Meeting 2nd of july

Lucas Condeza*

August 3, 2025

1 Estimation of production model.

Here we present the estimation algorithm fixing last's week mistakes. The main mistake was to consider each vjc as a different observation, without considering that two observations that share the same captain have the same captain random effect.

v indexes voyages, j products (bones, oil, sperm), τ_v is duration, then the production is:

$$y_{vj} \;=\; \begin{cases} (\tau_v\, w_{vj})^{\alpha_j}, & \text{with } P \;=\; \frac{1}{1+\exp\left(\gamma_0 \,-\, \gamma_1\, \log(w_{vj})\right)}, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$\log w_{vj} \;=\; X_f\, \beta_j \;+\; \delta_j\, a_c \;+\; \omega_{vj}. \tag{1}$$

Where

$$a_c \sim \mathcal{N}(0, \sigma_a) \quad \text{and} \quad \omega_{vj} \sim \mathcal{N}\big(0, \Sigma_\omega\big).$$

Denote y^*_{vj} the latent output -not always observed- hence:

$$\frac{y_{vj}^{*1/\alpha_j}}{\tau_v} = w_{vj} \implies \frac{\log(y^*_{vj})}{\alpha_j} = \log(\tau_v) + \log(w_{vj}) \tag{2}$$

replacing in the expression for w_{vj} we have:

$$\frac{\log(y^*_{vj})}{\alpha_j} = \log(\tau_v) + X_f\, \beta_j \;+\; \delta_j\, a_c \;+\; \omega_{vj}.$$

We also define:

$$d_{vj} = 1\{y_{vj} > 0\} \quad \hat{w}_{vj} = \log(w_{vj})$$

Estimation procedure

The vector of parameters $\theta = (\beta_j, \delta_j, \gamma_0, \gamma_1, \sigma_a, \Sigma_\omega)$ is estimated by maximizing the likelihood.

Let's start by constructing the likelihood of observing the production triplet for a particular voyage v, given that we know the captain fixed effect a_c.

*Yale University

$$L_v(\theta, a_c) = \prod_j \left[\Pr(y_{vj} > 0|y^*_{vj}) f(y^*_{vj})\right]^{d_{vj}} \Pr(y_{vj} = 0)^{1-d_{vj}} \tag{3}$$

where

$$\Pr(y_{vj} > 0|y^*_{vj}) = \frac{1}{1 + \exp\Big(\gamma_0 \; - \; \gamma_1 \left[\frac{\log(y^*_{vj})}{\alpha_j} - \log(\tau_v)\right]\Big)} \tag{4}$$

where the equality follows from equation 2. When product is positive we can replace the latent product by the observed one.

By a change of variables[1] we have that:

$$f_y(y^*_{vj}) = f_j\left(\omega_{vj} = \frac{\log(y^*_{vj})}{\alpha_j} - \log(\tau_v) - X_f\,\beta_j - \delta_j\,a_c\right)\left|\frac{1}{\alpha_j y^*_{vj}}\right|$$

where $f_j(\cdot)$ is the pdf of a normal distribution with standard deviation $\Sigma_\omega(j)$.

Hence, we have that when the product is positive the likelihood contribution is:

$$\frac{1}{1 + \exp\Big(\gamma_0 \; - \; \gamma_1 \left[\frac{\log(y_{vj})}{\alpha_j} - \log(\tau_v)\right]\Big)} f_j\left(\frac{\log(y_{vj})}{\alpha_j} - \log(\tau_v) - X_f\,\beta_j - \delta_j\,a_c\right)\left|\frac{1}{\alpha_j y_{vj}}\right| \tag{5}$$

The last element we need to calculate the voyage level likelihood is $\Pr(y_{vj} = 0)$ which can be calculated from:

$$\begin{aligned}
\Pr(y_{vj} = 0) &= \int \Pr(y_{vj} = 0|\hat{w}_{vj}) f(\hat{w}) d\hat{w} \\
&= \int \left(1 - \frac{1}{1 + \exp\big(\gamma_0 \; - \; \gamma_1 \hat{w}_{vj}\big)}\right) f(\hat{w}_{vj}) d\hat{w}_{vj} \\
&= \int \left(1 - \frac{1}{1 + \exp\big(\gamma_0 \; - \; \gamma_1 \hat{w}_{vj}\big)}\right) f_j(\hat{w}_{vj} - (X_f\beta_j + \delta_j a_c)) d\hat{w}_{vj}
\end{aligned} \tag{6}$$

the second equality uses: $\Pr(y_{vj} = 0|w_{vj}) = 1 - \frac{1}{1+\exp\big(\gamma_0 - \gamma_1\,\log(w_{vj})\big)}$ and the third equality uses the fact that $\hat{w}_{vj}$ is normally distributed with mean $X_f\beta_j + \delta_j a_c$ and standard deviation $\Sigma_\omega(j)$

Then, from equations 5 and 6 we can construct the voyage level likelihood.

Putting it all together we have:

$$\begin{aligned}
L_v(\theta, a_c) = &\prod_j \left[\frac{1}{1 + \exp\big(\gamma_0 - \gamma_1 \hat{w}_{vj}\big)} f_j\left(\hat{w}_{vj} - X_f\,\beta_j - \delta_j\,a_c\right)\left|\frac{1}{\alpha_j y_{vj}}\right|\right]^{d_{vj}} \\
&\left[\int \left(1 - \frac{1}{1 + \exp\big(\gamma_0 - \gamma_1 w\big)}\right) f_j(w - (X_f\beta_j + \delta_j a_c)) dw\right]^{1-d_{vj}}
\end{aligned} \tag{7}$$

[1] When $y = g(\omega)$ then $f_y(y) = f_\omega\left(g^{-1}(y)\right)\left|\frac{d}{dy} g^{-1}(y)\right|$, using the fact that a

Then the likelihood for a given captain c with N voyages is:

$$L_c(\theta) = \int_{-\infty}^{\infty} L_1(\theta, a) \cdot \ldots \cdot L_N(\theta, a) f_c(a) da \tag{8}$$

where $f_c(a)$ is the density of the captain random effect. And the likelihod of the sample is given by:

$$L(\theta) = \prod_c L_c(\theta) = \prod_c \int_{-\infty}^{\infty} L_1(\theta, a) \cdot \ldots \cdot L_N(\theta, a) f_c(a) da \tag{9}$$

Then, the log-likelihood is given by:

$$l(\theta) = \sum_c \log \left[\int_{-\infty}^{\infty} L_1(\theta, a) \cdot \ldots \cdot L_N(\theta, a) f_c(a) da \right] \tag{10}$$

Underflowing issues and the log-sum-exp trick

Replacing equation 3 into the integrand of equation 8 we have that the captain likelihood is the integral over a of the following expression:

$$L_c(\theta, a) \equiv \prod_{v=1}^{N} L_v(\theta, a_c) \tag{11}$$

this multiplication can cause problems since the numbers can be too small to be stored by matlab. For example, let's say that for a captain with $N = 20$ we have that he always produced a positive quantity and that $f(y_{vj})$ is around $1/1000$, which is realistic given our data, if the probability of positive production is around $1/2$ then we have the likelihood given a guess of (θ, a_c) is: $\left(\frac{1}{1000 \cdot 2}\right)^{20}$, that is evaluating the expression in the real value of θ, when searching over θ we will find even lower values. Moreover we are also integrating over a, and for values of a that are far from the real value we will have $L_c(\theta, a) \leq L_c(\theta, a^*)$. Hence in some of the simulations we ended up with $L_c(\theta, a) = 0$. Then to obtain the captain likelihood we use GH integration:

$$L_c(\theta) = \int_{-\infty}^{\infty} L_c(\theta, a) f_c(a) da = \sum_{k=1}^{K} w_k L_c(\theta, a_k) \tag{12}$$

Then we have:

$$L_c(\theta) = \sum_{k=1}^{K} w_k \exp(\log(L_c(\theta, a_k))) = \sum_{k=1}^{K} \exp(\log(w_k) + \log(L_c(\theta, a_k)) \tag{13}$$

Define $\ell_k = \log(L_c(\theta, a_k)), \gamma_k = \log(w_k)$ then we have:

$$L_c(\theta) = \sum_{k=1}^{K} \exp(\gamma_k + \ell_k) \tag{14}$$

Note that $\gamma_k + \ell_k$ does not pose any computationally problem, but as soon as we aplly the $\exp(\cdot)$ function we have underflow issues, because we are going back to the case where we are trying to calculate $L_c(\theta, a_k)$.

The trick is to use the log-sum-exp trick[2]. Let's start by applying logs to equation 14, then we have:

$$L_c(\theta) = \exp(M)\cdot\exp(-M)\sum_{k=1}^{K}\exp(\gamma_k+\ell_k) = \exp(M)\sum_{k=1}^{K}\exp(\gamma_k+\ell_k-M) \tag{15}$$

and applying logs we have:

$$\log(L_c(\theta)) = \log\left(\sum_{k=1}^{K}\exp(\gamma_k+\ell_k-M)\right)+M \tag{16}$$

Note that $\gamma_k+\ell_k \in (-\infty,0]$ We can choose $M=\max_k\{\gamma_k+\ell_k\}$ in which case $\gamma_k+\ell_k-M\in[-\infty,0]$ so that when applying $\exp(\cdot)$ we have a result between 0 and 1.

A second alternative is to choose $M=\min_k\{\gamma_k+\ell_k\}$

2 Allowing for correlation

In this section we relax the assumption that Σ_ω is a diagonal matrix. Define $\mathcal{J}_v^+$ the set of products with positive production and $\mathcal{J}_v^0$ the one with zero production.

Previously, conditioning on θ, a_c we had that: $\hat{w}_{vj}\sim N(X_f\beta_j+\delta_j a_c,\Sigma_\omega(j))$. Denote by $\hat{w}_v=(\hat{w}_{vj})_j$ then we have that:

$$\hat{w}_w|\theta,a_c\sim\mathcal{N}\big(X_f\beta+a_c\delta,\Sigma_\omega\big)$$

If we observe the latent production (y^*_{vj}), we can write equation 3 but conditioning on the latent production :

$$\begin{aligned}L_v(\theta,a_c,y^*_{vj}) &= \prod_j P(y^*_{vj})^{d_{vj}}\left[1-P(y^*_{vj})\right]^{1-d_{vj}}\\ &= \left[\prod_{j\in\mathcal{J}_v^+}\frac{1}{1+\exp\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}\right]\left[\prod_{j\in\mathcal{J}_v^0}\frac{\exp\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}{1+\exp\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}\right]\end{aligned} \tag{17}$$

where the only source of randomness is whether production is positive.

Then we can write:

$$\begin{aligned}L_v(\theta,a_c) &= \int_{\mathbb{R}^3} L_v(\theta,a_c,y^*_v)f(y^*_v)dy^*_v\\ &= \int_{\mathbb{R}^3}\left[\prod_{j\in\mathcal{J}_v^+}\frac{1}{1+e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}\right]\left[\prod_{j\in\mathcal{J}_v^0}\frac{e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}{1+e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}\right]f(y^*_v)dy^*_v\\ &= \int_{\mathbb{R}^{|J_v^0|}}\left[\prod_{j\in\mathcal{J}_v^+}\frac{1}{1+e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}\right]\left[\prod_{j\in\mathcal{J}_v^0}\frac{e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}{1+e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}\right]f\left(y^*_{vj\in J_v^+},y^*_{j\in J_v^0}\right)dy^*_{j\in J_v^0}\end{aligned} \tag{17}$$

[2] See here for some references

where the last line used the fact that for product with positive productions we in fact observe the production, hence it is not necessary to integrate over it.

We can also write:

$$f(y^*_{vj\in J_v^+}, y^*_{vj\in J_v^0}) = f(y^*_{vj\in J_v^0} \mid y^*_{vj\in J_v^+}) f(y^*_{vj\in J_v^+})$$

hence

$$\begin{aligned}
L_v(\theta, a_c) &= \int_{\mathbb{R}^{|J_v^0|}} \left[\prod_{j\in\mathcal{J}_v^+} \frac{1}{1+e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}\right]\left[\prod_{j\in\mathcal{J}_v^0} \frac{e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}{1+e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}\right] f(y^*_{vj\in J_v^0} \mid y^*_{vj\in J_v^+}) f(y^*_{vj\in J_v^+}) dy^*_{j\in J_v^0} \\
&= \left[\prod_{j\in\mathcal{J}_v^+} \frac{1}{1+e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}\right] f(y^*_{vj\in J_v^+}) \int_{\mathbb{R}^{|J_v^0|}} \left[\prod_{j\in\mathcal{J}_v^0} \frac{e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}{1+e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}\right] f(y^*_{vj\in J_v^0} \mid y^*_{vj\in J_v^+}) dy^*_{j\in J_v^0}
\end{aligned} \tag{17}$$

Define $y^*_v = g(\hat{w}_v) = ((\tau_v \exp(\hat{w}_{vj}))^{\alpha_j})_j$ then $g^{-1}(y^*_v) = (\frac{1}{\alpha_j}\log(y^*_{vj}) - \log\tau_v)_j$ then by a change of variables we have that[3]:

$$f(y^*_{vj\in J_v^+} = x) = f_{j\in J_v^+}(g^{-1}(x)) \left|\det \frac{\partial g^{-1}(x)}{\partial x}\right| = f_{j\in J_v^+}\left(\frac{1}{\alpha_j}\log(y^*_{vj}) - \log(\tau_v)\right) \prod_{j\in J_v^+} \frac{1}{\alpha_j y^*_{vj}}$$

where the second equality uses the fact that $\frac{\partial g^{-1}(x)}{\partial x}$ is a diagonal matrix where the jth diagonal element is $1/(\alpha_j y^*_{vj})$

Similarly we have that

$$f(y^*_{vj\in J_v^0} \mid y^*_{vj\in J_v^+}) = f(\hat{w}_{vj\in J_v^0} \mid \hat{w}_{vj\in J_v^+}) \prod_{j\in J_v^0} \frac{1}{\alpha_j y^*_{vj}} = f(\hat{w}_{vj\in J_v^0} \mid \hat{w}_{vj\in J_v^+}) \prod_{j\in J_v^0} \frac{1}{\alpha_j (\tau_v \exp(\hat{w}_{vj}))^{\alpha_j}}$$

Replacing in equation 2 we have [4]:

$$\begin{aligned}
L_v(\theta, a_c) = &\left[\prod_{j\in\mathcal{J}_v^+} \frac{1}{1+e\left(\gamma_0-\gamma_1\left[\frac{\log(y^*_{vj})}{\alpha_j}-\log(\tau_v)\right]\right)}\right] f_{j\in J_v^+}\left(\frac{1}{\alpha_j}\log(y^*_{vj}) - \log(\tau_v)\right) \prod_{j\in J_v^+} \frac{1}{\alpha_j y^*_{vj}} \\
&\cdot \int_{\mathbb{R}^{|J_v^0|}} \left[\prod_{j\in\mathcal{J}_v^0} \frac{e(\gamma_0-\gamma_1 \hat{w}_{vj})}{1+e(\gamma_0-\gamma_1\hat{w}_{vj})}\right] f(\hat{w}_{vj\in J_v^0} \mid \hat{w}_{vj\in J_v^+}) d\hat{w}_{j\in J_v^0}
\end{aligned} \tag{18}$$

Recall that $\hat{w}_v \sim \mathcal{N}\left(X_f\beta + a_c\delta,\ \Sigma_\omega\right)$., we partition the vector into the positive and zero production products: $\hat{w}_v^+ = \left(\hat{w}_{vj}\right)_{j\in J_v^+}, \hat{w}_v^0 = \left(\hat{w}_{vj}\right)_{j\in J_v^0}$ where $|J_v^+| = q, |J_v^0| = 3-q$

Accordingly, we partition the mean vector and covariance matrix as:

$$\mu = X_f\beta + a_c\delta = \begin{pmatrix}\mu^+ \\ \mu^0\end{pmatrix}, \qquad \Sigma_\omega = \begin{pmatrix}\Sigma_{++} & \Sigma_{+0} \\ \Sigma_{0+} & \Sigma_{00}\end{pmatrix}.$$

[3] The formula for the multivariate change of variables: $p_Y(y) = p_X(f^{-1}(y)) \left|\det J_{f^{-1}}(y)\right|$ comes from here

[4] Then the multivariate change-of-variables is: $E(h(y)) = \int_{\mathbb{R}^n} f_Y(y)\, h(y)\, dy = \int_{\mathbb{R}^n} f_Y\big(g(w)\big)\, h\big(g(w)\big)\, |\det J_g(w)|\, dw$ Moreover we have: $f_W(w) = f_Y\big(g(w)\big)\, |\det J_g(w)|$, substituting:

$$E(h(y)) = \int_{\mathbb{R}^n} h\big(g(w)\big) f_W(w) dw = E(h(g(w))$$

Note that in this case the jacobian does not appear in the integral. source

Where: μ^+ and Σ_{++} correspond to products with positive production, μ^0 and Σ_{00} correspond to products with zero production and Σ_{+0} is the covariance between positive and zero production products.

Then applying the multivariate normal conditional distribution [5]properties we have that:

$$\hat{w}_v^0 \mid \hat{w}_v^+ \sim \mathcal{N}(\bar{\mu} = \mu^0 + \Sigma_{0+}\Sigma_{++}^{-1}(\hat{w}_v^+ - \mu^+), \bar{\Sigma} = \Sigma_{00} - \Sigma_{0+}\Sigma_{++}^{-1}\Sigma_{+0})$$

2.1 Particular case: diagonal covariance matrix

Here we show that equation 18 when the covariance matrix is diagonal ends up being the same as equation 3.

With independence we have that: $f_{j\in J_v^+}\left(\frac{1}{\alpha_j}\log(y_{vj}^*) - \log(\tau_v)\right) = \prod_j f_j\left(\frac{1}{\alpha_j}\log(y_{vj}^*) - \log(\tau_v)\right)^{d_{vj}}$ and $f(\hat{w}_{vj\in J_v^0}|\hat{w}_{vj\in J_v^+}) = f(\hat{w}_{vj\in J_v^0}) = \prod_j f(\hat{w}_{vj})^{1-d_{vj}}$ then:

$$\begin{aligned}
L_v(\theta, a_c) = &\left[\prod_j \frac{1}{1+e\left(\gamma_0 - \gamma_1\left[\frac{\log(y_{vj}^*)}{\alpha_j} - \log(\tau_v)\right]\right)} \frac{1}{\alpha_j y_{vj}^*} f_j\left(\frac{1}{\alpha_j}\log(y_{vj}^*) - \log(\tau_v)\right)\right]^{d_{vj}} \\
&\cdot \int_{\mathbb{R}^{|J_v^0|}} \left[\prod_{j\in\mathcal{J}_v^0} \frac{e(\gamma_0 - \gamma_1\hat{w}_{vj})}{1+e(\gamma_0 - \gamma_1\hat{w}_{vj})}\right] \prod_j f(\hat{w}_{vj})^{1-d_{vj}} d\hat{w}_{j\in J_v^0} \\
= &\left[\prod_j \frac{1}{1+e\left(\gamma_0 - \gamma_1\left[\frac{\log(y_{vj}^*)}{\alpha_j} - \log(\tau_v)\right]\right)} \frac{1}{\alpha_j y_{vj}^*} f_j\left(\frac{1}{\alpha_j}\log(y_{vj}^*) - \log(\tau_v)\right)\right]^{d_{vj}} \\
&\cdot \prod_j \left[\int_{\mathbb{R}} \frac{e(\gamma_0 - \gamma_1\hat{w}_{vj})}{1+e(\gamma_0 - \gamma_1\hat{w}_{vj})} f(\hat{w}_{vj}) d\hat{w}_{vj}\right]^{1-d_{vj}}
\end{aligned} \tag{19}$$

which is the same expression as in equation 3 the second equality uses the fact that if the expected value of the product of independent variables is the product of their means[6]

2.2 Some tricks

In the estimating equation 18 the product within the integral generates some problems, hence we use:

$$\frac{e(\gamma_0 - \gamma_1\hat{w}_{vj})}{1+e(\gamma_0 - \gamma_1\hat{w}_{vj})} = \frac{1}{1+e(-\gamma_0 + \gamma_1\hat{w}_{vj})}$$

[5] See here

[6] See here