Online Appendix to: Heterogeneous Experience and Constant-Gain Learning

by John Duffy and Michael Shin

Alternative Derivation of Main Results

Here we do a re-derivation of the PYL learning rule. We focus on the case with no regressors for expositional purposes and then do an alternative derivation of the main formulation using matrices.

No Regressor Case

For the case with no regressors, Equation (1) with $\delta = 0$ implies that the DGP is given by:

$$x_t = \alpha + \beta x_t^e + \epsilon_t$$

where the REE in Equation (2) is given by

$$x_t = \bar{a} + \epsilon_t, \;\; \bar{a} = (1 - \beta)^{-1}\alpha$$

This implies that the PLM is:

$$x_{i,t}^e = a_{i,t-1}$$

That is, agents are regressing on a constant with their PLM, or alternatively just taking the sample mean of past realizations x_t. To show this more clearly note that:

$$\begin{aligned}
a_{1,t} &= a_{1,t-1} + \frac{1}{1}[x_t - a_{1,t-1}] = x_t \\
a_{2,t} &= a_{2,t-1} + \frac{1}{1}[x_t - a_{1,t-1}] = \frac{1}{2}(x_t + x_{t-1}) \\
&... \\
a_{i,t} &= a_{i,t-1} + \frac{1}{i}[x_t - a_{i,t-1}] = \frac{1}{i}(x_t + x_{t-1} + ... + x_{t-i})
\end{aligned}$$

As we can see, the recursive formulation is just a way of representing the sample mean, where agents are taking a sample mean relative to their own history. Here to tie in with the main derivation closer, we will use the following for the population weights where $g = q$:

$$n_i = q(1-q)^{i-1}$$

Then, the object of interest is $x_t^e = \sum_{i=1}^{\infty} n_i x_{i,t}^e$ where plugging in for the individual PLMs implies:

$$x_{i,t}^e = \sum_{i=1}^{\infty} n_i \frac{1}{i}(x_{t-1} + x_{t-2} + ... + x_{t-i})$$

There are two equivalent ways of solving this equation. The first done in the paper is to rearrange the infinite sums. The second way is to represent all the objects as matrices. Both are equivalent but the matrices also provide visual intuition while also showing a connection between our results and models where agents use their full information sets, which is why we will re-derive the main results in the next subsection. Here, we will use the first approach with infinite sums.

Here we know the geometric-harmonic series itself is convergent. We are allowed to rearrange the infinite series as long as the series is absolutely convergent. The most common way to show this is to show that the absolute value of the terms is convergent as well. Assuming sufficient regularity of the x_t, it is clear that the geometric-harmonic terms of the $x_{i,t}^e$ are absolutely convergent as the terms are all positive.[1] Then, writing out the terms:

$$x_{i,t}^e = n_1 x_{t-1} + \frac{1}{2} n_2 (x_{t-1} + x_{t-2}) + ... + \frac{1}{i} n_i (x_{t-1} + x_{t-2} + ... + x_{t-i}) + ...$$

[1] Again, as in the paper, the key assumption that is needed is that the discounted sum of future x_t i.e., x_t is convergent, which is satisfied if x_t is stationary. Another sufficient condition is that x_t and has a growth rate that is bounded by some constant, c i.e., they are not growing faster than the denominator.

Note that every cohort places some weight on x_{t-1}. Then the pattern is every cohort besides cohorts of age 1 places weight on x_{t-2} and so on. Rearranging, we get:

$$x^e_{i,t} = \left(\sum_{i=1}^{\infty} n_i\right) x_{t-1} + \left(\sum_{i=2}^{\infty} \frac{n_i}{2}\right) x_{t-2} + ... + \left(\sum_{j=i}^{\infty} \frac{n_j}{j}\right) x_{t-j} + ...$$

That is, the individual infinite sum terms, which we can define as S_i are infinite sums to infinity but starting from $j = i$ i.e. removing all the first i terms. Then:

$$\begin{aligned} S_i &= \sum_{j=i}^{\infty} \frac{q(1-q)^{j-1}}{j} \\ &= q(1-q)^{i-1} * \left(\sum_{k=0}^{\infty} \frac{(1-q)^k}{i+k}\right) \\ &= q(1-q)^{i-1}\Phi(1-q, 1, i) \end{aligned}$$

where the second equality above follows from splitting the series, factoring out $q(1-q)^{i-1}$, and setting the new index from k and where $\Phi(1-q, 1, i)$ again is known as a Lerch transcendent. Using this result, we can then plug in and get our analogue to Equation (10):

$$x^e_{i,t} = \sum_{i=1}^{\infty} q(1-q)^{i-1}\Phi(1-q, 1, i)x_{t-i}$$

Matrices

Here, we solve for aggregate expectations, x^e_t but using the matrices approach. We will again assume the population weights case where $g = q$ for expositional purposes. Then, we can define aggregate expectations, $x^e_t \equiv \sum_{i=1}^{\infty} n_i x^e_{i,t}$. In order to solve for aggregate expectations, we need to aggregate across each individual's expectations. To solve for aggregate expectations, we adopt the following strategy. We define the infinite-dimensional matrices W, X, and N:

$$W = \begin{bmatrix} 1 & 0 & ... & ... \\ \frac{1}{2} & \frac{1}{2} & 0 & ... \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & ... \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & ... \\ ... & ... & ... & ... \end{bmatrix} \quad X = \begin{bmatrix} x_{t-1} - b'w_{t-2} \\ x_{t-2} - b'w_{t-3} \\ x_{t-3} - b'w_{t-4} \\ x_{t-4} - b'w_{t-5} \\ ... \end{bmatrix} \quad N = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ ... \end{bmatrix}$$

where W is a weighting matrix that summarizes how much weight each cohort places on past observations, X is the vector of past residualized means, and N is the vector of population weights by age. Note that X is indexed with $x_{t-1} - b'w_{t-2}$ because x_t is a function of w_{t-1} and not w_t. Then we can define x_t^e:

$$\begin{aligned} x_t^e &= N'WX \\ &= \sum_{i=1}^{\infty} n_i * \frac{1}{i} * x_{i,t}^e \\ &= \sum_{i=1}^{\infty} q(1-q)^{i-1} * \frac{1}{i} * (a_{i,t-1} + b'w_{t-1}). \end{aligned}$$

Let X_i denote the ith observation of vector X. Note that the rows of W correspond to the cohort's weights on past observation X_i, that is, the rows of W have weight 1 on the first element and 0 elsewhere because the cohort of agents of age 1 places all their weight on the most recent observation X_1.

We use the insight that the DGL learning rule under stochastic gradient learning is essentially a weighted average of past observations. Then aggregate expectations are equivalent to summing up the columns of W and then using these as weights for the individual elements of X. That is, solving for the coefficient term $a_{i,t}$ is equivalent to finding the coefficients on the past $x_{t-1} - b'w_{t-2}$ terms.

In particular, we can define the solution for aggregate expectations x_t^e for the constant a_{t-1} in the aggregate perceived law of motion (PLM) as:

$$\begin{aligned} a_{t-1} &= S_1(x_{t-1} - b'w_{t-2}) + S_2(x_{t-2} - b'w_{t-3}) + ... \\ &= \sum_{i=1}^{\infty} S_i(x_{t-i} - b'w_{t-i-1}) \end{aligned}$$

where the individual elements i of the row vector $N'W$, that is $(N'W)_i = \sum_{j=i} n_i * \frac{1}{i}$, are equivalent to S_i i.e.:

$$\begin{aligned} S_i &= (N'W)_i \\ &= \sum_{j=i}^{\infty} q(1-q)^{j-1} * \frac{1}{j} \end{aligned}$$

Note that each element of $N'W$ is itself an infinite series. S_i is essentially the partial sum of a geometric-harmonic series. Summing up each infinite series in $N'W$, leads to the following:

$$\begin{bmatrix} \sum_{i=1}^{\infty} q(1-q)^{i-1} * \frac{1}{i} & \sum_{i=2}^{\infty} q(1-q)^{i-1} * \frac{1}{i} & \sum_{i=3}^{\infty} q(1-q)^{i-1} * \frac{1}{i} & ... \end{bmatrix} \begin{bmatrix} x_{t-1} - b'w_{t-2} \\ x_{t-2} - b'w_{t-3} \\ x_{t-3} - b'w_{t-4} \\ x_{t-4} - b'w_{t-5} \\ ... \end{bmatrix}$$

To solve for aggregate expectations requires solving for the composition of two infinite series, n_i and $\frac{1}{i}$, specifically, $\sum_{i=1}^{\infty} q(1-q)^{i-1} * \frac{1}{i}$. Under PYL, aggregate expectations for the constant a_{t-1} and x_t^e follow:

$$a_{t-1} = \sum_{i=1}^{\infty} S_i(x_{t-i} - b'w_{t-i-1})$$

$$x_t^e = a_{t-1} + b'w_{t-1}$$

where,

$$S_i = \sum_{j=i}^{\infty} q(1-q)^{j-1} * \frac{1}{j}$$

Plugging in for S_i implies the following result which is the same as our main result.

Again, while the disadvantage of this approach is that the derivation is more convoluted, the matrices show that our PYL model is related to the CGL model but with a restricted history matrix.

Alternative Learning Rules for Different q

Here we provide different implied effective constant gain parameters for different q values and a wider variety of parameters. Note that $g = q$ for this exercise which implies that birth rates equal death rates. We do this for $q = 0.001, 0.002, 0.003, 0.004, 0.005$. More specifically, we choose 0.0031 as this is the case we use in the benchmark PYL model.

Additional Data

Table 1: Implied Effective CGL for Additional Data by Birth Rates

h	$q = 0.001$	$q = 0.002$	$q = 0.003$	$q = 0.004$	$q = 0.005$
$h = 0$ (PYL)	0.007	0.013	0.018	0.022	0.027
2	0.005	0.010	0.013	0.016	0.019
4	0.005	0.008	0.011	0.014	0.016
6	0.004	0.008	0.010	0.012	0.015
8	0.004	0.007	0.009	0.011	0.013
10	0.004	0.007	0.009	0.011	0.012
12	0.004	0.006	0.008	0.010	0.011
14	0.004	0.006	0.008	0.009	0.011

General Decreasing-gain Learning (DGL)

Table 2: Implied Effective CGL for General DGL by Birth Rates

α	$q = 0.001$	$q = 0.002$	$q = 0.003$	$q = 0.004$	$q = 0.005$
$\alpha = 0.25$	0.217	0.258	0.289	0.306	0.323
0.5	0.055	0.077	0.092	0.107	0.119
0.75	0.017	0.028	0.038	0.044	0.051
1 (PYL)	0.007	0.013	0.018	0.022	0.027
1.25	0.004	0.007	0.011	0.014	0.017
1.5	0.003	0.005	0.008	0.009	0.012
1.75	0.002	0.004	0.006	0.008	0.009
2	0.002	0.003	0.005	0.007	0.008
2.25	0.001	0.003	0.005	0.006	0.007
2.5	0.001	0.003	0.004	0.005	0.007
2.75	0.001	0.003	0.004	0.005	0.006
3	0.001	0.002	0.004	0.005	0.006

Biased Weights (MN)

For the model with biased weights, we also provide the implied effective constant gain under the second and third terms S_2^{MN} and S_3^{MN}.

Table 3: Implied Effective CGL for Biased Weights by Birth Rates

Gain Term	$q = 0.001$	$q = 0.002$	$q = 0.003$	$q = 0.004$	$q = 0.005$
1st Gain S_1^{MN}	0.012	0.021	0.031	0.037	0.044
2nd Gain S_2^{MN}	0.010	0.017	0.024	0.029	0.034
3rd Gain S_3^{MN}	0.009	0.015	0.021	0.025	0.029

where the formulas for the second and third gains in the MN sequence are given by:

$$\gamma_{CGL}^{MN} = S_2^{MN} = \sum_{i=2}^{\infty} q(1-q)^{i-1} * \frac{i-1}{\frac{1}{2}i(i+1)} = \frac{2q(2q - q\ln q - \ln q - 2)}{(1-q)^2}$$

$$\gamma_{CGL}^{MN} = S_3^{MN} = \sum_{i=3}^{\infty} q(1-q)^{i-1} * \frac{i-2}{\frac{1}{2}i(i+1)} = \frac{q(q^2 + 4q - 4q\ln q - 2\ln q - 5)}{(1-q)^2}$$

We will not prove the results here but they can be verified via programs such as Mathematica. In general, the implied effective constant gains for the second and third gain terms are lower than for the first gain as we would expect.