ONLINE APPENDIX:

"The Impact of Incomplete Financial Markets on Crop-Choice, Agricultural Productivity, and Welfare"
by Albert Rodriguez-Sala
September 2024

A SOLVING THE BENCHMARK MODEL

In this section, I outline the steps to solve the benchmark model. The code for solving the model (in Python) is available in my Github repository for the project you can find the code to solve the model (in Python).

To simplify the problem, I first assume that agricultural shocks are independent and identically distributed (iid), meaning households don't need to track past realizations to form expectations. As a result, investment in both crops, shocks in both crops, and accumulated savings can be grouped into a single state variable, "cash-on-hand" (x). However, to better match the income and wealth distribution, I assume that non-agricultural income (y_{na}) has a persistent component, requiring households to track it in order to form accurate expectations.

Computing Expectations. The problem presents some complexity in terms of computing expectations. Households need to form expectations about agricultural shocks, which follow a multivariate normal distribution. To solve this, I use the Gauss-Hermite quadrature, a numerical integration method that approximates integrals using a relatively small number of nodes. Given a fixed y_{na} and z, this method allows us to compute the expected value function under shocks.

θ, ε is

$$\mathbb{E}_{\theta,\varepsilon}[V(x', z, y'_{na})] = \int_{\mathbb{R}^2_+} V\left((1-\delta)a' + \theta' zA(m'_h)^{\alpha} + \varepsilon' zB(m'_l)^{\gamma}, z, y'_{na}\right) lnMN(\theta', \varepsilon') d(\theta', \varepsilon')$$

in which $lnMN(\theta', \varepsilon')$ is the density of the log-Multinormal distribution on θ', ε'. The Gauss-Hermite quadrature rule consists in approximating this integral by the following weighted sum

$$\mathbb{E}_{\theta,\varepsilon}[V(x', z, y'_{na})] \approx \sum_{j_\theta=1}^{J_\theta} \sum_{j_\varepsilon=1}^{J_\varepsilon} \omega_{j\theta} \cdot \omega_{j\varepsilon} \cdot V\left((1-\delta)a' + \theta_{j\theta} zA(m'_h)^{\alpha} + \varepsilon_{j\varepsilon} zB(m_l)'^{\gamma}, z, y'_{na}\right) \tag{1}$$

in which $\{\omega_{j\theta},\omega_{j\varepsilon}\}$, $\{\theta_{j\theta}, \varepsilon_{j\varepsilon}\}$ are, respectively, the weights and nodes in the dimensions θ, ε, derived from the unidimensional Gauss-Hermite quadrature rule. Note that, since in this exercise it is a multidimensional integral, I use the tensor product version of the Gauss-Hermite quadrature rule, and because the shocks are correlated, I rewrite the integral in terms of uncorrelated variables using The Cholesky decomposition. To do so, I follow the steps in Maliar and Maliar (2014). To have a correct approximation of the variance of the shocks distribution, I use a relatively large number of nodes in each dimension,7, so that the total number is 49, the approximation of the multidimensional integral has good performance. Then, for the non-agricultural income process, I discretize the AR(1) process into a five-state Markov chain using the Rouwenhorst procedure. Thus, the expectation of the value function of the problem given for each permanent component z is approximated as follows

$$\mathbb{E}_{\theta,\varepsilon,y_n a}[V(x',z,y'_{na})] \approx \sum_{i=1}^{5} \pi_{,y_{na},y'_{nai}} \cdot \sum_{j_\theta=1}^{J_\theta} \sum_{j_\varepsilon=1}^{J_\varepsilon} \omega_{j\theta} \cdot \omega_{j\varepsilon} \cdot \\ V\left((1-\delta)a' + \theta_{j\theta} z A(m'_h)^\alpha + \varepsilon_{j\varepsilon} z B(m_l)'^\gamma, z, y'_{nai}\right). \tag{2}$$

Household's Problem Solution. To solve the household's problem, I use value function iteration in continuous form. This method relies on the contraction mapping of the Bellman equation to iteratively find the fixed point of the value function.

For each iteration, I follow these steps: For each permanent state z (discretized into 5 points) and possible future shocks to y_{na} (also discretized into 5 points), I interpolate the value function for tomorrow using cubic splines along the x dimension, with 100 node points on x. This gives us $V^k(x,z,y_{na}) = V^k_{f(x)}(x,z,y_{na})$, where k represents the iteration number. Next, I maximize the right-hand side of the Bellman operator over the variables c, m'_h, m'_l, and a'. The effect of m'_h, m'_l, and a' on tomorrow's expected value function is captured by tomorrow's cash-on-hand x', which is given by $x' = w_j(1-\delta)a' + \theta_{j_\theta} m'^\alpha_h + \varepsilon_{j_\varepsilon} B m'^\gamma_l$. This feeds into the interpolated value function $V^k_{f(x)}(x,z,y_{na})$ for each state z and y_{na}. Finally, the expected value function based on the choice variables is computed as $\sum_{k=1}^{5} \pi_{y_{na},y'_{na,k}} V^k_{f(x)}(x',z,y'_{na,k})$.

To maximize this function under four continuous, non-intratemporal choice variables and the linear budget constraint, I use SciPy's trust-region algorithm for constrained optimization. This algorithm implements the Byrd-Omojokun Trust-Region SQP method (Omojokun, 1989; Byrd, Gilbert, and Nocedal, 2000). After testing various optimization routines, I found that the trust-region algorithm provided the most robust and accurate results, particularly for large-scale problems.

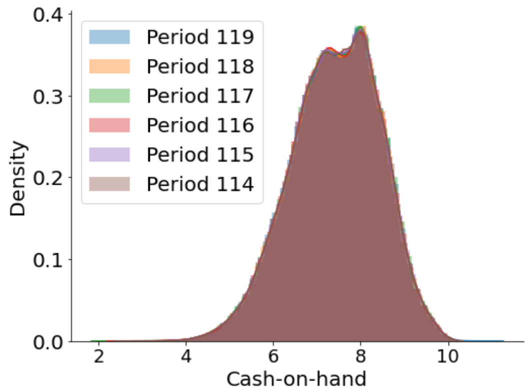
Once the solution is obtained, the maximized value becomes the new guess for the value function $V^{k+1}(x,z,y_{na})$. I repeat this process until $|V^{k+1}(x,z,y_{na}) - V^k(x,z,y_{na})|_\infty < \epsilon$, where the tolerance level ϵ is set to 9^{-5}. Lower tolerance levels caused convergence issues.

To accelerate computation, I parallelized the algorithm. In recursive problems, the solution for each state is independent of others, making this problem "embarrassingly parallel" in computing terms. I followed the approach of Fernández-Villaverde and Valencia (2018) to parallelize the value function iteration.

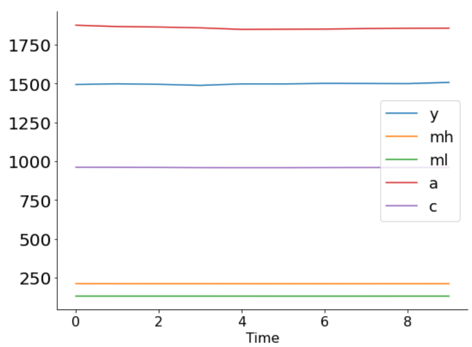
Solving the Stationary Equilibrium. After solving the household problem, I compute the stationary invariant distribution using Monte Carlo simulation. I simulate an economy with 1,000,000 households over 120 periods. Initially, all agents are identical in assets, input investment, and non-agricultural shock realizations but differ in their permanent components. This gives the initial states for households: x_0, z_0, and $y_{na,0}$. For subsequent periods, I proceed as follows: given the previous period's non-agricultural shock $y_{na,t}$, I determine the next period's realization $y_{na,t+1}$ by random sampling according to the transition matrix $\pi_{y_{na,t},y_{na,t+1}}$. Agricultural shocks θ_{t+1} and ε_{t+1} are drawn using the Gauss-Hermite weighting vector for each state (θ, ε).

For each state $s_t = x_t, z, y_{na,t}$, I calculate the next period's state for the endogenous variable as: $x_{t+1} = (1-\delta)a'(x_t, z, y_{na,t}) + \theta_t A m_h'(x_t, z, y_{na,t})^{\alpha} + \varepsilon_t B m_l'(x_t, z, y_{na,t})^{\gamma}$. This is based on the shock samples and interpolated policy functions $a'(x_t, z, y_{na,t})$, $m_l'(x_t, z, y_{na,t})$, and $m_h'(x_t, z, y_{na,t})$. The process is repeated until the state probability distribution converges. I again use parallelization to speed up this process, as each household's time series is independent and can be computed in parallel.

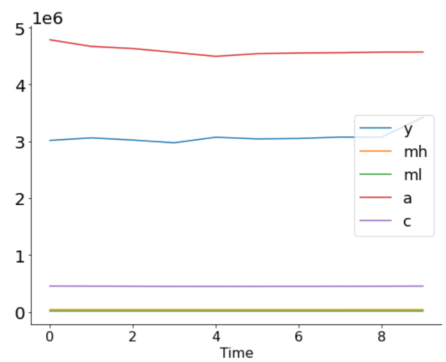
Figure 1 shows the convergence to the stationary distribution using the Monte Carlo method described above. Subplot (1a) displays the distribution of the state variable "cash-on-hand" ($\lambda(x_t)$) for the last 6 periods, demonstrating that the distribution has stabilized (i.e., $\lambda(x_t) = \lambda(x_{t-1}) = \lambda^*(x)$). Subplot (**??**b) plots the average and variance of key economic variables—income (y), high-technology intermediate usage (m_h), low-technology intermediate usage (m_l), assets (a), and consumption (c)—over the last 10 periods, showing clear stationarity in both means and variances.



(a) Cvg. in Distribution $\lambda^*(x)$



(b) Cvg. in Means



(c) Cvg. in Variances

Figure 1: Stationary Distribution (λ^*)

Notes: Subplot (a) plots the distribution of the state variable cash-on-hand, $\lambda(x_t)$, for the last 6 periods of the Monte Carlo simulation (with $N = 1,000,000$ and $T = 120$). Sub-plot (1a) plots the average and sub-plot (1b) the variance of income (y), high-technology intermediates usage (m_h), low-technology intermediates usage (m_l), assets (a), and consumption (c) across the last 10 periods of the simulation.

For completeness, Table 1 presents summary statistics for the state variables, along with other key variables of the economy.

	x	z	y_{na}	y	y_h	y_l	m_h	m_l	a	c
Mean	2,668.71	1.00	497.84	1,506.67	653.82	355.01	211.60	131.72	1,855.65	959.58
Std	2,794.00	0.61	814.12	1,850.35	1,368.04	599.53	200.96	124.08	2,138.22	673.87
Min	12.25	0.33	10.34	18.70	0.86	1.10	3.08	3.07	0.00	29.20
5%	292.68	0.33	10.34	146.82	18.79	14.38	31.95	19.74	62.56	227.65
25%	845.78	0.52	44.11	400.00	78.77	60.20	57.78	37.68	428.67	469.91
50%	1,766.90	0.82	188.18	944.23	242.86	151.82	135.14	86.92	1,099.56	757.39
75%	3,517.07	1.29	802.80	1,808.21	694.80	383.53	309.96	192.17	2,573.85	1,263.64
95%	7,922.40	2.03	3,424.96	4,721.50	3,107.22	1,581.84	662.31	402.30	6,015.03	2,379.12
99%	13,576.37	2.03	3,424.96	7,907.88	5,713.67	2,374.27	778.25	438.28	10,175.83	3,089.27
Max	70,553.58	2.03	3,424.96	61,792.94	58,090.57	19,267.85	937.94	469.66	22,697.33	4,236.64

Table 1: Benchmark Stationary Equilibria: Summary Statistics

Notes: Distributional statistics of the economy in the stationary equilibria. x denotes cash-on-hand, z permanent component, $y_n a$ the non-agricultural earnings, y total individual's income, y_h output from high-technology, y_l output from low-technology, m_h intermediates expenditure high-technology, m_l intermediates expenditure low-technology, a risk-free asset, and c consumption.

B TABLES

Table 2: Summary Main Crops Uganda. Household-Crop Level Observations.

cropID	Obs (N)	Yield ($/Acres)	Outp ($)	Land (Acres)	Interm ($)	Risk (CV)
Beans	2370	253.61	51.32	0.40	11.10	0.24
Maize	2095	294.41	54.29	0.52	13.73	0.68
Plantain Bananas	1626	770.34	231.16	0.57	22.25	0.73
Sweet Potatoes	967	257.84	75.23	0.39	13.94	0.30
Cassava	954	307.13	68.27	0.50	11.62	0.47
Groundnuts	623	657.12	158.87	0.49	15.01	0.52
Sorghum	388	187.23	58.87	0.68	5.88	0.61
Banana Beer	319	676.69	207.46	0.42	18.67	0.56
Finger Millet	295	287.94	111.61	0.54	10.71	0.20
Simsim	287	310.08	142.80	0.92	12.36	0.53
Irish Potatoes	206	392.16	111.41	0.34	22.42	0.25
Soya Beans	169	131.90	60.54	0.62	12.37	0.46
Banana Sweet	154	824.38	108.73	0.21	3.83	0.72
Sun Flower	116	115.76	109.21	0.98	18.49	0.33
Pigeon Peas	108	187.35	40.99	0.50	4.68	0.43
Yam	88	1,132.30	107.93	0.18	9.83	0.72
Cotton	83	163.18	124.26	1.25	22.62	0.34
Field Peas	76	208.48	80.67	0.58	7.28	0.43
Rice	74	1,089.37	336.97	0.80	30.22	1.02
Sugarcane	69	930.29	173.53	0.66	75.52	1.29
Tomatoes	61	492.19	166.67	0.42	42.31	0.43
Pumpkins	51	2,204.70	133.87	0.20	4.34	1.41
Cow Peas	34	207.87	63.99	0.37	14.02	0.58
Onions	27	458.12	113.70	0.41	33.74	0.60
Cabbage	26	904.85	296.26	0.42	23.38	0.71
Eggplants	20	1,428.43	223.57	0.22	8.01	0.96
Pineapples	19	540.10	164.68	0.66	24.95	0.54

Notes: Crops are ranked by most to least cultivated crops. N denotes number observations (households) by pooling all waves. Risk is measured with the Coefficient of Variation of averaging crop observations at the wave level. *Source:* UNPS: 09/10–15/16.

Table 3: Household Economic Levels and the Crop Portfolio

Log\Log	High-Yield Crops Output over Total Output ($\hat{\beta}$)	($\hat{\beta}$)	($\hat{\beta}$)	($\hat{\beta}$)
Consumption	0.9865			
	(0.0457)			
Income		0.9022		
		(0.0200)		
Wealth			0.4270	
			(0.0186)	
Land Area				0.4531
				(0.0163)
Intercept	-1.8898	-1.0110	1.9827	5.0689
	(0.3390)	(0.1437)	(0.1504)	(0.0219)
N	5956	5956	5956	7038
R2	0.07	0.25	0.08	0.10

Notes: Coefficients and their SE of the OLS estimates on: $logy^h_{it} - logy_{it} = \beta_0 + \beta_1 logx_i + u_{it}$. Where y^h_{it} is the household's agricultural production in high-yield crops and x_i is the time average household economic levels variable such as consumption or income. *Source:* UNPS: 09/10–15/16.

Table 4: High-yield and Low-yield Crops at Household Level

Wave	y_h (\$)	y_l (\$)	A_h (Acres)	A_l (Acres)	m_h (\$)	m_l (\$)
09–10	674.73	328.91	0.88	2.22	32.23	41.96
10–11	402.86	354.04	1.31	1.76	21.69	18.30
11–12	554.84	406.39	1.75	2.89	56.93	44.24
13–14	729.94	345.12	1.63	2.84	55.52	41.04
15–16	893.22	332.48	1.33	2.32	37.38	36.77
aVG	651.12	353.39	1.38	2.41	40.75	36.46

Notes: Where y represents output in $, A represents land size in acres, m represents intermediates expenditure in $. h denotes variables for high-yield crops, l for low-yield crops. Last row presents the targeted moments for high-yield crops output and input usage and low-yield crops. *Source:* UNPS: 09/10–15/16.

References

Byrd, Richard H, Jean Charles Gilbert, and Jorge Nocedal (2000). “A trust region method based on interior point techniques for nonlinear programming”. In: *Mathematical programming* 89.1, pp. 149–185.

Fernández-Villaverde, Jesús and David Zarruk Valencia (2018). *A Practical Guide to Parallelization in Economics*. Tech. rep. National Bureau of Economic Research.

Maliar, Lilia and Serguei Maliar (2014). “Numerical methods for large-scale dynamic economic models”. In: *Handbook of Computational Economics* 3, pp. 325–477.

Omojokun, Emmanuel Omotayo (1989). “Trust region algorithms for optimization with nonlinear equality and inequality constraints”. PhD thesis. University of Colorado at Boulder.