MATHEMATICAL WEB APPENDIX: LIKELIHOOD FUNCTION

Profit function (1) includes includes four differentiated structural error components $(\epsilon_{a\,i}, \epsilon_{h\,i}, \epsilon_{y\,i}, \epsilon_{\pi\,i})$, whose realizations uniquely determine the observed optimal choice of $(x_{a\,i}, x_{h\,i}, x_{y\,i}, \pi_i)$. To estimate the model we assume that the vector of unobservable returns follows an unrestricted multivariate normal distribution.

The Joint Density of Scale and Profits.—To write the likelihood function we first condition on the two continuous variables of the model, *i.e.*, scale and profits. First, from equation (3), the unobserved return associated to scale is

$$\epsilon_{y\,i} = \gamma x_{y\,i} - \theta_y - \delta_{ay} x_{a\,i} - \delta_{hy} x_{h\,i}, \tag{A1}$$

and next we rewrite the profit equation (1) as follows

$$\epsilon_{p\,i} = \pi_i - \theta_\pi - \theta_a x_{a\,i} - \theta_h x_{h\,i} - \delta_{ah} x_{a\,i} x_{h\,i} + (\gamma/2) x_{y\,i}^2, \tag{A2}$$

where we define $\epsilon_{p\,i}$ as the total unobserved return to adopting any strategy other than the scale, that is

$$\epsilon_{p\,i} = \epsilon_{\pi\,i} + \epsilon_{a\,i} x_{a\,i} + \epsilon_{h\,i} x_{h\,i}. \tag{A3}$$

Because of our normality assumptions on the distribution of $\boldsymbol{\epsilon}_i$, it follows that $\epsilon_{p\,i}$ is also normally distributed with zero mean and variance

$$\sigma_{p\,i}^2 = \sigma_\pi^2 + (\sigma_a^2 + 2\sigma_a \sigma_\pi \rho_{a\pi}) x_{a\,i} + (\sigma_h^2 + 2\sigma_h \sigma_\pi \rho_{h\pi}) x_{h\,i} + 2\sigma_a \sigma_h \rho_{ah} x_{a\,i} x_{h\,i}. \tag{A4}$$

Thus, the joint density of $\epsilon_{y\,i}$ and $\epsilon_{p\,i}$ is given by

$$g(\epsilon_{y\,i}, \epsilon_{p\,i}) = (\sigma_y \sigma_{p\,i})^{-1} \phi_2(\epsilon_{y\,i}/\sigma_y, \epsilon_{p\,i}/\sigma_{p\,i}; \rho_{yp\,i}), \tag{A5}$$

where the correlation coefficient between $\epsilon_{y\,i}$ and $\epsilon_{p\,i}$ is

$$\rho_{yp\,i} = (\sigma_\pi \rho_{y\pi} + \sigma_a \rho_{ay} x_{a\,i} + \sigma_h \rho_{hy} x_{h\,i})/\sigma_{p\,i}. \tag{A6}$$

Notice that given the distribution of $\boldsymbol{\epsilon}_i$, and making use of (A3), equations (1) and (3) define a transformation from $(\epsilon_{y\,i}, \epsilon_{p\,i})$ to $(x_{y\,i}, \pi_i)$. The determinant of the Jacobian of the inverse transformation given by equations (A1) and (A2) is

$$\boldsymbol{J} = \begin{vmatrix} \dfrac{\partial \epsilon_{y\,i}}{\partial x_{y\,i}} & \dfrac{\partial \epsilon_{y\,i}}{\partial \pi_i} \\ \dfrac{\partial \epsilon_{p\,i}}{\partial x_{y\,i}} & \dfrac{\partial \epsilon_{p\,i}}{\partial \pi_i} \end{vmatrix} = \begin{vmatrix} \gamma & 0 \\ -\gamma x_{y\,i} & 1 \end{vmatrix} = \gamma > 0. \tag{A7}$$

The determinant of the Jacobian of the inverse transformation is strictly positive because of the assumption that profits are concave in $x_{y\,i}$. Thus, equations (1) and (3) define a one-to-one transformation from $(\epsilon_{y\,i}, \epsilon_{p\,i})$ to $(x_{y\,i}, \pi_i)$ so that the joint density of $(x_{y\,i}, \pi_i)$ is

$$g(x_{y\,i}, \pi_i) = (\sigma_y \sigma_{p\,i})^{-1} \phi_2(\epsilon_{y\,i}/\sigma_y, \epsilon_{p\,i}/\sigma_{p\,i}; \rho_{yp\,i})\gamma, \tag{A8}$$

which depends on the values of $x_{a\,i}$ and $x_{h\,i}$ through equations (A1) and (A2).

Probability of Innovation Profile Choice.—The adoption of innovations is determined by conditions (13a)–(13c), which also depends on the unobserved returns to scale and innovations. Therefore, we first rewrite those equations conditioning on $\epsilon_{y\,i}$ and $\epsilon_{p\,i}$, and second, we derive the probabilities of observing each of the four possible innovation profiles. Thus, we write

$$\epsilon_{a\,i} = m_{a\,i} + \epsilon_{a.yp\,i}, \tag{A9a}$$

$$\epsilon_{h\,i} = m_{h\,i} + \epsilon_{h.yp\,i}, \tag{A9b}$$

where $m_{a\,i}$ and $m_{h\,i}$, are the expectations of $\epsilon_{a\,i}$ and $\epsilon_{h\,i}$, conditional on $\epsilon_{y\,i}$ and $\epsilon_{p\,i}$ respectively, *i.e.*,

$$m_{a\,i} = \sigma_a \frac{(\rho_{ay} - \rho_{ap\,i}\rho_{yp\,i})\epsilon_{y\,i}/\sigma_y + (\rho_{ap\,i} - \rho_{ay}\rho_{yp\,i})\epsilon_{p\,i}/\sigma_{p\,i}}{1 - \rho_{yp\,i}^2}, \tag{A10a}$$

$$m_{h\,i} = \sigma_h \frac{(\rho_{hy} - \rho_{hp\,i}\rho_{yp\,i})\epsilon_{y\,i}/\sigma_y + (\rho_{hp\,i} - \rho_{hy}\rho_{yp\,i})\epsilon_{p\,i}/\sigma_{p\,i}}{1 - \rho_{yp\,i}^2}, \tag{A10b}$$

and where the correlations between $\epsilon_{p\,i}$ and $\epsilon_{a\,i}, \epsilon_{h\,i}$ are

$$\rho_{ap\,i} = (\sigma_\pi \rho_{a\pi} + \sigma_a x_{a\,i} + \sigma_h \rho_{ah} x_{h\,i})/\sigma_{p\,i}, \tag{A11a}$$

$$\rho_{hp\,i} = (\sigma_\pi \rho_{h\pi} + \sigma_h x_{h\,i} + \sigma_a \rho_{ah} x_{h\,i})/\sigma_{p\,i}, \tag{A11b}$$

so that $\epsilon_{a.yp\,i}, \epsilon_{h.yp\,i}$ are normal variables that, by construction, are independent of $\epsilon_{y\,i}$ and $\epsilon_{p\,i}$. They have variances

$$\sigma_{a.yp\,i}^2 = \sigma_a^2 \left[1 - \frac{\rho_{ay}^2 + \rho_{ap\,i}^2 - 2\rho_{yp\,i}\rho_{ay}\rho_{ap\,i}}{1 - \rho_{yp\,i}^2}\right], \tag{A12a}$$

$$\sigma_{h.yp\,i}^2 = \sigma_h^2 \left[1 - \frac{\rho_{hy}^2 + \rho_{hp\,i}^2 - 2\rho_{yp\,i}\rho_{hy}\rho_{hp\,i}}{1 - \rho_{yp\,i}^2}\right], \tag{A12b}$$

and covariance given by

$$\text{cov}(\epsilon_{a.yp\,i}, \epsilon_{h.yp\,i}) = \sigma_a \sigma_h \left[\rho_{ah} - \frac{\rho_{ay}\rho_{hy} + \rho_{ap\,i}\rho_{hp\,i} - \rho_{yp\,i}(\rho_{ay}\rho_{hp\,i} + \rho_{ap\,i}\rho_{hy})}{1 - \rho_{yp\,i}^2}\right]. \tag{A13}$$

Next, we substitute the unobserved returns to innovations given by equations (A9a) and (A9b) into conditions (13a)–(13c) and after rearranging terms we get

$$q_{ai}\epsilon_{a.ypi} > -q_{ai}(k_{ai} + \delta x_{hi}), \tag{A14a}$$
$$q_{hi}\epsilon_{h.ypi} > -q_{hi}(k_{hi} + \delta x_{ai}), \tag{A14b}$$
$$q_{hi}\epsilon_{s.ypi} > -q_{hi}\left[k_{hi} + \delta/2 + s_i(k_{ai} + \delta/2)\right], \tag{A14c}$$

where

$$k_{ai} = \kappa_{ai} + m_{ai}, \tag{A15a}$$
$$k_{hi} = \kappa_{hi} + m_{hi}, \tag{A15b}$$
$$\epsilon_{s.ypi} = \epsilon_{h.ypi} + s_i\epsilon_{a.ypi}, \tag{A15c}$$

which is a normal variable with zero mean and variance equal to

$$\sigma^2_{s.ypi} = \sigma^2_{a.ypi} + \sigma^2_{h.ypi} + 2s_i\sigma_{a.ypi}\sigma_{h.ypi}\rho_{ah.ypi}. \tag{A16}$$

Furthermore, the correlation coefficients among $\epsilon_{s.ypi}$ and $\epsilon_{a.ypi}, \epsilon_{h.ypi}$ are

$$\rho_{as.ypi} = (\sigma_{h.ypi}\rho_{ah.ypi} + s_i\sigma_{a.ypi})/\sigma_{s.ypi}, \tag{A17a}$$
$$\rho_{hs.ypi} = (\sigma_{h.ypi} + s_i\sigma_{a.ypi}\rho_{ah.ypi})/\sigma_{s.ypi}. \tag{A17b}$$

Consider now the probability that firm i adopts both innovations, *i.e.*, $x_{ai} = 1$, and $x_{hi} = 1$. Then, conditional on ϵ_{yi} and ϵ_{pi}, conditions (8a)–(8c) must hold; that is

$$\epsilon_{a.ypi} > -k_{ai} - \delta, \tag{A18a}$$
$$\epsilon_{h.ypi} > -k_{hi} - \delta, \tag{A18b}$$
$$\epsilon_{s.ypi} > -k_{ai} - k_{hi} - \delta. \tag{A18c}$$

There are two cases of interest depending on the value of δ:

1) $\delta \leq 0$. In this case the last of the above inequalities does not bind. This case corresponds to the bottom of Figure 1 where $S_i(1,1)$ is rectangular and thus, the probability of adopting both innovations becomes

$$\Pr(x_{ai} = 1, x_{hi} = 1) = \Pr(\epsilon_{a.ypi} > -k_{ai} - \delta, \epsilon_{h.ypi} > -k_{hi} - \delta), \tag{A19}$$

which, given our assumption of joint normal distribution, leads to

$$\Pr(x_{ai} = 1, x_{hi} = 1) = \Phi_2\left(\frac{k_{ai} + \delta}{\sigma_{a.ypi}}, \frac{k_{hi} + \delta}{\sigma_{h.ypi}}; \rho_{ah.ypi}\right). \tag{A20}$$

$\Phi_2(\cdot;\rho)$ is the cumulative density function of a standard bivariate normal distribution with correlation coefficient ρ, which in this case is the correlation coefficient between $\epsilon_{a.ypi}$ and $\epsilon_{h.ypi}$ (or, equivalently, the correlation between ϵ_{ai} and ϵ_{hi} conditional on $\epsilon_{y\,i}, \epsilon_{p\,i}$):

$$\rho_{ah.yp\,i} = \text{cov}(\epsilon_{a.yp\,i}, \epsilon_{h.yp\,i})/(\sigma_{a.yp\,i}\sigma_{h.yp\,i}). \tag{A21}$$

2) $\delta > 0$. Now the three inequalities (A18a)–(A18c) bind. This case corresponds to the top of Figure 1 where $\mathcal{S}_i(1,1)$ is no longer rectangular. To compute the probability of adopting both innovations, we split the region defined by the three inequalities (A18a)–(A18c) into the following two disjoint areas defined by

$$\epsilon_{a.yp\,i} > -k_{a\,i}, \tag{A22a}$$
$$\epsilon_{h.yp\,i} > -k_{h\,i} - \delta, \tag{A22b}$$

and by

$$-k_{a\,i} > \epsilon_{a.yp\,i} > -k_{a\,i} - \delta, \tag{A23a}$$
$$\epsilon_{s.yp\,i} > -k_{a\,i} - k_{h\,i} - \delta, \tag{A23b}$$

where the second set of inequalities make use of a change of basis so that the integration region defined in the $(\epsilon_{a.yp\,i}, \epsilon_{s.yp\,i})$ plane is rectangular. Integrating the probability density function of $(\epsilon_{a.yp\,i}, \epsilon_{h.yp\,i})$ over the area defined by (A22a) and (A22b) we get

$$\Pr(\epsilon_{a.yp\,i} > -k_{a\,i}, \epsilon_{h.yp\,i} > -k_{h\,i} - \delta) = \Phi_2\left(\frac{k_{a\,i}}{\sigma_{a.yp\,i}}, \frac{k_{h\,i} + \delta}{\sigma_{h.yp\,i}}; \rho_{ah.yp\,i}\right), \tag{A24}$$

and integrating the probability density function of $(\epsilon_{a.yp\,i}, \epsilon_{s.yp\,i})$ over the region defined by (A23a)–(A23b) we have

$$\begin{aligned}\Pr(-k_{a\,i} > \epsilon_{a.yp\,i} > -k_{a\,i} - \delta, \epsilon_{s.yp\,i} > -k_{a\,i} - k_{h\,i} - \delta) = \\ \Phi_2\left(\frac{k_{a\,i} + \delta}{\sigma_{a.yp\,i}}, \frac{k_{a\,i} + k_{h\,i} + \delta}{\sigma_{s.yp\,i}}; \rho_{as.yp\,i}\right) - \\ \Phi_2\left(\frac{k_{a\,i}}{\sigma_{a.yp\,i}}, \frac{k_{a\,i} + k_{h\,i} + \delta}{\sigma_{s.yp\,i}}; \rho_{as.yp\,i}\right).\end{aligned} \tag{A25}$$

Finally, combining (A24), and (A25) we obtain the probability that a firm engages in both product and process innovation as

$$\begin{aligned}\Pr(x_{a\,i} = 1, x_{h\,i} = 1) = \Phi_2\left(\frac{k_{a\,i}}{\sigma_{a.yp\,i}}, \frac{k_{h\,i} + \delta}{\sigma_{h.yp\,i}}; \rho_{ah.yp\,i}\right) + \\ \Phi_2\left(\frac{k_{a\,i} + \delta}{\sigma_{a.yp\,i}}, \frac{k_{a\,i} + k_{h\,i} + \delta}{\sigma_{s.yp\,i}}; \rho_{as.yp\,i}\right) -\end{aligned} \tag{A26}$$

$$\Phi_2\left(\frac{k_{a\,i}}{\sigma_{a.yp\,i}}, \frac{k_{a\,i}+k_{h\,i}+\delta}{\sigma_{s.yp\,i}}; \rho_{as.yp\,i}\right).$$

We can determine the probabilities of adopting each innovative profile in a similar manner. To provide a general notation, let's define the indicator variable I_i as

$$I_i = \begin{cases} 1 & \text{if} \quad s_i\delta > 0, \\ 0 & \text{if} \quad s_i\delta \leq 0. \end{cases} \tag{A27}$$

Then, since $x_{a\,i}$ and $x_{h\,i}$ may take only values in $\{0, 1\}$, we have

$$\begin{aligned} \text{(A28)} \quad & \Pr(x_{a\,i}, x_{h\,i}) = \\ & \Phi_2\left(q_{a\,i}\frac{k_{a\,i}+\delta[I_i - x_{h\,i}(2I_i-1)]}{\sigma_{a.yp\,i}}, q_{h\,i}\frac{k_{h\,i}+\delta x_{a\,i}}{\sigma_{h.yp\,i}}; s_i\rho_{ah.yp\,i}\right) \\ & + I_i s_i\left[\Phi_2\left(\frac{k_{a\,i}+\delta}{\sigma_{a.yp\,i}}, q_{h\,i}\frac{k_{h\,i}+\delta/2+s_i[k_{a\,i}+\delta/2]}{\sigma_{s.yp\,i}}; q_{h\,i}\rho_{as.yp\,i}\right)\right. \\ & \quad \left. - \Phi_2\left(\frac{k_{a\,i}}{\sigma_{a.yp\,i}}, q_{h\,i}\frac{k_{h\,i}+\delta/2+s_i[2k_{a\,i}+\delta/2]}{\sigma_{s.yp\,i}}; q_{h\,i}\rho_{as.yp\,i}\right)\right]. \end{aligned}$$

The Likelihood Function.—Finally, we write the unconditional probability of observing a firm with specific strategy choices by multiplying the conditional probability of a given innovation profile (A28), by the join density of the distribution of scale and profits from other activities (A8), to obtain the contribution of observation i to the logarithm of the likelihood function

$$\begin{aligned} \text{(A29)} \quad & \ln L_i(\boldsymbol{\Theta}|x_{y\,i}, \pi_i, x_{a\,i}, x_{h\,i}) = \\ & \ln\gamma - \ln\sigma_y - \ln\sigma_{p\,i} + \ln\phi_2(\epsilon_{y\,i}/\sigma_y, \epsilon_{p\,i}/\sigma_{p\,i}; \rho_{yp\,i}) + \\ & \ln\left[\Phi_2\left(q_{a\,i}\frac{k_{a\,i}+\delta[I_i - x_{h\,i}(2I_i-1)]}{\sigma_{a.yp\,i}}, q_{h\,i}\frac{k_{h\,i}+\delta x_{a\,i}}{\sigma_{h.yp\,i}}; s_i\rho_{ah.yp\,i}\right) + \right. \\ & I_i s_i \Phi_2\left(\frac{k_{a\,i}+\delta}{\sigma_{a.yp\,i}}, q_{h\,i}\frac{k_{h\,i}+\delta/2+s_i[k_{a\,i}+\delta/2]}{\sigma_{s.yp\,i}}; q_{h\,i}\rho_{as.yp\,i}\right) - \\ & \left. I_i s_i\ \Phi_2\left(\frac{k_{a\,i}}{\sigma_{a.yp\,i}}, q_{h\,i}\frac{k_{h\,i}+\delta/2+s_i[2k_{a\,i}+\delta/2]}{\sigma_{s.yp\,i}}; q_{h\,i}\rho_{as.yp\,i}\right)\right], \end{aligned}$$

where $\boldsymbol{\Theta} = (\theta_a, \theta_h, \theta_y, \theta_\pi, \delta_{ah}, \delta_{ay}, \delta_{hy}, \gamma, \sigma_a, \sigma_h, \sigma_y, \sigma_\pi, \rho_{ah}, \rho_{ay}, \rho_{a\pi}, \rho_{hy}, \rho_{h\pi}, \rho_{y\pi})'$ is the vector of parameters of the model.

Normalization.—Directly maximizing the log-likelihood function (A29) with respect to the elements of parameter vector $\boldsymbol{\Theta}$ is complicated by the non-negativity of some parameters (as is the case of γ and the standard deviations σ_a, σ_h, σ_y, and σ_π) and by the restrictions on the correlation coefficients (ρ_{ah}, ρ_{ay}, …, $\rho_{y\pi}$) imposed by the positive definiteness of correlation matrix $\boldsymbol{R}$ given in equation (15). We thus reparameterize the likelihood function in order to avoid these restrictions. First, we define the parameter φ as

$$\varphi = \ln(\gamma), \tag{A30}$$

so that φ is unrestricted although γ is restricted to be positive. Second, we can write the covariance matrix of unobservables $(\epsilon_{a\,i}, \epsilon_{h\,i}, \epsilon_{y\,i}, \epsilon_{\pi\,i})'$ as

$$\boldsymbol{\Sigma} = \boldsymbol{DRD}, \tag{A31}$$

where $\boldsymbol{D}$ is a diagonal matrix whose main diagonal entries are $(\sigma_a, \sigma_h, \sigma_y, \sigma_\pi)$ and $\boldsymbol{R}$ is the correlation matrix given in equation (15). Consider now the Cholesky decomposition of $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma} = \boldsymbol{LL}', \tag{A32}$$

where $\boldsymbol{L}$ is a lower triangular matrix with strictly positive diagonal entries. From l_{ij}, the non-zero entries of matrix $\boldsymbol{L}$, we define

$$\lambda_{ij} = \begin{cases} l_{ij}, & j < i; i = 1, \ldots, 4, \\ \ln(l_{ii}), & i = 1, \ldots, 4. \end{cases} \tag{A33}$$

The λ_{ij} parameters are continuous differentiable functions of the standard deviations and correlation coefficients of the unobservables $(\epsilon_{a\,i}, \epsilon_{h\,i}, \epsilon_{y\,i}, \epsilon_{\pi\,i})'$. But, the λ_{ij} are not restricted and could take any value. The remaining parameters in $\boldsymbol{\Theta}$, *i.e.*, the direct returns to the choice variables, $\theta_a, \theta_h, \theta_y, \theta_\pi$, and the complementarity parameters, δ_{ah}, δ_{ay}, δ_{hy}, are not restricted.

The above transformations define a function between the unrestricted parameter vector $\boldsymbol{\Theta}^* = (\theta_a, \theta_h, \theta_y, \theta_\pi, \delta_{ah}, \delta_{ay}, \delta_{hy}, \varphi, \lambda_{11}, \lambda_{21}, \lambda_{22}, \lambda_{31}, \lambda_{32}, \lambda_{33}, \lambda_{41}, \lambda_{42}, \lambda_{43}, \lambda_{44})'$ and $\boldsymbol{\Theta}$:

$$\boldsymbol{\Theta} = c(\boldsymbol{\Theta}^*). \tag{A34}$$

The log-likelihood function can finally be written as

$$\ln L_i(\boldsymbol{\Theta} | x_{y\,i}, \pi_i, x_{a\,i}, x_{h\,i}) = \ln L_i(c(\boldsymbol{\Theta}^*) | x_{y\,i}, \pi_i, x_{a\,i}, x_{h\,i}) = \ln L_i^*(\boldsymbol{\Theta}^* | x_{y\,i}, \pi_i, x_{a\,i}, x_{h\,i}), \tag{A35}$$

so that we obtain our estimates by maximizing $\ln L_i^*$ with respect to $\widehat{\boldsymbol{\Theta}}^*$ and applying the transformation $\widehat{\boldsymbol{\Theta}} = c(\widehat{\boldsymbol{\Theta}}^*)$.

Simulations.—The simulations are performed by drawing $\boldsymbol{\Theta}^{*(i)}$ from a multivariate normal distribution with mean $\widehat{\boldsymbol{\Theta}}^*$ and a cluster-robust covariance matrix estimator of

$\widehat{\mathbf{\Theta}}^*$. For each draw of the parameter vector, $\mathbf{\Theta}^{(i)} = c\left(\mathbf{\Theta}^{*(i)}\right)$, we perform a random draw of $(\epsilon_{a\,i}, \epsilon_{h\,i}, \epsilon_{y\,i}, \epsilon_{\pi\,i})'$ for every observation in the sample, using a multivariate normal distribution with mean 0 and covariance matrix $\mathbf{\Sigma}^{(i)}$ obtained from the corresponding elements of $\mathbf{\Theta}^{(i)}$. Finally, with these simulated sample of the unobservables we solve for the values of the endogenous variables $x_{a\,i}, x_{h\,i}, x_{y\,i}, x_{\pi\,i}$, as discussed in Section II.

This procedure allows us to compute the expected change in the endogenous variables that result from changes in exogenous variables that are discussed in Section III.C. For each scenario we used 2,000 different parameter vectors and for each of those draws we generate 150 random samples. In this way, we take into account the impact of the estimation uncertainty into the simulation results.