# Latent Dynamic Health Production Function

Fan Wang

2022-05-27

## Contents



# 1 Latent Dynamic Health Production Function

Go to the **RMD**, **R**, **PDF**, or **HTML** version of this file. Go back to fan's REconTools Package, R Code Examples Repository (bookdown site), or Intro Stats with R Repository (bookdown site).

## 1.1 Latent Health and Observed Health

First, output and input relationship is described generally by:

$$Y_t^* = H(X_t, Z_t, Y_t^*) + \epsilon_t \, ,$$

where we assume the separability between the error term and equation $f$.

Additionally, assume that:

$$\epsilon \sim \mathrm{N}\left(\mu, \sigma\right) \, .$$

We do not observe $Y_t^*$, however, we observe discretized integers $Y_{i,t}$:

$$Y_{i,t} \equiv \left\{ j \in \mathcal{N} : G_{j-1} \le Y_{i,t}^* \le G_j , \text{ for } 0 < j \le J+1) \right\} \, .$$

Our parameters of interest are: - The $H(X_t, Z_t, Y^{*,t})$ function. - What are $\left\{G_j^*\right\}_{j=1}^J$, assume that $G_0^* = -\infty$ and $G_{J+1}^* = \infty$.

If the function is known, we understand the dyanmic evolution of health, and can analyze the effects of changing input on health status for the given population or for alternative populations with different distributions of current inputs.

## 1.2 Standard Estimation with Observed Health Status

Suppose we observe the dataset: $\{Y_i, X_i, Z_i\}_{i=1}^N$. Suppose that $E[\epsilon|X, Z] = 0$.

$$P(Y = j|X, Z) = \Phi\left(\frac{G_j - H(X, Z) - \mu}{\sigma}\right) - \Phi\left(\frac{G_{j-1} - H(X, Z) - \mu}{\sigma}\right)$$

Note that we can construct the log-likelihood, to estimate thresholds and the parameters of for example lienarized $f()$ function. Note that we can not identify $\mu$ and also can not identify $\sigma$. Note that this is ordered-logit/ordered-probit framework. In this framework, would assume homogeneous parameters across individuals.

## 1.3 Observed Probabilities

We now observe the probabilities of observing the discrete $Y_i$ values given $X$ and $Z$. From the data, we have

$$\{P(Y = j|X, Z)\}_{j=1}^J .$$

Additionally, we observe probabilities when input changes. These allow us to estimate/identify the model with individual specific parameters and thresholds, and given the nature of the question, do not have to worry about correlation between error and input changes.

$$Y_t^* = H_i(X_t, Z_t, Y_{t-1}^*) + \epsilon .$$

We have the individual-specific $H_i$ function. Our thresholds now can also be individual-specific $\{G_{ji}\}_{j=1}^J$.

## 1.4 Single Contemporaneous Input Model New

Suppose there is no dynamics and we only observe one input. For each individual, we know:

$$\{P(Y = j|Z)\}_{j=1}^J .$$

Assuming separability between the error term and the observed component of the model, we have:

$$Y^* = H_i(Z) + \epsilon .$$

### 1.4.1 Probability given Current Inputs

For each individual, it might seem like that we can normalize given their current $X$ choices so that this is equal to 0, but we can not. So the probability of observering aparticular discrete health outcome is equal to:

$$P(Y = j|Z) = \Phi(G_{ji} - H_i(Z)) - \Phi(G_{j-1,i} - H_i(Z))$$

Alternatively, this can be written as:

$$\sum_{\widehat{j}=1}^{j} \left(P(Y = \widehat{j}|Z)\right) = \Phi(G_{ji} - H_i(Z))$$

For any $j$, just using information given current choices, we can not identify $\{G_{ji}\}_{j=1}^J$, because we do not know what $H_i(Z)$ is.

### 1.4.2 Probability at Two Input Levels

Now, we observe both:

$$\{P(Y=j|Z)\}_{j=1}^{J}, \text{ and } \{P(Y=j|Z+1)\}_{j=1}^{J}$$

Since the individual thresholds, $\{G_{ji}\}_{j=1}^{J}$, applies to both probabilities, now we are able to identify the change in the $H_i$ from $Z$ to $Z+1$. Specifically, because:

$$\Phi^{-1}\left(\sum_{\widehat{j}=1}^{j}\left(P(Y=\widehat{j}|Z)\right)\right) = G_{ji} - H_i(Z)$$

Hence:

$$\Phi^{-1}\left(\sum_{\widehat{j}=1}^{j}\left(P(Y=\widehat{j}|Z)\right)\right) - \Phi^{-1}\left(\sum_{\widehat{j}=1}^{j}\left(P(Y=\widehat{j}|Z+1)\right)\right) = (G_{ji}-H_i(Z))-(G_{ji}-H_i(Z+1)) = H_i(Z+1)-H_i(Z), \forall j \in \{1$$

Note that the above will equal to the same number for any $j$ under the assumption that the error term is distributed normal.

Define the difference as:

$$\beta_i^{+} = H_i(Z+1) - H_i(Z)$$

If $H_i$ is linear in $Z$, then, $H_i(Z) = \beta_i^{+} \cdot Z$. If rather than adding 1 to $Z$ we evaluate probabilities after subtracting 1 from $Z$ and compute following the above procedure $\beta_i^{-}$, we would have $\beta_i^{+} = \beta_i^{-} = \beta_i$.

Having identified the $\beta_i$ parameter, which is individual-specific. We can now identify all $\{G_{ji}\}_{j=1}^{J}$:

$$G_{ji} = \Phi^{-1}\left(\sum_{\widehat{j}=1}^{j}\left(P(Y=\widehat{j}|Z)\right)\right) + Z \cdot \beta_i$$

We observe $Z$ and the probability, and we have just found $\beta_i$, so the thresholds are now known.

To conclude, we have Result 1 below:

> Given a latent health production function with a single observed input, a normal additive error term, and given linearity of input and latent health outcome relationship, to $T$ identify individual specific threshold and the effect of input on the latent outcome, we need to observe, for each individual, $\{P(Y=j|Z)\}_{j=1}^{J}$ and $P(Y=1|Z+1)$.

As a corrolary of Result 1, Result 2 is:

> More flexibly, given deviations in outcomes of $\beta_i^{+}$, 0, and $\beta_i^{-}$, at $Z-1$, $Z$, and $Z+1$, we can fit a quadratic function such that $H_i(Z) = \beta_0 + \beta_1 \cdot Z + \beta_2 \cdot Z^2$.

## 1.5 Mixture of Normals

For the exercse above, for $Z+1$, we only needed $P(Y=1|Z+1)$. Suppose we observe both $\{P(Y=j|Z),P(Y=j|Z+1)\}_{j=1}^{J}$ for both $Z$ and $Z+1$, then we could allow for a mixture of normal for the error term. Another way to think about this is if we obtain very different $\widehat{\beta}_i$ depending on which $P(Y=j|Z+1)$ is observed for which $j$, we can potentially explain the observed data with more flexible distributional assumptions. A mixture of normal can be assumed, rather than a normal distribution. Adding a mixture introduces three additional parameter, the mean of the second normal, its standard deviation, as well as its weight.

Given these, suppose we have $J\times 2$ probabilities, and $J=4$ ($J=5$ is one minus the rest), this means we have 8 probabilities available to use, and we have 4 threshold parameters, one $\beta_i$ parameter, and the 3 parameters associated with the 2nd normal of the mixture.

Under the mixture of two normals, we have $J$ equations:

$$\sum_{\widehat{j}=1}^{j}\left(P(Y=\widehat{j}|Z)\right)=(1-\omega_i)\cdot\Phi\left(G_{ji}\right)+\omega_i\cdot\Phi\left(\frac{G_{ji}-\mu_{\epsilon_2}}{\sigma_{\epsilon_2}}\right),\forall j\in\{1,...,J\}$$

And $J$ additional equations:

$$\sum_{\widehat{j}=1}^{j}\left(P(Y=\widehat{j}|Z+1)\right)=(1-\omega_i)\cdot\Phi\left(G_{ji}-\beta_i\right)+\omega_i\cdot\Phi\left(\frac{G_{ji}-\beta_i-\mu_{\epsilon_2}}{\sigma_{\epsilon_2}}\right),\forall j\in\{1,...,J\}$$

When $J=4$, we solve the above system of 8 equations with the following 8 unknowns:

$$\{G_{1i},G_{2i},G_{3i},G_{4i},\beta_i,\omega_i,\mu_{\epsilon_2},\sigma_{\epsilon_2}\}$$

This is Result 3.

Additionally normals can be added into the mixture if there are additional moments to fit when other inputs change for example.

## 1.6 Multiple Inputs

Suppose we have two inputs, and still assuming separability between the error term and the observed component of the model, we have:

$$Y^*=H_i(Z,X)+\epsilon\,.$$

Suppose we have probabilities

$$\{P(Y=j|Z,X),P(Y=j|Z+1,X),P(Y=j|Z,X+1),P(Y=j|Z+1,X+1)\}_{j=1}^{J}\,.$$

We can follow the prior procedures to identify

$$\beta_i=H_i(Z+1,X)-H_i(Z,X)\alpha_i=H_i(Z,X+1)-H_i(Z,X)\zeta_i=H_i(Z+1,X+1)-H_i(Z,X)$$

Given different parametric assumptions on $H_i$, we can back out different underlying production function parameters. To illustrate, suppose we have:

$$H_i=\overline{\alpha_i}\cdot X+\overline{\beta_i}\cdot Z+\overline{\zeta_i}\cdot X\cdot Z$$

Then, we can back out the underlying production function parameters with the following three equations and three unknowns:

$$\beta_i = \overline{\beta_i} + \overline{\zeta}_i \cdot X \alpha_i = \overline{\alpha_i} + \overline{\zeta}_i \cdot Z \zeta_i = \overline{\alpha_i} + \overline{\beta_i} + \overline{\zeta_i} \cdot (X + Z + 1)$$

This is Result 4.

## 1.7 Dynamics and Contemporaneous Input

Now we generalize the structure above, to account for dynamics. Our model still has only one input, but we also know the lagged input. For each individual, we know:

$$\left\{ P(Y_t = j | Z_t, Y_{t-1}) \right\}_{j=1}^{J} .$$

Suppose that we have the following dynamic relationship in the latent variable and linear input/output relationship:

$$Y_t^* = \rho \cdot Y_{t-1}^* + \beta_i \cdot Z_t + \epsilon .$$

Note that the $\rho$ is not individual-specific, but $\beta_i$ is.

### 1.7.1 Identifying $\beta_i$

Despite the inclusion of dynamics, we identify $\beta_i$ in the same way as for the problem without dynamics. This is possible because the $\rho \cdot Y_{t-1}^*$ is invariant across probabilities of arriving in different health status tomorrow whether input is $Z$ or $Z+1$. Specifically, we have, for each individual:

$$\Phi^{-1}\left( \sum_{\widehat{j}=1}^{j} \left( P(Y_t = \widehat{j} | Z_t, Y_{t-1}^*) \right) ) \right) - \Phi^{-1}\left( \sum_{\widehat{j}=1}^{j} \left( P(Y_t = \widehat{j} | Z+1, Y_{t-1}^*) \right) ) \right) = (G_{ji} - \rho \cdot Y_{t-1}^* - \beta_i \cdot Z_t) - (G_{ji} - \rho \cdot Y_{t-1}^* - \beta_i \cdot (Z_t + 1))$$

Below, we try to identify persistence $\rho$ and thresholds $\{G_{ji}\}_{j=1}^{J}$

### 1.7.2 Identify Gaps In Thresholds

The presence of potential lagged persistence, however, means that we can no longer directly obtain the individual-specific threshold coefficients, $\{G_{ji}\}_{j=1}^{J}$, as prior.

For each individual, conditional on the same lagged outcome, we have probabilities of arriving at different $j$ health status next period:

$$G_{ji} + \rho \cdot Y_{t-1}^* = \Gamma_{ji} = \Phi^{-1}\left( \sum_{\widehat{j}=1}^{j} \left( P(Y_t = \widehat{j} | Z_t, Y_{t-1}^*) \right) ) \right) + \beta_i \cdot Z_t$$

Differencing, we have:

$$\forall (j' - 1) = j \geq 1,\ \Gamma_{j',i} - \Gamma_{j,i} = (G_{j',i} + \rho \cdot Y_{t-1}^*) - (G_{ji} + \rho \cdot Y_{t-1}^*) = G_{j',i} - G_{ji} = \Delta G_{j',i}$$

So we know the individual specific threshold gaps, but we do not know $G_{j=1,i}$. We know, $\Gamma_{ji}$, from the right-hand side in the equation above, but can not distinguish the left-hand side components. Importantly, we do not observe the latent lagged variable $Y_{t-1}^* \in \mathcal{R}$, but only observed the discretized value $Y_{t-1} \in \{1, .., J\}$.

### 1.7.3 Linear Restrictions on Threshold and Lagged Latent Values given $\rho$

From the last section, we have, for $j = 1$:

$$\Gamma_{j=1,i} = G_{j=1,i} + \rho \cdot Y^*_{t-1}$$

Note that $j = 1$ is not for the lagged choice, but relates to the probability of going to $j$ in $t$.

Note that the $G_{j=1,i}$ is a bound on feasible values for $Y^*_{t-1}$ if we observe $Y_{t-1} = 1$. Suppose we observed $Y_{t-1} = 1$, we have two equations:

$$G_{j=1,i} = \Gamma_{j=1,i} - \rho \cdot Y^*_{t-1} G_{j=1,i} > 0 + 1 \times Y^*_{t-1}$$

Think of $G_{j=1,i}$ as the Y-axis value and $Y^*_{t-1}$ as the X-axis value. The first equation is a downward sloping line whose slope is controled by $\rho \in [0, 1)$, and whose Y-intercept we already know. The second equation says that we can restrict our attention to the area to the top left of the 45 degree upward sloping line through the origin. With these two equations, while we do not know the exactly values of threshold $G_{j=1,i}$ and the latent lagged value $Y^*_{j=1,i}$, we have substantially restricted the sub-set of jointly valid values that is consistent with observed probabilities given $\rho$.

Given $\rho$, and observing $Y_{t-1} = 1$, note that the two equations above provides a minimum threshold and a maximum latent value.

$$G^{\min}_{j=1,i}(Y_{t-1} = 1) = \frac{\Gamma_{j=1,i}}{1-\rho} Y^{*,\max}_{t-1}(Y_{t-1} = 1) = \frac{\Gamma_{j=1,i}}{1+\rho}$$

If we observe for the lagged discrete value $Y_{t-1} = j > 1$, then rather than having 2 equations as above, we have 3:

$$G_{j,i} = \Gamma_{j,i} - \rho \cdot Y^*_{t-1} G_{j,i} > 0 + 1 \times Y^*_{t-1} G_{j,i} < (\Gamma_{j,i} - \Gamma_{j-1,i}) + 1 \times Y^*_{t-1}$$

Which gives us bounds on the lagged latent value:

$$Y^{*,\max}_{t-1}(Y_{t-1} = j) = \frac{\Gamma_{j,i}}{1+\rho} Y^{*,\min}_{t-1}(Y_{t-1} = j) = \frac{\Gamma_{j-1,i}}{1+\rho}$$

Note that we are explicitly using the information provided by the lagged discrete health status, because that's what we are relying on to generate the inequality restrictions.

### 1.7.4 The Latent Distribution

$\rho$ and $\beta_i \cdot Z$ jointly determine the distribution of $f(Y^*)$. For simplificty, suppose that an individual who uses $Z$ level of input at $t$ uses the same level of input in all past periods, then we know that (given that $\epsilon$ is normalized as a standard normal):

$$Y^* \sim \mathcal{N}\left(\mu_{Y^*} = \left(\frac{\beta_i \cdot Z}{1-\rho}\right), \sigma^2_{Y^*} = \left(\frac{1}{1-\rho^2}\right)\right)$$

Regardless of where threshold cutoff values are, we can compute the chance of observing latent lagged value between some range $P(a < Y^*_{t-1} < b)$.

### 1.7.5 Estimating Persistence $\rho$

The intuition for looking for the persistence parameter is that even though given $\rho$, there is a continuum of threshold and lagged latent values that can explain observed probabilities. However, values of $\rho$ controls, as just shown, the distribution of the latent variable. So as $\rho$ changes, the chance for observing different ranges of latent values changes.

So given that we observe $Y_{t-1} = j$, we look for $\rho$ to maximize the probability of observing $Y_{t-1} = j$ (without needing to know what the threshold values should be):

$$\widehat{\rho}^* = \arg\max_{\widehat{\rho}} \left( \int_{\frac{\Gamma_{j-1,i}}{1+\widehat{\rho}}}^{\frac{\Gamma_{j,i}}{1+\widehat{\rho}}} \phi \left( \frac{Y^* - \left(\frac{\beta_i \cdot Z}{1-\widehat{\rho}}\right)}{\left(\frac{1}{1-\widehat{\rho}^2}\right)} \right) dY^* \right)$$

Note that: 1. $\Gamma_{j,i}$ and $\Gamma_{j-1,i}$ are based on observed probabilities 2. the integration bounds comes from the lagged discrete outcome which generates the inequality conditions 3. Discrete thresholds are unrelated to the distribution of the latent variable.

Maximizer $\widehat{\rho}$ is an individual-specific maximizer. Not entirely clear if $\widehat{\rho}^* = \rho$, that is, not clear if the procedure above identifies the true $\rho$.