Supplemental Material for Selection and Comparative Advantage in Technology Adoption: A Reconsideration

John M. Antle

May 2018

1. DATA AND SAS PROGRAMS

The data were obtained from the Tegemeo Institute at Egerton University, Kenya. The Tegemeo Institute provides the data on the condition that the data are not shared by researchers. Thus, the data used by Suri were not shared with me, nor was Suri's code shared, although descriptions of the variable construction were provided. Using this information, the data used in this study were constructed from the raw data using a SAS program available online at https://agsci.oregonstate.edu/tradeoff-analysis-project/applications-library, along with all the other SAS programs used to make the computations presented here. The variable construction was made as similar to Suri's as possible, given the information provided by Suri (see "Data Documentation for Suri 2011.pdf" that was provided to me upon request; this information is not in the *Econometrica* online supplemental material). Comparison of summary statistics from my data construction to Suri's shows that most variables are very close but many do not match exactly, possibly due to data revisions made after Suri's analysis, and also due to slightly different implementation of data construction in cases where missing values and other data issue arose. Because the original program code were not shared with me, it was not possible for me to resolve all of these data issues.

The low, medium and high productivity agro-ecozones used in this study were constructed based on the yield distributions in each of 9 zones identified in the Tegemeo surveys. Zone 6 is identified as "High Potential Maize Zone" and is the high zone in this study. The low zone in this study includes zone 2 (Coastal Lowlands), zone 3 (Eastern Lowlands) and zone 4 (Western Lowlands). All other zones with maize producing household were included in the medium zone in this study, these are zone 5 (Western Transitional), zone 7 (Western Highlands) and zone 8 (Central Highlands). The seven observations in zone 9 (Marginal Rain Shadow) were also included in the medium zone.

2. LINEAR PROBABILITY MODEL AND COBB-DOUGLAS PRODUCTION ESTIMATES

Table S1 presents linear probability models estimated using the full panel data set with hybrid use as the dependent variable, for all farms and for permanent and transitory hybrid users. These and other models with subsets of regressors were used to generate the data presented in Table II. Table S2 presents Cobb-Douglas production function estimates for the full panel with data pooled across agro-ecozones and by zone.

Table S1. Linear Probability Model Estimates for Hybrid Use, Full Panel

	All Farms		Permanent Hybrid/Non-hybrid		Transitory	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Intercept	0.451	(0.327)	1.630	(0.262)	0.269	(0.585)
Low zone	-0.612	(0.023)	-0.845	(0.019)	-0.388	(0.053)
High zone	0.193	(0.026)	0.123	(0.018)	0.116	(0.057)
Year 2000	0.009	(0.019)	0.005	(0.014)	0.013	(0.035)
Year 2004	-0.085	(0.019)	0.000	(0.014)	-0.180	(0.036)
Year 2007	0.018	(0.019)	-0.001	(0.014)	0.061	(0.036)
Year 2010	0.110	(0.019)	0.002	(0.014)	0.260	(0.036)
Province 1	0.321	(0.041)	0.078	(0.036)	0.249	(0.085)
Province 3	0.304	(0.034)	0.151	(0.024)	0.211	(0.080)
Province 4	0.227	(0.032)	0.172	(0.022)	0.154	(0.075)
Province 5	0.022	(0.026)	0.028	(0.018)	-0.029	(0.060)
Province 6	0.073	(0.032)	0.111	(0.023)	-0.018	(0.064)
Area	0.010	(0.009)	0.010	(0.007)	0.015	(0.019)
Farm size	0.038	(0.011)	0.020	(0.008)	0.022	(0.021)
Rain	0.014	(0.013)	-0.004	(0.012)	-0.001	(0.022)
Milk	0.099	(0.012)	0.068	(0.010)	0.069	(0.021)
Intercrop	-0.045	(0.016)	-0.013	(0.011)	-0.064	(0.035)
Double crop	0.012	(0.016)	0.004	(0.012)	0.005	(0.033)
Household Size	0.005	(0.002)	0.008	(0.001)	0.000	(0.003)
Max temp	0.053	(0.099)	-0.276	(0.079)	0.098	(0.178)
Dist. to fert market	-0.004	(0.001)	-0.001	(0.001)	-0.004	(0.002)
Dist. to road	-0.004	(0.004)	-0.001	(0.003)	-0.004	(0.007)
Dist. To extension	-0.004	(0.001)	-0.004	(0.001)	-0.002	(0.002)
Credit	0.012	(0.012)	-0.014	(0.009)	0.017	(0.023)
R2	0.365		0.705		0.164	
Number of Obs.	5136		2885		2251	

Table S2. Additive-Error Cobb-Douglas Production Functions, Full Panel

	All Zones		Low Zone		Medium Zone		High Zone	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Intercept	6.720	(0.652)	-3.575	(1.945)	4.923	(1.160)	4.669	(0.966)
Year 2000	0.335	(0.047)	0.299	(0.183)	0.427	(0.072)	0.139	(0.059)
Year 2004	0.326	(0.044)	0.374	(0.228)	0.489	(0.065)	0.170	(0.055)
Year 2007	0.474	(0.042)	0.749	(0.198)	0.647	(0.060)	0.259	(0.060)
Year 2010	0.238	(0.042)	0.589	(0.173)	0.615	(0.064)	-0.208	(0.063)
Province 1	-0.269	(0.078)	-0.580	(0.130)				
Province 3	0.013	(0.048)			0.844	(0.315)		
Province 4	-0.403	(0.036)	-0.740	(0.120)	0.508	(0.316)		
Province 5	-0.067	(0.035)			0.734	(0.314)	-0.007	(0.046)
Province 6	-0.131	(0.044)			0.572	(0.311)		
Hybrid	0.322	(0.042)	0.239	(0.067)	0.189	(0.059)	0.385	(0.061)
Fert Intercept	-0.646	(0.077)	-0.153	(0.139)	-0.500	(0.101)	-1.023	(0.137)
Fertilizer Quantity	0.210	(0.020)	0.127	(0.043)	0.162	(0.025)	0.238	(0.030)
Seed	0.373	(0.035)	0.352	(0.088)	0.349	(0.062)	0.147	(0.061)
Area	-0.154	(0.035)	-0.540	(0.089)	-0.307	(0.071)	-0.036	(0.032)
Farm Size	0.068	(0.034)	0.198	(0.076)	0.168	(0.057)	0.022	(0.035)
Ave. Rain	0.146	(0.035)	0.056	(0.054)	-0.101	(0.052)	0.254	(0.079)
Milk	0.102	(0.026)	0.054	(0.073)	0.131	(0.042)	0.110	(0.039)
Intercrop	0.008	(0.030)	-0.137	(0.121)	0.053	(0.048)	0.004	(0.038)
Double crop	0.017	(0.036)	0.025	(0.084)	0.013	(0.060)	-0.022	(0.047)
Household Size	0.012	(0.004)	0.019	(0.018)	0.023	(0.007)	0.008	(0.005)
Max Temp	-0.456	(0.198)	2.609	(0.630)	-0.269	(0.336)	0.480	(0.287)
R2	0.400		0.474		0.442		0.223	
Number of Obs	5225		1470		2465		1290	

3. ESTIMATION OF AES PRODUCTION MODELS

Table S3 presents estimates of the AES production models. The models for transitory adopters and non-adopters are for all zones, using zone dummies to control for zone differences, based on the fact that average hybrid returns in each zone are similarly low. Exogenous regressors include time dummies, farm characteristics, fertilizer and seed use. The average yield in all years except the year of the dependent variable is used as a proxy for unobserved site-specific productivity. Labor variables are not included due to large outliers suggesting data errors, and data inconsistency across years in the data. Models for high and medium zones estimate production functions for permanent adopters using transitory non-adopters as control observations; the model for the low zone estimates the production for permanent non-adopters, using observations of transitory adopters as control observations.

Table S3. Additive-Error Switching Regression Model Estimates

	Transitory				High Zone				Medium Zone				Low Zone			
	Non-Hybrid		Hybrid		Non-Hybrid		Hybrid		Non-Hybrid		Hybrid		Non-Hybrid		Hybrid	
Parameter	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Intercept	3.790	1.362	5.274	1.552	14.809	4.735	2.506	1.007	0.857	10.021	5.283	1.023	13.061	8.037	4.013	1.724
Low Zone	-0.203	0.070	0.040	0.093												
High Zone	0.031	0.104	0.218	0.083												
Year 2000	0.326	0.109	0.235	0.102	-0.162	0.237	0.086	0.062	0.154	0.808	0.529	0.069	0.359	0.193	0.643	0.163
Year 2004	0.413	0.101	0.347	0.133	-0.175	0.283	0.127	0.056	1.027	0.528	0.533	0.076	0.100	0.197	0.662	0.161
Year 2007	0.618	0.104	0.646	0.103	0.205	0.325	0.191	0.063	0.986	0.640	0.701	0.073	0.858	0.191	1.137	0.135
Year 2010	0.588	0.103	0.417	0.096	-0.901	0.323	-0.196	0.065	1.094	0.661	0.642	0.078	0.600	0.176	1.028	0.129
Maize Acres	-0.282	0.049	-0.320	0.058	-0.056	0.105	0.003	0.030	-0.845	0.400	-0.213	0.036	-0.447	0.129	-0.273	0.075
Farm Size	0.134	0.051	0.156	0.060	-0.194	0.134	-0.013	0.035	0.179	0.248	0.124	0.040	0.277	0.135	0.055	0.090
Average Rain	0.018	0.054	0.045	0.054	1.016	0.362	0.121	0.078	-0.480	0.421	0.030	0.055	-0.174	0.120	-0.024	0.069
Milk	0.074	0.062	0.049	0.055	0.092	0.172	0.080	0.041	0.281	0.475	0.110	0.055	0.069	0.117	0.079	0.074
Intercrop	-0.017	0.087	0.052	0.100	0.157	0.172	0.018	0.039	0.698	0.551	0.041	0.046	0.182	0.157	-0.307	0.117
Double Crop	0.108	0.070	-0.031	0.109	-0.100	0.298	-0.036	0.050	-0.317	0.483	0.024	0.057	0.125	0.134	0.233	0.108
Household Size	0.012	0.010	0.035	0.016	0.037	0.024	0.004	0.006	0.006	0.055	0.012	0.008	0.045	0.015	0.001	0.011
Max Temperature	-0.182	0.362	-0.557	0.449	-2.808	1.412	0.223	0.304	0.062	2.722	-0.926	0.282	-2.614	2.314	-0.067	0.486
Fertilizer Intercept	0.413	0.159	0.377	0.162	0.754	0.333	0.658	0.154			0.624	0.149			0.170	0.159
Fertilizer Quantity	0.201	0.045	0.121	0.042	0.159	0.087	0.156	0.033			0.146	0.029			0.094	0.048
Seed	0.141	0.055	0.325	0.076	-0.047	0.103	0.236	0.068	0.057	0.304	0.429	0.063	0.296	0.125	0.302	0.078
Average yield	0.266	0.066	0.183	0.066	0.071	0.147	0.368	0.052	0.364	0.366	0.322	0.049	-0.038	0.113	0.169	0.059
R2		0.444				0.233				0.439				0.453		
Number of Observations		1712				1020				1734				689		

4. ESTIMATION OF ADDITIONAL COST OF HYBRID SEED AND FERTILIZER

The data show that hybrid seed quantities and fertilizer use vary over time and space, and differ between farmers using hybrid and non-hybrid seed. Although Suri assumed that fertilizer is used in fixed proportions to seed, the data show that many farmers not using hybrid use fertilizer, and vice versa. Fertilizer use varies by soil and climate conditions and other factors affecting productivity. To estimate the additional cost of seed and fertilizer associated with hybrid use, I estimated a regression of expenditure on seed and fertilizer in maize yield units in 2004 and 2007 (i.e., the expenditure normalized by the price of maize) on a hybrid dummy and the average maize yield in other years to represent average productivity. The difference in predicted values of this regression for hybrid use and non-use were used to generate the distribution of hybrid seed and fertilizer cost show in the right-hand panel of figure 3. The mean cost was 99 kg/ac with a standard deviation of 18, a minimum of 74 and a maximum of 230.

5. A SPECIFICATION TEST FOR MULTIPLICATIVE-ERROR MODELS WITH FIXED EFFECTS

As noted above, one potential limitation of Suri's approach to specification and estimation of the CRC model is the requirement of linearity in the log of yield, implying a multiplicative error model. Just and Pope (1978) observed that whereas a stochastic production function can always be expressed in additive-error form (11) without imposing restrictions on the conditional distribution of output given inputs, this is not true for multiplicative-error specifications. It follows that production models with log-transformed dependent variables impose restrictions on the distribution of output conditional on inputs and other covariates. Antle (1983, 2010) showed that these restrictions can be expressed as a set of restrictions on the second and all higher-order moments of output that can be tested by parameterizing and estimating, e.g., the second and third-order moment functions. Here I develop a simpler approach and use it to test the restrictions implied by the multiplicative-error model with time-invariant fixed-effects.

For $\tilde{u}_{it}^{s} \equiv \frac{u_{it}^{s}}{E[Y_{it}^{s}|Z_{it}^{s}]}$, the additive-error model (11) can be transformed into a multiplicative-error model as:

(A1) $Y_{it}^{s} = E[Y_{it}^{s}|Z_{it}^{s}] + u_{it}^{s} = E[Y_{it}^{s}|Z_{it}^{s}](1+\tilde{u}_{it}^{s})$,

or letting $y_{it}^{s} = ln\, Y_{it}^{s}$, in log form,

(A2) $y_{it}^{s} = ln\, E\, [Y_{it}^{s}|Z_{it}^{s}] + ln(1+\tilde{u}_{it}^{s})$.

By (11) we know that $E[u_{it}^{s}|Z_{it}^{s}] = 0$, thus the expectation of $ln(1+\tilde{u}_{it}^{s})$ is generally non-zero and a function of Z_{it}^{s} as can be shown, e.g., by using a Taylor's series expansion of $ln(1+\tilde{u}_{it}^{s})$ and taking the expectation. Now suppose that the expectation conditional on Z_{it}^{s} takes the form $E[ln(1+\tilde{u}_{it}^{s})|\, Z_{it}^{s}] = k^{s} + \theta_{i}^{s}$, where $k^{s}$is a constant and θ_{i}^{s} is an individual-specific fixed effect. This will be true if the second and higher moments of u_{it}^{s} are constants for each individual, as would make sense if the factors affecting the higher moments are time-invariant factors such as soil and climate, but will not be true if time-varying factors such as input use affect the higher moments of yield. Thus the multiplicative-error fixed-effects model implies

(A3) $ln(1+\tilde{u}_{it}^{s}) = k^{s} + \theta_{i}^{s} + \varepsilon_{it}^{s}$, $E[\varepsilon_{it}^{s}|Z_{it}^{s}] = 0$.

However, the more general additive-error model implies $ln(1+\tilde{u}_{it}^{s})$ is a function of Z_{it}^{s}, e.g., letting φ^{s} be a parameter vector conformable to Z_{it}^{s}, a linear approximation is

(A4) $ln(1+\tilde{u}_{it}^{s}) = k^{s} + \theta_{i}^{s} + \varphi^{s}Z_{it}^{s} + \varepsilon_{it}^{s}$, $E[\varepsilon_{it}^{s}|Z_{it}^{s}] = 0$.

Thus, the error specification used by Suri (equations S8-S11) imposes the restriction $\varphi^s = 0$. This restriction can be tested by consistently estimating (11) via non-linear least squares, and then using the estimated model and residuals to calculate a consistent estimate of $ln(1 + \tilde{u}_{it}^s)$. This estimate of $ln(1 + \tilde{u}_{it}^s)$ can then be used to estimate (A4) as a linear fixed-effects model. Statistical significance of φ^s would be evidence against specification (A3).

6. PRODUCTION RISK ANALYSIS

In general, the mean function in (11) as well as higher moments of yield may be functions of Z_{it}^s. Following Antle (2010), both agronomic and economic considerations suggest that inputs may have asymmetric effects on the yield distribution. For example, inputs such as fertilizer have been found to be upside-risk increasing, whereas inputs such as pesticides are downside-risk reducing (Antle 2010). Partial moments of yield can be used to characterize these downside and upside dimensions of risk. Here, I use the negative and positive partial second moments of yield to represent downside and upside risk; more generally, other higher moments can be used. These partial moments can be estimated using a two-step procedure: first equation (11) is specified and estimated to obtain the residuals; second, the squared negative and positive residuals are used as dependent variables in regressions on the covariates. In the results reported here, the square root of the partial moments are used as measures of downside and upside risk in yield units.

Figure S1 shows plots of predicted expected hybrid returns against the effect of hybrid use on downside and upside risk in the low zone. The positive relationship between expected returns to hybrid and downside risk indicates that farmers with relatively high gross hybrid returns also experience high downside risk, thus, high downside risk could inhibit adoption.

Figure S2 shows plots of predicted expected hybrid returns against the effect of hybrid use on downside and upside risk in the high zone. The positive relationship between expected returns to hybrid and downside risk, combined with high upside risk for farms with low expected returns, indicates that farmers with relatively low gross hybrid returns experience low downside risk and high upside risk, thus indicating that farmers who are downside risk averse but upside risk seeking could adopt even when expected gross returns are low and expected net returns are negative.

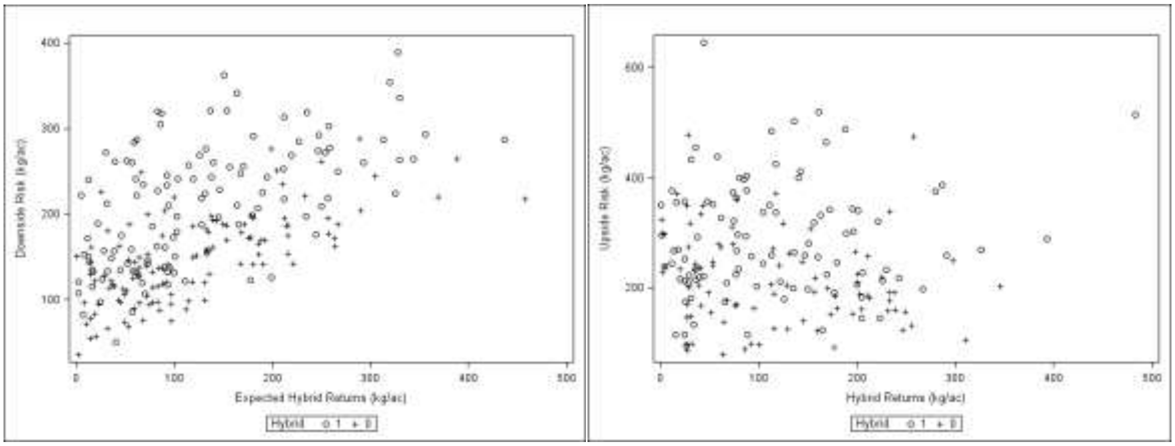

Figure S1. Low Zone Permanent Non-adopters, Hybrid Returns vs Downside Risk, Estimated with AES model

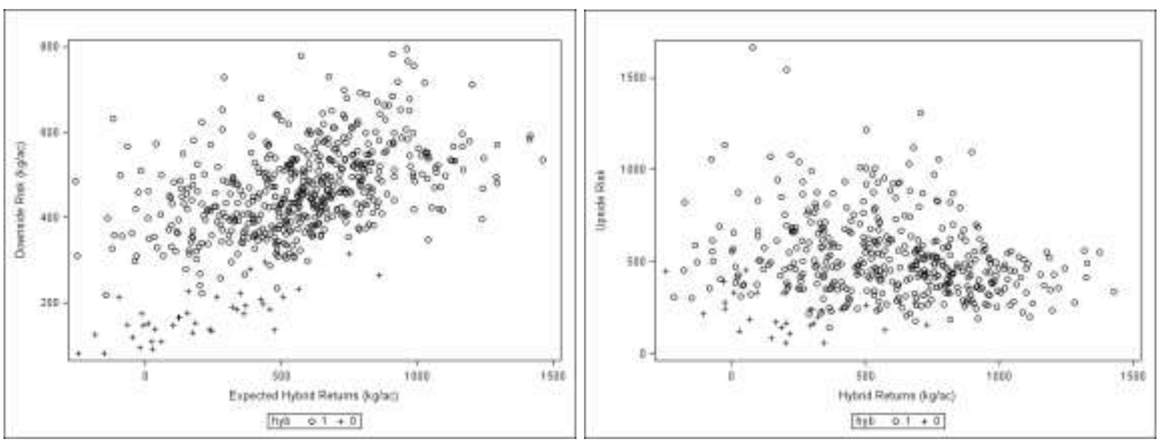

Figure S2. High Zone Permanent Adopters, Hybrid Returns vs Downside Risk, Estimated with AES model