected. On the other hand, the nested versions of the DHM that assume underlying parameters are the same, can be useful if the underlying mechanisms are sufficiently similar and samples are restrictively small.

These tradeoffs notwithstanding, a certain belief has evolved around the comparison of restricted and unrestricted hurdle models. Specifically, a frequent assertion is that more flexible hurdle models can be restricted into their nested counterparts to test the hypothesis that decisions are made simultaneously against the alternative that they are made sequentially. The argument has been made across a spectrum of agricultural economics archives, including in recent issues of highly ranked peer-reviewed journals (e.g., Bellemare and Barrett 2006; Ricker-Gilbert and Jayne 2009; Muamba 2011; Akpan, Nkanta and Essian 2012; Dehinenet et al. 2014; Fischer and Qaim 2014; Abu, Issahaku and Kwame Nkegbe 2016). Moreover, many publications cite these works as having tested for sequential decision making (e.g., Boughton et al. 2007; Reyes et al. 2012; Miteva et al. 2017; Fan and Salas Garcia 2018), and together these studies have tallied hundreds of citations. In short, confidence in the belief that DHMs can be used to test for sequential decision making is not uncommon.

If the more flexible version is a statistically significantly better fit to the data, these studies argue that decisions about whether and how much to participate/adopt must have been made sequentially. Conversely, the argument goes, decisions must have been made simultaneously if we reject the hypothesis that the more flexible version is a better fit. In this paper I argue that neither conclusion is necessarily correct. To be clear, the question of whether restricted models test the hypothesis that decisions are made simultaneously versus sequentially is the only issue at hand in this present analysis. All of the aforementioned studies have made contributions to our knowledge about correlations and causal relationships that cannot be discounted, irrespective of whether the statistical tests regarding the sequence of decision-making processes were appropriate.

This article will show how no assumptions regarding the order of decisions are made when restricting flexible DHMs into their nested counterparts. I argue the only hypothesis that can be tested when comparing two such estimates is whether the underlying processes are the same or different. Further arguments regarding the simultaneousness or sequence of decisions may reasonably be made by introspection, but, strictly speaking, cannot be considered either consistent or inconsistent with results of a statistical comparison of Craggit and Tobit specifications.

I will further submit that sample size is often a likely determinant of whether the Cragg-like flexible DHMs or Tobin-like restricted DHMs are a statistically better fit. This proposal is examined using randomly chosen sub-samples of data on demand for subsidized fertilizer in Zambia and showing how the likelihood of rejecting Tobin's model in favor of Cragg's more flexible alternative increases with sample size. Finally, since there are reasons we might reject parameter restrictions in the DHM in favor of flexible alternatives that are unrelated to the difference in the underlying coefficient parameters (Wooldridge 2010, page 693), I will present similar results from simulated data with a known DGM that shows the same relationship between sample size and the likelihood of rejecting restriction assumptions. Data and Stata syntax for both examples are provided in online appendices so that this exercise can be replicated with or without any arbitrarily defined alterations.

Conceptual Hurdle Model

Before comparing Craggit and Tobit, it may be useful to begin with a conceptual framework that they share. Hurdle models were introduced largely to address the empirical reality that in many applications we are faced with strictly non-negative data that have a non-trivial number of observations where the variable we are trying to explain is zero (throughout this section I rely heavily on Smith (2002) and chapter 17 of Wooldridge (2010)). For a given application, "corners" or "pile-ups" other than zero may exist, or indeed a multitude of them as in Bellemare and Barrett (2006) or Burke, Myers and Jayne (2015), but for simplicity this discussion will focus on models with one corner at zero. The underlying model has been described where the outcome, y, is the product of an indicator variable that determines whether y is positive (w) and a latent variable (y^*) that is only observed when w=1, such that:

(1) $$y = w \cdot y^*$$

(2) $$w = 1[y > 0]$$

From this, it is a logical extension to see why there may be two equations of interest when investigating how other factors ($\boldsymbol{x}$) determine *y* – one for each component of the observed outcome. Because y^* is not always observable, it can be helpful when implementing a DHM to make the assumptions that, conditional on $\boldsymbol{x}$, the variables *w* and y^* are independent, so that we can say the following about the distribution (*D*):

(3) $D(y^*|\boldsymbol{x}, w) = D(y^*|\boldsymbol{x})$

The reason this is helpful is that it implies the following is also true about the conditional expectations of y^* and *y*:

(4) $E(y^*|\boldsymbol{x}, w) = E(y^*|\boldsymbol{x}) \;\therefore\; E(y|\boldsymbol{x}, w) = w \cdot E(y^*|\boldsymbol{x}, w) = w \cdot E(y^*|\boldsymbol{x})$

Therefore, where we are able to observe y^*, that is when $w = 1$, we can say:

(5) $E(y|\boldsymbol{x}, w = 1) = E(y^*|\boldsymbol{x})$

In words, the assumption in equation (3) allows us to say we can estimate how determinants affect the latent variable y^* by specifying a model for its expected value and using only the positive outcomes for y^* that we are able to observe. Moreover, we can separately estimate parts of the model for the probability that $w = 1$ and the expected value of *y* given that it is positive. With these results in hand, we can also easily estimate the expected value of *y* for any observation (i.e., *not* conditional on *y* being positive) as the product of the estimated probability that $w = 1$ and the estimated expected value of y^*.

As a brief digression, it is worth pointing out that much can be said about what it implies to assume *w* and y^* are conditionally independent. However, that discussion is not germane to the present question regarding whether outcomes are determined simultaneously or sequentially. The only benefit the conditional independence assumption is that it easily motivates the separate specification of equations relating determinants to the outcome and, under this and other usual assumptions, separate estimations are unbiased and consistent. As experienced practitioners may know, Tobin's model does not require separability, and correspondingly does not require the assumption of conditional independence. This is not, however, a difference between Tobin's model and more flexible alternatives generally because flexible alternatives can also relax the assumption in equation 3.[1]

The Difference Between Craggit and Tobit

The key difference between Tobit and Craggit, rather, only occurs when proceeding from the underlying framework to estimation. Although Cragg's estimation procedure was introduced later than Tobin's, it is more logical for the present purpose to begin with Cragg. Specifically, Cragg operationalizes the framework such that *w* follows the probit model and, to be consistent

---

[1] For example, one method of doing so applied to Cragg's Tobit alternative requires a two-stage procedure and an adjustment to standard error estimates that results in a procedure that is operationally identical to Heckman's (1979) method of controlling for selection bias (Lin and Schmidt 1984). Similar methods have been applied to other double hurdles and three stage models that have Cragg-like flexibility (e.g., Bellemare and Barrett 2006; Burke, Myers and Jayne 2015).

with the fact that the support of $y^*$is strictly positive, y^* follows the truncated normal model. In notation form this can be written as:

(6) $P(w = 1|\boldsymbol{x}) = \Phi(\boldsymbol{x\gamma})$

(7) $y^* = \boldsymbol{x\beta} + u$

(8) $D(y|\boldsymbol{x}, w = 1) \sim Normal(\boldsymbol{x\beta}, \sigma^2)$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $(u|\boldsymbol{x})$ has a truncated normal distribution with a lower limit of $-\boldsymbol{x\beta}$ (predicted values of y are strictly positive). Conveniently, the $\boldsymbol{\gamma}$ parameters can be estimated with a probit regression of w on $\boldsymbol{x}$, and the $\boldsymbol{\beta}$ and σ parameters can be estimated with a truncated normal regression of y on $\boldsymbol{x}$ using only those observations for which $y > 0$. This proposal for estimating the DHM is succinctly encompassed in the likelihood function:

(9) $$f(w, y|\boldsymbol{x}) = [1 - \Phi(\boldsymbol{x\gamma})]^{1[w=0]} \left[\Phi(\boldsymbol{x\gamma}) \frac{(2\pi)^{-\frac{1}{2}}\sigma^{-1}\exp\left(-(y-\boldsymbol{x\beta})^2/2\sigma^2\right)}{\Phi(\boldsymbol{x\beta}/\sigma)}\right]^{1[w=1]}$$

All parameters could also be estimated simultaneously via maximum likelihood, as described in Burke (2009). Results will be the same whether estimating parameters simultaneously or with separate maximum likelihood regressions.

The key point of this discussion is that if we substitute $\boldsymbol{x\beta}/\sigma$ for $\boldsymbol{x\gamma}$ in equation (9), the likelihood function is identical to Tobin's. In other words, the only assumption necessary to move from Cragg's proposal to Tobin's original proposal is that $\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$. By imposing this assumption there are nearly half as many parameters to estimate. Imposing this assumption and relaxing the assumption in equation (3) would mean estimating a flexible alternative such as Cragg's may require three times as many parameter estimates. Depending on the ratio of the number of parameters to the number of observations available, imposing Tobin's assumption could grant useful degrees of freedom. On the other hand, as described earlier, Tobin's assumption could be inconsistent with the theoretical underpinnings that motivate including certain variables in the model.

Importantly, the assumptions under which Cragg's estimate of $\boldsymbol{\gamma}$ is unbiased and consistent do not include any assumption regarding the actual value of the $\boldsymbol{\gamma}$ parameters. It follows that even if Tobin's assumption holds, Cragg's estimation procedure will provide unbiased and consistent estimates, though Cragg's estimator would be inefficient. Since Cragg's procedure also provides unbiased and consistent estimates under Tobin's assumptions, there can be no implication in Tobin's assumptions regarding the sequence of decisions that lead to the observed outcome that is not also true when the restriction is not imposed. Therefore, whether or not a statistical test leads to rejecting the restriction assumptions cannot, in itself, imply whether decisions are sequential or simultaneous.

Empirical Comparisons of Craggit and Tobit

Since imposing Tobin's assumption results in a special case of Cragg's model, it is relatively easy to compare the two using a likelihood ratio test where Craggit is the unrestricted model, Tobit is the restricted model, and the number of restrictions is equal to the number of explanatory variables (plus one if there is an intercept). As usual, the null hypothesis is that the restricted Tobit is "correct" versus the alternative that the relatively flexible Craggit is a statistically significantly better fit.

For example, if we consider a simple version of the model in equations 6-8 where there is just one explanatory variable, so that $P(w = 1|x) = \Phi(\gamma_0 + \gamma_1 x)$ and $y^* = \beta_0 + \beta_1 x + u$, testing the null hypothesis that Tobit is a better fit against the Craggit alternative can be written as $H_0\colon \gamma_0 = \beta_0/\sigma\,; \gamma_1 = \beta_1/\sigma$ versus the alternative that at least one of the null equalities does not hold. The likelihood ratio test statistic (*LR*) is twice the difference between the log-likelihood of the unrestricted model $(L_{Craggit})$ and the restricted model (L_{Tobit}), or $LR = 2(L_{Craggit} - L_{Tobit})$. *LR* is asymptotically χ^2_Kdistributed, where *K* is the difference in the number of parameters estimated in the unrestricted and restricted models, or 2 in this example. Any statistical package that estimates Tobit will provide L_{Tobit} in the estimation results. In Stata, $L_{Craggit}$ is provided after employing the "craggit" command described in Burke (2009). Alternatively, $L_{Craggit}$ is equal to the sum of the log-likelihoods from separate probit and truncated normal regressions described above. Executing this likelihood ratio test can be done straightforwardly following the examples in this article's online appendices. If we reject the null hypothesis (i.e., if the p-value from the likelihood ratio test is considered meaningfully "low"), we would conclude Craggit is preferred to Tobit.

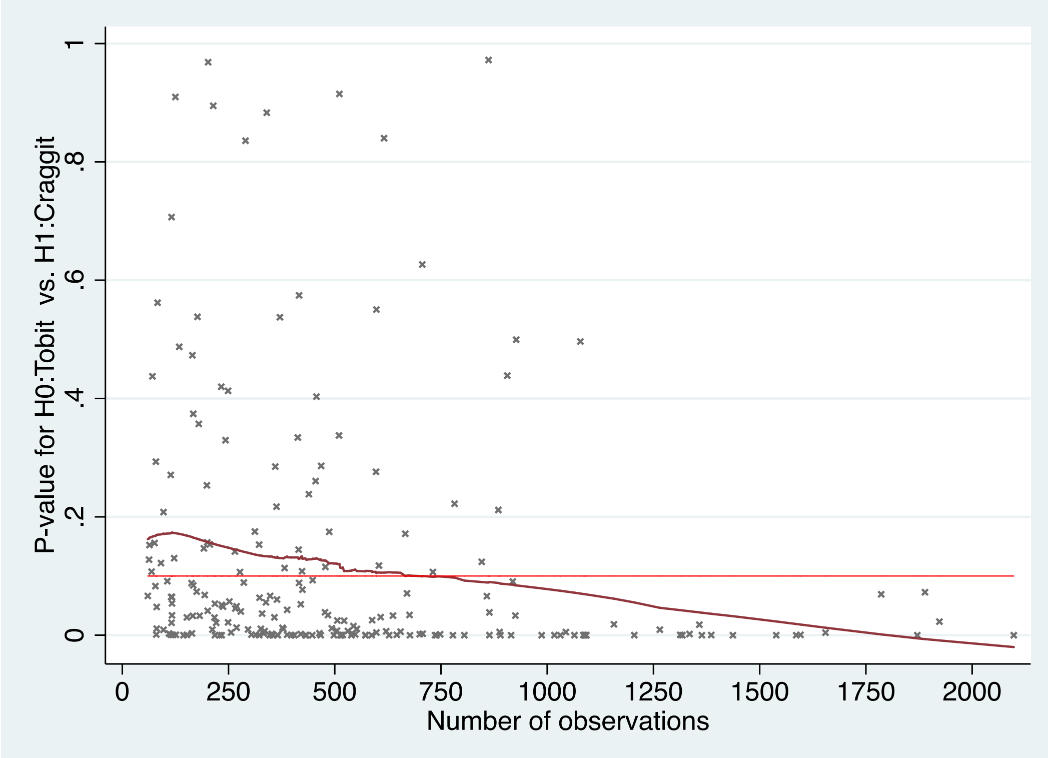
Barring the cases where Tobin's assumption actually holds for the DGM, whether we reject the null hypothesis is most likely determined by three factors: 1) how similar the underlying processes truly are, 2) how much of the variation in the outcome can be explained (compared to how much variation is determined stochastically), and 3) how big our datasets are.[2]

To the first point, the closer the true DGM is to matching Tobin's assumption that $\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$, the more likely we would be to fail to reject the Tobit hypothesis for a given sample size and level of stochasticity. In the simple, one determinant model, the closer $(\gamma_0\sigma - \beta_0)$ and $(\gamma_1\sigma - \beta_1)$ are to zero, the less likely we are to reject the null hypothesis of Tobin's assumption.

To the second point, for any given sample size, $\boldsymbol{\gamma}$**,** and $\boldsymbol{\beta}$ values in the DGM, the degree of stochasticity in the underlying model, σ, and the likelihood of rejecting Tobin's assumptions are negatively correlated. In other words, the less variation in the outcome that can be explained, the less likely we are to reject Tobin's assumptions. This is because, if all else is equal, a higher degree of stochasticity in the DGM leads to less precision in parameter estimates. Less precision means the confidence intervals around the $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}/\sigma$ estimates would be wider and more likely to overlap, in which case we would fail to reject Tobin's assumptions. These first two arguments

[2] Related to the first factor, if n_0 is the number of observations where $y = 0$, it is also the case that as $n_0 \to 0$, the likelihood of rejecting Tobin's hypothesis vs. Cragg's alternative decreases for any given values of $\beta, \gamma,$ and σ. This is because as $n_0 \to 0$ both Tobit and Craggit approach OLS estimates of regressing *y* on *x*, and hence become more similar to each other. This is demonstrated in online Appendix E.

Figure 1. P-values for Likelihood Ratio Tests of Tobin's Assumption vs. Cragg's Alternative Using Data on Subsidized Fertilizer Purchases in Zambia (2000-2003)



Source: Zambia Post Harvest and Supplemental surveys 1999/2000 & 2002/2003.
Notes: Hypothesis testing done using 215 random draws of between 60 and 2100 observations, each represented by an "x". The horizontal reference line indicates 10% significance level for the null hypothesis test that $\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$. The curved line is a LOWESS regression of p-values on the value of N.

can be demonstrated using simulated data with known DGMs.[3] In any natural setting, however, the true coefficients and level of stochasticity are unobservable. The third point regarding the relationship between sample size and the likelihood of rejecting Tobin's hypothesis, on the other hand, can be discussed in the context of "real" data.

Holding the first two factors constant (and assuming $\boldsymbol{\gamma} \neq \boldsymbol{\beta}/\sigma$), larger datasets are more likely to reject the hypothesis that Tobin's assumptions hold. Within a given dataset the first two factors are indeed constant (albeit unknown), which is why we can examine this proposition empirically. To do so I will be using the dataset that are publicly available and the model described by Burke (2009).[4]

The data are from a panel of nationally representative smallholders collected during the Post-Harvest and Supplemental Surveys of 1999/2000, and 2002/2003 conducted by the Central

[3] Online Appendix C demonstrates the relationship between the "closeness" of Tobit to the true DGM and the likelihood of failing to reject Tobin's assumption. Online Appendix D demonstrates the relationship between the level of stochasticity and the likelihood of failing to reject Tobin's assumptions.
[4] The data can be downloaded directly through Stata by entering "findit craggit" in the command window, then clicking on the "st0179" and getting the ancillary files.

Statistics Office of the Government of Zambia. Because the data are proprietary and were collected under conditions of confidentiality, identifying information such as household numbers, district indicators, the survey year and so on are not available. These data are not very useful, therefore, for any substantive policy analysis, but are sufficiently fit for the present purpose. The variables we can use are an indicator for whether subsidized fertilizer was purchased (13.2% of 6,378 observations), the quantity acquired (179 kgs on average amongst those who purchased), the distance to the nearest district capital (33.7 km on average), the quantity of cultivated land (2 ha on average), and the age and education of the household head (46.3 years and 5.2 years on average respectively).

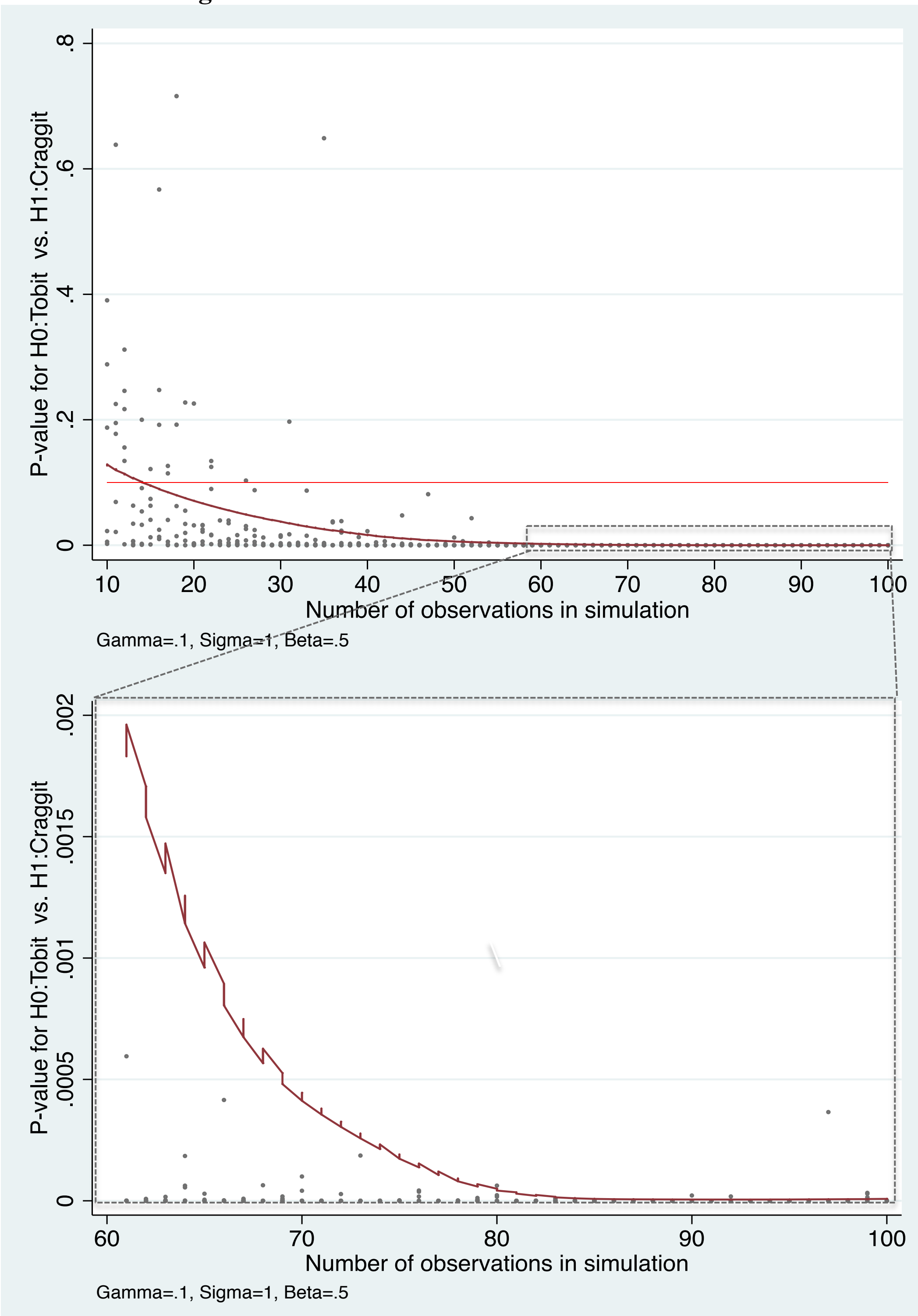
To conduct this examination, figure 1 is a scatter plot and LOWESS regression of the p-values from likelihood ratio tests of the null hypothesis that Tobin's assumption holds on the size of random sub-samples. Two-hundred and fifteen (215) random draws of between 60 and 2,100 observations are used to generate figure 1.[5] Both the scatter plot and the nonparametric trend show a noticeable negative relationship between p-values and sample size that is consistent with the proposal laid out earlier. Simple linear regressions (not shown) of either the p-value (p) or $ln\ (p)$ on either N or ln (N) also predict negative and statistically significant trends (statistical significance of each trend is 0.011 or lower).

Of course, the relationship between sample size and likelihood of rejecting Tobit can also be examined using simulated data. For example, let us suppose x is one variable, and $x{\sim}N(5, 5^2)$, $\beta = 0.5$, $\sigma = 1$, $\gamma = 0.1$, and truncated normal stochastic variation is added to $x\beta$ (and the intercept is 0). In other words, the simulated DGM follows a Craggit specification. Figure 2 shows the p-values from likelihood ratio tests using 546 random draws for the null hypothesis that Tobin's assumption holds versus Cragg's alternative changing as the sample size increases from 10 observations to 100 observations. The top panel shows the trend predicted earlier, where we would have failed to reject the null hypothesis for several of the smaller samples, but we rejected it for larger samples (at the 10% significance level, for example, which is indicated by the horizontal reference line). The fractal nature of this pattern is demonstrated in the bottom panel of the figure, which show an enlargement of the region highlighted in the top panel.[6]

One could argue that the DGM for these samples is indeed sequential because it is in fact the result of lines of code that are organized in a sequence. On the other hand, it is trivial to alter the simulation parameters arbitrarily,[7] meaning that for all intents and purposes the outcomes may have well been determined simultaneously. However, whether this DGM is actually sequential or simultaneous is immaterial. The important takeaway is that the DGM is exactly the same for every sample whose results are shown in figure 2, yet for smaller samples we are more likely to fail to reject Tobin's assumption, irrespective of the order in which the dependent variables are generated.

[5] Syntax to replicate results using Stata are in online Appendix A.
[6] Note, if the DGM indeed follows the specification assumed by Tobin, the likelihood of rejecting Tobin's hypothesis is unrelated to sample size, which can be seen in the results of Appendix C.
[7] All of the code used to run this simulation using Stata is available in online Appendix B and requires no actual data to start with.

Figure 2. P-values for Likelihood Ratio Tests of Tobin's Assumption vs. Cragg's Alternative Using Simulated Data as N increases



Source: the author and data simulated using Stata code in online Appendix B.
Notes: Hypothesis testing done using 546 random draws of between 10 and 100 observations, each represented by a "dot". The horizontal reference line indicates 10% significance level for the null hypothesis test that $\gamma = \beta/\sigma$. The curved line is a LOWESS regression of p-values on the value of N.

Conclusion

The objective of this article is to highlight and present evidence against a prevalent belief surrounding a popular model in agricultural economics: that we can restrict hurdle models to explicitly test the null hypothesis that decisions are made simultaneously versus the alternative that decisions are made sequentially. I argue the theoretical underpinnings of the model do not support this belief. Either a Tobin-like restricted model or a more flexible Cragg-like alternative can be used to estimate corner solution models, whether the data used to estimate them are generated from a process that is either sequential or simultaneous, and thus both may still be useful. Comparing these models, however, cannot help us understand whether the DGM is simultaneous or sequential. Conceptually, this is evidenced by the fact that models with Cragg-like flexibility are inefficient but unbiased and consistent even when Tobin-like restriction assumptions hold. Therefore, anything that is true about the sequence of the DGM under Tobin's assumptions can also be true under Cragg's and vice versa.

In this article, I further argue that sample size is negatively correlated with the likelihood of failing to reject Tobin's assumptions irrespective of the chronology of decisions in the DGM. This proposition is examined empirically using data from Zambian smallholders' purchases of subsidized fertilizer as well as with simulated data with a known DGM. In both cases the evidence is highly consistent with the argument. Of course, this trend would not hold (in fact the opposite would be true) if Tobin's assumptions were actually and completely correct. Moreover, there may be applications where some combination of a "small" sample size, a DGM that is "close" to Tobin's assumptions, and/or a "high" degree of stochasticity such that tests will fail to reject Tobin's assumptions even when they are not completely correct. In other words, there may be situations where Tobit is a better fit for a given data set and model. However, since "small", "close" and "high" are all relative terms, no rule-of-thumb can tell us whether a test to compare the two specifications is not statistically necessary.

That said, it does seem very unlikely that Tobin's assumptions could ever be completely correct for describing any relationship in nature. To my knowledge, there is no obvious case wherein the coefficients determining the probability of a positive outcome, $\boldsymbol{\gamma}$, should be expected to be the specific and uniform function of the coefficients determining the expected value of a given positive outcome, $\boldsymbol{\beta}$, and the standard deviation of the stochastic residual, σ, proposed by Tobin. In the absence of any natural context, Tobin's assumption that $(\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma)$ as an economic concept seems, to me, very specific and somewhat arbitrary. It seems more like, as it was in 1958, an assumption based on statistical convenience, not behavioral principles. Therefore, barring applications for which underlying theory actually suggests Tobin's very restrictive assumptions should hold, one might regard the likelihood ratio test comparing Craggit and Tobit as more of a test for whether data limitations favor estimating a Tobit model as opposed to the alternative of estimating Cragg's more flexible model. Most importantly, regardless of which model is the better fit, the comparison of Craggit and Tobit does not add to our understanding of the chronology of decision making.

It is worth taking a moment to highlight what appears to be a common misunderstanding of how DHM results should be interpreted. The claims that a hurdle model has been used to test for sequential decision making are not rare. Moreover, a model that properly tests for simultaneous versus sequential decision making could be a useful analytical tool. Bellemare and

Barrett (2006) convincingly describe a scenario where the difference can have important economic policy implications. Unfortunately, it does not seem that DHMs as they are most often employed are fit for this purpose. Developing more appropriate models to test for simultaneous decision making is likely to occur more slowly as long as there is not consensus around whether the current efforts do so.

References

Abu, B.M., H. Issahaku and P.K. Nkegbe. 2016. Farmgate Versus Market Centre Sales: a Mulit-Crop Approach. *Agricultural and Food Economics* 4(21): 1-16

Akpan, S.B., V. S. Nkanta and U.A. Essien. 2012. A Double-Hurdle Model of Fertilizer Adoption and Optimum Use Among Farmers in Southern Nigeria. *Tropicultura* 30(4): 249-253.

Bellemare, M.F. and C.B. Barrett. 2006. An Ordered Tobit Model of Market Participation: Evidence from Kenya and Ethiopia. *American Journal of Agricultural Economics* 88(2): 324-337.

Boughton, D., D. Mather, C.B. Barrett, R. Benfica, D. Abdula, D. Tschirley and B. Cunguara. 2007. Market Participation by Rural Households in a Low-Income Country: An Asset-Based Approach Applied to Mozambique. *Faith and Economics* 50(Fall): 64-101.

Burke, W.J. 2009. Fitting and Interpreting Cragg's Tobit Alternative Using Stata. *Stata Journal* 9(4): 584-592

Burke, W.J., R.J. Myers and T.S. Jayne. 2015. A Triple-Hurdle Model of Production and Market Participation in Kenya's Dairy Market. *American Journal of Agricultural Economics* 97(4): 1227-1246.

Cragg, J.G. 1971. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica* 39: 829-844.

Dehinenet, G., H. Mekonnen, M. Kidoido, M. Ashenafi and E. Guerne Bleich. 2014. Factors Influencing Adoption of Dairy Technology on Small Holder Dairy Farmers in Selected Zones of Amhara and Oromia National Regional States, Ethiopia. *Discourse Journal of Agriculture and Food Sciences* 2(5): 126-135.

Fan, Q and V.B. Salas Garcia. 2018. Information Access and Smallholder Farmers' Market Participation in Peru. *Journal of Agricultural Economics* 69(2): 476-494.

Fischer, E. and M. Qaim. 2014. Smallholder Farmers and Collective Action: What Determines the Intensity of Participation? *Journal of Agricultural Economics* 65(3): 683-702.

Heckman, J.J. 1979. Sample Selection Bias as Specification Error. *Econometrica* 47: 931-959

Lin, Tsai-Fen and P. Schmidt. 1984. A Test of the Tobit Specification Against an Alternative Suggested by Cragg. *The Review of Economics and Statistics* 66(1): 174-177

Miteva, D.A., R.A. Kramer, Z.S. Brown and M.D. Smith. 2017. Spatial Patterns of Market Participation and Resource Extraction: Fuelwood Collection in Northern Uganda. *American Journal of Agricultural Economics* 99(4): 1008-1026.

Muamba, F.M. 2011. Selling at the Farmgate or Traveling ot the Market: A Conditional Farm-Level Model. *The Journal of Developing Areas* 44(2): 95-107.

Reyes, B., C. Donovan, R. Bernsten and M. Maredia. 2012. "Market Participation and Sale of Potatoes by Smallholder Farmers in the Central Highlands of Angola: A Double Hurdle Approach." Paper presented at the International Association of Agriculturl Economists Triennial Conference, Foz do Iguacu, Brazil, 16-22 August.

Ricker-Gilbert, J. and T.S. Jayne. "Do Fertilizer Subsidies Affect Demand for Commercial Fertilzer? An Example from Malawi." Paper presented at the International Association of Agriculturl Economists Triennial Conference, Beijing, China, 18-24 August.

Smith, M.D. 2002 "On Specifying Double-Hurdle Models." In A. Ullah, A.T.K. Wan and A. Chaturvedi, eds. *Handbook of Applied Econometrics and Statistical Inference*, Marcel Dekker, New York, USA.

Tobin, J. 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica* 26: 24-36.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data.* 2nd ed. Cambridge, MA: MIT Press.

AJAE Appendices for "Evidence Against Imposing Restrictions on Hurdle Models as a Test for Simultaneous vs. Sequential Decision Making"

William J. Burke
March 25, 2019

Note: The material contained herein is supplementary to the article named in the title and published in *The American Journal of Agricultural Economics* (*AJAE*).

Appendix A: Stata code for duplicating results in Figure 1

```
use "zam_fert_ex.dta", clear
mat LRgraph=(.,.,.)
set seed 2019
local sel=.01
local sel2=0
local stop=.4
local group=1

while `sel'<=`stop' {
preserve
       gen sel=runiform(0,1)
       quietly keep if sel<`sel' & sel>`sel2'
       quietly sum basal_g
       if r(mean)<0.15 | r(mean)>0.75          {
              local sel=`sel'+.0001
              restore
                                               }
       else                                    {
       quietly tobit qbasal_g disttown cland educ age, ll
       scalar Lr=e(ll)                                              /*Restricted log-likelihood*/
       scalar k=e(k)-1          /*Number of restrictions in the LR test (minus 1 because of sigma)*/
       quietly probit basal_g disttown cland educ age
       scalar Lu1=e(ll)                                /*Unrestricted log-likelihood from first hurdle*/
       quietly truncreg qbasal_g disttown cland educ age, ll(0) iter(10)
       scalar Lu2=e(ll)                               /*Unrestricted log-likelihood from second hurdle*/
       scalar LR=2*(Lu1+Lu2-Lr)    /*Compute LR stat (note the unrestricted ll is sum of Lu1, Lu2)*/
       scalar pval=chi2tail(k,LR)        /*Recall k is the number of parameters in tobit, minus 1)*/
       scalar lnp=ln(pval)    /*This is only necessary if you want to keep ln(p) in its own column*/
       quietly mat lrgraph=([_N] , lnp, pval)
       quietly mat LRgraph=LRgraph\lrgraph
restore
di "completeded for sel=`sel'"        /*Keeps you updated on progress while running - OK to block*/
local sel=`sel'+.0001
       }
       }
/*
  Note that for these "seed", "sel" and "stop" parameters the code will
  run through to the end.  If you change any these initial values
  you may encounter an iteration that fails to converge, which will
  stop the "while" loop.  At that point you could skip to the
  "svmat" command and proceed (the completed iterations will have
  stored results), or you could try to change the seed or increase the
 skip criterion on line 18 (for example to 0.15) and start over.
*/
clear
svmat LRgraph
rename LRgraph1 N
rename LRgraph2 lnp
rename LRgraph3 pval
lab var pv "P-value for H0:Tobit  vs. H1:Craggit"
label variable N "Number of observations"
lowess pv N, bw(.8) m(x) msize(small) gen(lowess)
gen refline=0.1
graph twoway  (scatter pv N, m(x) xlabel(0 (250) 2000) mc(gs7))                 ///
                      (line lowess N, sort legend(off)                          ///
                      ytitle("P-value for H0:Tobit  vs. H1:Craggit"))           ///
                      (line refline N, lw(vthin) lcolor(red))
reg p N
reg lnp N
gen lnN=ln(N)
reg lnp lnN
reg p lnN
```

Appendix B: Stata code for duplicating results in Figure 2

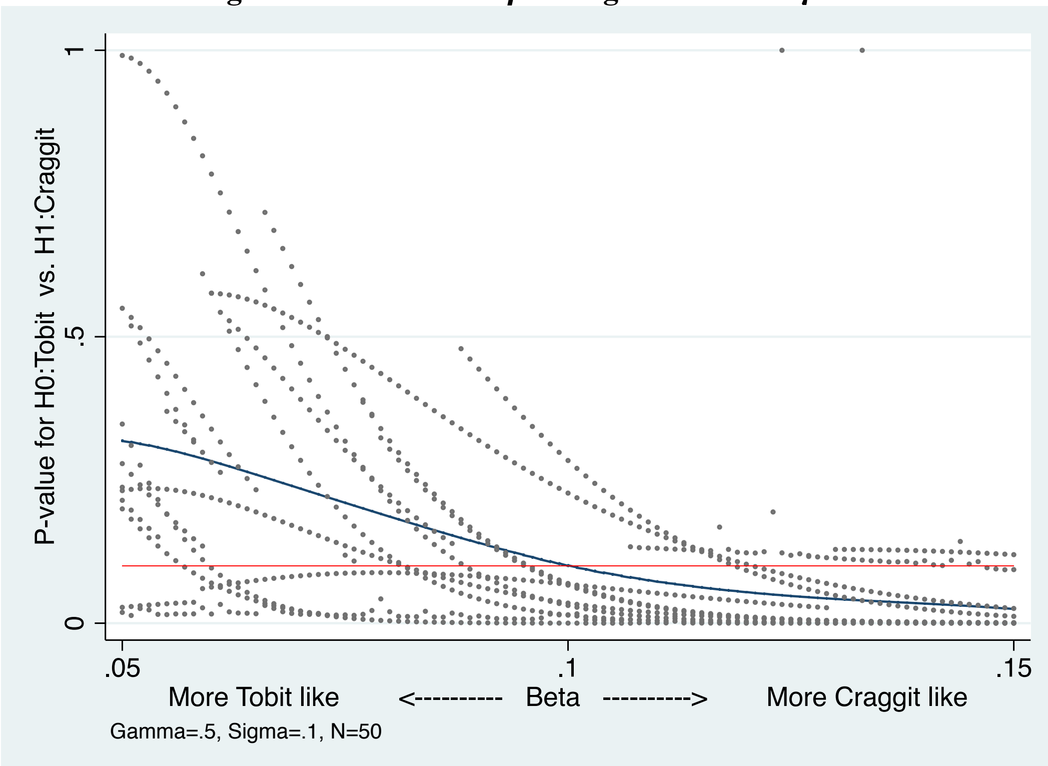
```
clear all
mat LRgraph=(.,.,.)
local b=.5                              /*Establish parameters*/
local s=1
local g=.1
global g=`g'                            /*Also making global macros to use in graphs*/
global s=`s'
global b=`b'
local obs=10                            /*Initiate N for simulations*/
local end=100
local seed=1978                         /*Set to duplicate results – change to show robustness*/
/*Generate simulated data for several “groups” (each group starts over with small
  samples of 10), estimate and compare tobit & craggit models over an increasing N           */
local group=1
while `group' <=6               {
while `obs'<=`end'              {
       clear
       quietly set obs `obs'
       quietly set seed `seed'
       quietly generate x=rnormal(5,5)                        /*Generate the determinant variable */
       quietly generate pr=normal(x*`g')                              /*Generate probability y>0 */
       quietly generate ystar=x*`b'+max(-x*`b',rnormal(0,`s')) /*Generate “latent” variable       */
       quietly gen y1=runiform(0,1)            /*These 3 lines assure y is ystar with probability */
       quietly gen y=ystar if pr>y1 & ystar>0  /*equal to normal(xg), given that ystar is also >0 */
       quietly replace y=0 if y==.             /*and equal to zero otherwise                      */
       quietly tobit y x , ll(0)
       quietly estimates store tobit
       quietly craggit y x , sec(y x)        /*Unlike Appendix A, this shows how the unrestricted */
       quietly estimates store craggit       /*log-likelihood can come from one command, -craggit-*/
       quietly lrtest tobit craggit, force
       quietly mat lrgraph=([_N],r(p),`group')
       quietly mat LRgraph=LRgraph\lrgraph
       di "Completed N is `obs' of `end'”           /*Keeps you updated on progress- OK to block*/
       local obs=`obs'+1
       local seed=`seed'+39
       }
       local obs=10
       local group=`group'+1
       }
svmat LRgraph
rename LRgraph1 N
rename LRgraph2 Pval
lab var Pval "P-value for H0:Tobit  vs. H1:Craggit"
lab var N "Number of observations in simulation"
lowess Pval N, gen(lowess)
gen refline=0.1
graph twoway  (scatter Pval N, m(oh) msize(vsmall) mc(gs7)                                 ///
                      note("Gamma=$g, Sigma=$s, Beta=$b") legend(off))                     ///
                      (line lowess N, sort ytit("P-value for H0:Tobit  vs. H1:Craggit"))   ///
                      (line refline N, lw(vvthin) lco(red) xscale(r(10 100))               ///
                      xlabel(10 (10) 100) name(dN))

graph twoway  (scatter Pval N if N>60, m(oh) msize(vsmall) mc(gs7)                         ///
                      note("Gamma=$g, Sigma=$s, Beta=$b")  legend(off))                    ///
                      (line lowess N if N>60, sort ytit("P-value for H0:Tobit  vs. H1:Craggit") ///
                       xscale(r(60 100)) xlabel(60 (10) 100) name(dN60))

/*The line “local seed=1978” is necessary to replicate results.
 To change it to something random, replace this line with
 “local seed=abs(int(rnormal(0,100)))” and re-run this code
 to see the robustness of the conclusion that p-values tend
 to decrease as N increases without changing the DGM.  */
```

Appendix C: Stata code to show how LR tests change as β changes relative to γ

```
clear all
mat LRgraph=(.,.,.)
/*Establish parameters that will not change*/
local g=.5
local s=.1
local obs=50
global g=`g' /*Also using global macros for the purpose of generating graphs*/
global s=`s'
global obs=`obs'

/* Initiate Beta parameters "close" to tobit*/
local Db=0
local end=1
local seed=1978
local group=1
*local seed=abs(int(rnormal(0,100)))
/*Generate simulated data for several “groups” (each group starts over with parameters
  that are “close” to tobit), estimate and compare tobit & craggit models over changing Beta*/

while `group'<=10                {
di "Working on group `group'"                   /*Keeps you updated on progress- OK to block*/
while `Db'<=`end'                {
       clear
       quietly set obs `obs'
       quietly set seed `seed'
       quietly generate x=rnormal(5,5)
       local b=(`g'+`Db')*`s'
       quietly generate pr=normal(x* `g')
       quietly generate ystar=x*`b'+max(-x*`b',rnormal(0,`s'))
       quietly gen y1=runiform(0,1)
       quietly gen y=ystar if pr>y1 & ystar>0
       quietly replace y=0 if y==.
       quietly generate w=0
       quietly replace w=1 if y>0
       quietly tobit y x , ll(0)
       quietly estimates store tobit
       quietly craggit y x , sec(y x) iter(10)
       quietly estimates store craggit
       quietly lrtest tobit craggit, force
       quietly mat lrgraph=(`b',r(p),`group')
       quietly mat LRgraph=LRgraph\lrgraph
       local Db=`Db'+.01                        /*Changes Beta for next round of simulation*/
                                               }
local Db=0                             /*So that next “group” will be begin “close” to tobit*/
local seed=`seed'+39                  /*So that next “group” will be different than previous*/
di "Finished group `group'"                    /*Keeps you updated on progress- OK to block*/
local group=`group'+1
}
svmat LRgraph
rename LRgraph1 b
rename LRgraph2 Pval
lab var Pval "P-value for H0:Tobit  vs. H1:Craggit"
lab var b "More Tobit like        <----------   Beta   ---------->        More Craggit like"
lowess Pv b, gen(lowess)
sum P
global pmax=r(max)
gen refline=0.1
graph twoway (line lowess b, sort ytit("P-value for H0:Tobit  vs. H1:Craggit"))   ///
                     (scatter Pval b,msize(vsmall)   m(oh) mc(gs7)                ///
                     ylabel(0 ( `=round(($pmax)/2, .01)' ) `=round(($pmax),.01)') ///
                     note("Gamma=$g, Sigma=$s, N=$obs"))                          ///
                     (line refline b, lw(vvthin) lco(red) legend(off))
```

```
/*The line "local seed=1978" is necessary to replicate results. Change it to something random,
 by replacing this line with "local seed=abs(int(rnormal(0,100)))" and re-run this code many
 times to see the robustness of the conclusion (p-values tend to decrease as B moves farther from
 G*s */
```

Figure C1. P-values for Likelihood Ratio Tests of Tobin's Assumption vs. Cragg's Alternative Using Simulated Data as β changes relative to γ



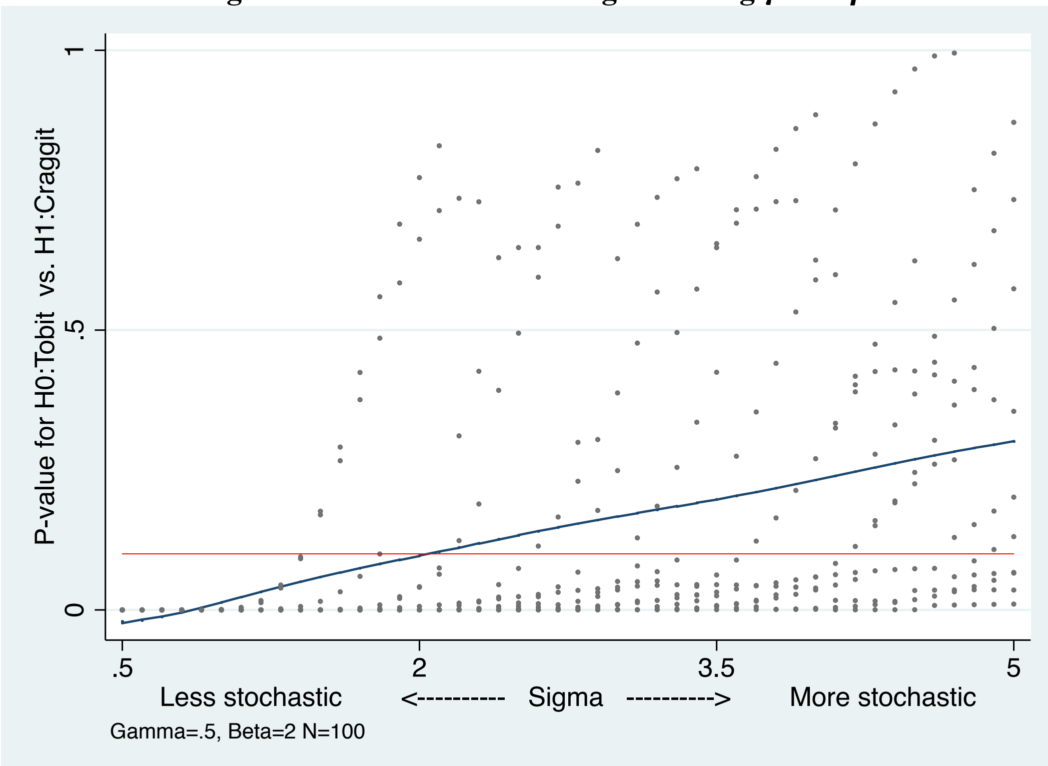
Source: the author and data simulated using Stata code in online Appendix C.
Notes: The reference line indicates 10% significance level for the null hypothesis test that $\gamma = \beta/\sigma$. Each marker on this graph indicates a unique dataset where $x \sim N(5, 5^2)$; $\Pr(y > 0|x) = \Phi(x\gamma)$ and $E(y|y > 0, x) = x\beta + u$; $u \sim N(0, \sigma^2)$ with a lower truncation point at $-x\beta$. The solid curved line is a LOWESS regression of p-values on the value of β.

Appendix D: Stata code to show how LR tests change as σ changes holding γ and β constant

```
clear all
mat LRgraph=(.,.,.,.)
local g=.5            /*Establish initial parameters*/
local s=.5
local b=2
local obs=100
global g=`g'          /*Also using global macros for the purpose of generating graphs*/
global obs=`obs'
global b=`b'
/* Initiate parameters with "low" stochasticity*/
local Ds=0
local end=4.5
local seed=1978
*local seed=abs(int(rnormal(0,100)))
local group=1
/*Generate simulated data for several "groups" (each group starts over with parameters
  that have "low" stochasticity), estimate and compare tobit & craggit models over changing sigma*/

while `group'<=10                {
di "Working on group `group'"                   /*Keeps you updated on progress- OK to block*/
while `Ds'<=`end'               {
       clear
       quietly set obs `obs'
       quietly set seed `seed'
       quietly generate x=rnormal(5,5)
       quietly generate pr=normal(x* `g')
       local S=`s' + `Ds'
       quietly generate ystar=x*`b'+rnormal(0,`S')
       quietly gen y1=runiform(0,1)            /*These 3 lines assure y is ystar with probability */
       quietly gen y=ystar if pr>y1 & ystar>0  /*equal to normal(xg), given that ystar is also >0 */
       quietly replace y=0 if y==.             /*and equal to zero otherwise                      */
       quietly tobit y x , ll(0)
       quietly estimates store tobit
       quietly craggit y x , sec(y x) iter(10)
       quietly estimates store craggit
       *Generate a Pseudo-R^2 - This is for interest only, not used in the graph
       quietly predict xg, eq(Tier1)
       quietly predict xb, eq(Tier2)
       quietly predict sig, eq(sigma)
       quietly gen IMR=normalden(xb/sig)/normal(xb/sig)
       quietly gen Ey=normal(xg)*(xb+sig*IMR)
       quietly pwcorr Ey y
       scalar Rsq=r(rho)^2
       *
       quietly lrtest tobit craggit, force
       quietly mat lrgraph=(`S',r(p),Rsq,`group')
       quietly mat LRgraph=LRgraph\lrgraph
       local Ds=`Ds'+.1
       }
di "Completed group `group'"
local group=`group'+1
local Ds=0
local seed=`seed'+39
}
svmat LRgraph
rename LRgraph1 s
rename LRgraph2 Pval
rename LRgraph3 PseudoR2
lab var Pval "P-value for H0:Tobit  vs. H1:Craggit"
lab var s "Less stochastic        <----------   Sigma   ---------->        More stochastic"
lab var Pseudo "Corr[y, E(y|x)]^2  - Pseudo R-squared"
lowess Pval s, gen(lowess)
```

```
gen refline=0.1
sum Pval
global pmax=r(max)

graph twoway (line lowess s, sort ytitle("P-value for H0:Tobit  vs. H1:Craggit"))              ///
       (scatter Pval s, msize(vsmall) m(oh) mc(gs7) xscale(r(0.5 5)) xlabel(0.5 (1.5) 5)      ///
                     ylabel(0 ( `=round(($pmax)/2, .01)' ) `=round(($pmax),.01)')             ///
                     note("Gamma=$g, Beta=$b N=$obs"))                                        ///
                     (line refline s, lw(vvthin) lcolor(red) legend(off))

/*The line "local seed=1978" is necessary to replicate results.
 To change it to something random, replace this line with
 "local seed=abs(int(rnormal(0,100)))" and re-run this code many
 times to see the robustness of the conclusion that p-values tend
 to increase as sigma increases without otherwise changing the DGM.  */
```

Figure D1. P-values for Likelihood Ratio Tests of Tobin's Assumption vs. Cragg's Alternative Using Simulated Data as σ changes holding γ and β constant



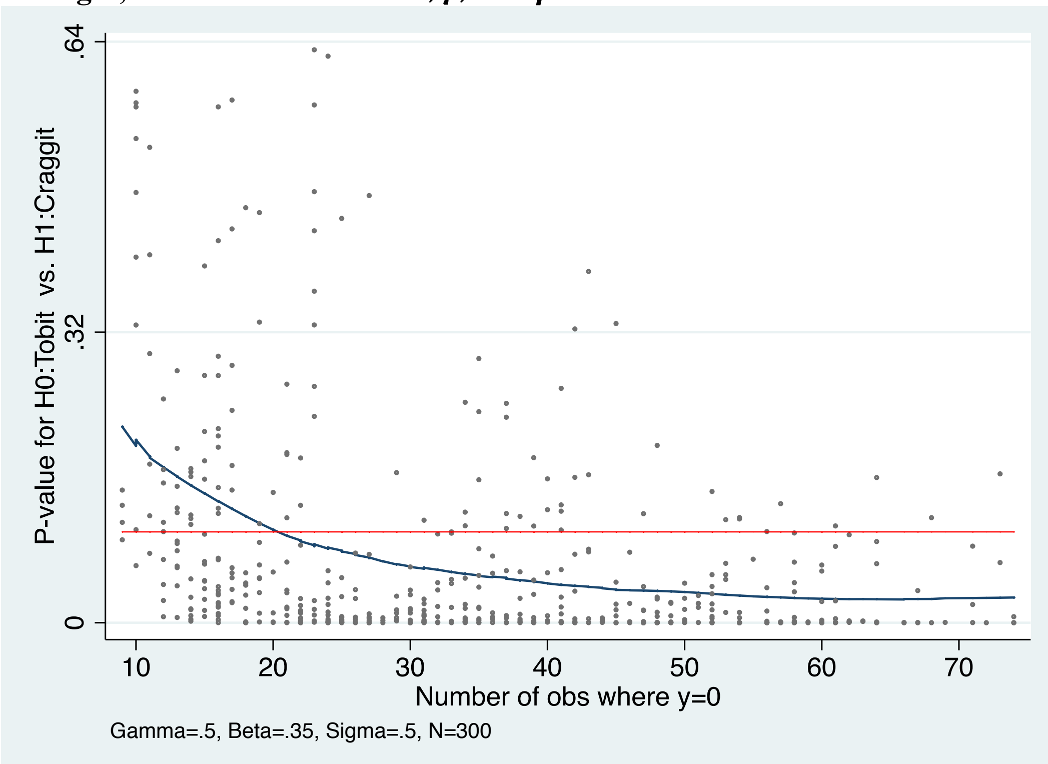
Source: the author and data simulated using Stata code in online Appendix D.
Notes: The reference line indicates 10% significance level for the null hypothesis test that $\gamma = \beta/\sigma$. Each marker on this graph indicates a unique dataset where $x \sim N(5, 5^2)$; $\Pr(y > 0|x) = \Phi(x\gamma)$ and $E(y|y > 0, x) = x\beta + u$; $u \sim N(0, \sigma^2)$ with a lower truncation point at $-x\beta$. The solid curved line is a LOWESS regression of p-values on the value of σ.

Appendix E: Stata code to show how LR tests change as the number of observations where $y = 0$ changes, holding σ, γ, and β constant

```
clear all
mat LRgraph=(.,.,.)
local g=.5          /*Establish parameters*/
local s=.5
local b=.35         /*NB: For this g & s, Tobit is when b=.25*/
local obs=300
global g=`g'        /*Also using global macros for graphing results*/
global b=`b'
global s=`s'
global obs=`obs'
/* Initiate with a "low" n_sub_0 */
local Dn0=0
local end=5
local seed=1978
local group=1
*local seed=abs(int(rnormal(0,100)))

/*generate simulated data, estimate and compare
  tobit & craggit models over changes in n_sub_0*/
while `group'<=10                 {
di "Working on group `group'"
while `Dn0'<=`end'               {
       clear
       quietly set obs `obs'
       quietly set seed `seed'
       quietly generate x=rnormal((5+`Dn0'),5)
       quietly generate pr=normal(x* `g')
       quietly generate ystar=x*`b'+rnormal(0,`s')
       quietly gen y1=runiform(0,1)
       quietly gen y=ystar if pr>y1 & ystar>0
       quietly replace y=0 if y==.
       quietly generate w=0
       quietly replace w=1 if y>0
       quietly sum w if w==0
       local n0=r(N)
       quietly tobit y x , ll(0)
       quietly estimates store tobit
       quietly craggit y x , sec(y x) iter(10)
       quietly estimates store craggit
       quietly lrtest tobit craggit, force
       quietly mat lrgraph=(`n0',r(p),`group')
       quietly mat LRgraph=LRgraph\lrgraph
       local Dn0=`Dn0'+.1
       }
di "Completed group `group'"
local group=`group'+1
local Dn0=0
local seed=`seed'+39
}
svmat LRgraph
rename LRgraph1 n0
rename LRgraph2 Pval
lab var Pval "P-value for H0:Tobit  vs. H1:Craggit"
lab var n0 "Number of obs where y=0"

lowess Pval n0, gen(lowessn0)
gen refline=0.1
sum Pval
global pmax=r(max)
sum n0
global nmin10=round(r(min),10)
```

```
global nmax10=round(r(max),10)
global nmin=r(min)
global nmax=r(max)

graph twoway (line lowessn0 n0, sort ytit("P-value for H0:Tobit  vs. H1:Craggit"))     ///
                     (scatter Pval n0 , msize(vsmall) m(oh) mc(gs7)                    ///
                     xscale(r($nmin $nmax )) xlabel($nmin10 (10) $nmax10 )             ///
                     note("Gamma=$g, Beta=$b, Sigma=$s, N=$obs")                        ///
                     ylabel(0 ( `=round(($pmax)/2, .01)' ) `=round(($pmax),.01)'))     ///
                     (line refline n0, sort lw(vthin) lco(red) legend(off) name(Dn0))
```

Figure E1. P-values for Likelihood Ratio Tests of Tobin's Assumption vs. Cragg's Alternative Using Simulated Data as the Number of Observations Where $y = 0$ (n_0) Changes, for Given Values of $\sigma, \gamma,$ and β



Source: the author and data simulated using Stata code in online Appendix E.
Notes: The reference line indicates 10% significance level for the null hypothesis test that $\gamma = \beta/\sigma$. Each marker on this graph indicates a unique dataset where $x \sim N(\mu, 5^2)$; $\mu \in [5, 10]$; $\Pr(y > 0|x) = \Phi(x\gamma)$ and $E(y|y > 0, x) = x\beta + u$; $u \sim N(0, \sigma^2)$ with a lower truncation point at $-x\beta$. The solid curved line is a LOWESS regression of p-values on the value of n_0.