Economics Working Paper Series

Faculty of Business, Economics and Law, AUT

Information weighting under least squares adaptive learning

Jaqueson K. Galimberti

2020/04

Information weighting under least squares adaptive learning

Jaqueson K. Galimberti*
Auckland University of Technology

Abstract

This note evaluates how adaptive learning agents weight different pieces of information when forming expectations with a recursive least squares algorithm. The analysis is based on a new and more general non-recursive representation of the learning algorithm, namely, a penalized weighted least squares estimator, where a penalty term accounts for the effects of the learning initials. The paper then draws behavioral implications of different specifications of the learning mechanism, such as the cases with decreasing-, constant-, regime-switching, and age-dependent gains. The latter is shown to imply the emergence of "dormant memories" as the agents get old.

Keywords: bounded rationality, expectations, adaptive learning, memory.

JEL codes: E70, D83, D84, D90, E37, C32, C63.

"The longer you can look back,
the farther you can look forward."
–Winston Churchill

1 Introduction

Adaptive learning can generate out-of-equilibrium expectations that help explain deviations from rational expectations and an economy's transitional dynamics towards equilibrium. Here agents' beliefs are modeled through the assumption of a recursive learning mechanism that updates agents' perceptions about the economy

*Corresponding author: School of Economics, Auckland University of Technology, Auckland, New Zealand.
Personal website: https://sites.google.com/site/jkgeconoeng/.
Contact e-mail: jaqueson.galimberti@aut.ac.nz.

as new data observations become available. The weight given to these observations is a key determinant of the degree of persistence introduced by adaptive learning in the evolution of expectations, and, hence, it is an important factor in the explanation of deviations from rational expectations predictions. This note studies the behavioral implications of alternative specifications of the learning mechanism, showing how these assumptions can affect the way agents weight information when forming expectations according to a recursive least squares algorithm.

The first contribution is the proposal of a penalized weighted least squares non-recursive representation of the learning algorithm, where a penalty term accounts for the effects of the learning initial estimates. This framework, outlined in Section 2, provides flexible analytical expressions for the calculation of the weights given to different pieces of information under alternative specifications of the learning gains, including the traditional decreasing (Marcet and Sargent, 1989) and constant-gain (Sargent, 1999) specifications, as well as more sophisticated mechanisms such as endogenous gain-switching (Marcet and Nicolini, 2003) and age-dependent (Malmendier and Nagel, 2016) specifications.

Several interesting results are outlined in Section 3, including the key findings that: (i) the assumption of a diffuse initial under constant-gain implies that the profile of weights given to past observations is time-varying, hence distorting the behavioral interpretation of this mechanism in small samples of data; and, (ii) the application of decreasing-gains to cohort-level data, which leads to age-dependent gains, also leads to the emergence of "dormant memories," namely, that experiences earlier in the agent's lifetime can have a comeback and receive an increasing weight as the individual ages. Detailed derivations are provided in an Online Appendix.

2 Framework

In models with adaptive learning a perceived law of motion (PLM) is specified relating the variables agents are assumed to observe and those variables they care and need to form expectations about. Focusing on a univariate case a typical PLM specification is given by a linear regression model of the form

$$y_t = \mathbf{x}_t'\phi_t + \varepsilon_t, \tag{1}$$

where y_t is assumed to be related to a vector of (pre-determined) variables, $\mathbf{x}_t = (x_{1,t}, \ldots, x_{k,t})'$, through the vector of coefficients $\phi_t = (\phi_{1,t}, \ldots, \phi_{k,t})'$, and ε_t denotes a white noise disturbance term.

Recursive learning A recursive estimator is assumed to represent how agents update their PLM estimates as new observations become available. One popular algorithm (see Berardi and Galimberti, 2014) is given by the Recursive Least Squares (RLS),

$$\hat{\phi}_t = \hat{\phi}_{t-1} + \gamma_t \mathbf{R}_t^{-1} \mathbf{x}_t \left(y_t - \mathbf{x}_t' \hat{\phi}_{t-1} \right), \tag{2}$$

$$\mathbf{R}_t = \mathbf{R}_{t-1} + \gamma_t \left(\mathbf{x}_t \mathbf{x}_t' - \mathbf{R}_{t-1} \right), \tag{3}$$

where γ_t is a learning gain parameter, $\mathbf{R}_t$ stands for an estimate of regressors' matrix of second moments, $E[\mathbf{x}_t \mathbf{x}_t']$, and the $\left\{ \hat{\phi}_0, \mathbf{R}_0 \right\}$ initial estimates are set to be consistent with plausible agents' beliefs at the beginning of the modeled sample (see Berardi and Galimberti, 2017b). The learning gain is an important parameter of this learning mechanism because it determines how quickly new information is incorporated into the recursive estimates, and hence, how quickly agents react to different pieces of information. More precisely, the sequence of learning gains can be related with the relative weights given to sample observations in the estimation process. In order to draw this relationship it is useful to consider the non-recursive formulation corresponding to this estimation problem.

Non-recursive form When initialized from arbitrary initials, $\hat{\phi}_0$ and $\mathbf{R}_0$, the RLS has a non-recursive form given by

$$\hat{\phi}_t = \arg\min \sum_{i=1}^{t} \omega_{t,i} \left(y_i - \mathbf{x}_i' \hat{\phi}_t \right)^2 + \omega_{t,0} \left(\hat{\phi}_0' - \hat{\phi}_t' \right) \mathbf{R}_0 \left(\hat{\phi}_0 - \hat{\phi}_t \right), \tag{4}$$

$$= \left[\sum_{i=1}^{t} \omega_{t,i} \mathbf{x}_i \mathbf{x}_i' + \omega_{t,0} \mathbf{R}_0 \right]^{-1} \left[\sum_{i=1}^{t} \omega_{t,i} \mathbf{x}_i y_i + \omega_{t,0} \mathbf{R}_0 \hat{\phi}_0 \right], \tag{5}$$

where the weights are related to the sequence of learning gains according to

$$\omega_{t,i} = \begin{cases} \prod_{j=1}^{t} \left(1 - \gamma_j \right) & for\ i = 0\,(initial), \\ \gamma_i \prod_{j=i+1}^{t} \left(1 - \gamma_j \right) & for\ 0 < i < t, \\ \gamma_t & for\ i = t, \end{cases} \tag{6}$$

Interestingly, when the initial is taken into account, the RLS is equivalent to a Weighted Least Squares (WLS) estimation problem augmented with a penalty on squared deviations between estimates and initials.

Relation to literature This non-recursive formulation of the RLS for arbitrary initials has never been outlined in the previous literature. Berardi and Galimberti (2013), for example, have drawn a correspondence between the RLS and the standard (without penalty) WLS, except that under the assumption of a diffuse initial prior, when $\mathbf{R}_0 \to \mathbf{0}$. From a Bayesian point of view, $\mathbf{R}_t$ is inversely related to the uncertainty in the corresponding Kalman filter estimates of ϕ_t (see Evans et al., 2010). Hence, $\mathbf{R}_0 \to \mathbf{0}$ can be interpreted as increasing the uncertainty about the initial estimates, and it is only under this diffuse prior assumption that the standard WLS is equivalent to the RLS. Also, notice that if $\mathbf{R}_0 = \mathbf{0}$ (exactly rather than as a limit), (2)-(3) implies that $\hat{\phi}_1 = \left(\mathbf{x}_1\mathbf{x}_1'\right)^{-1}\mathbf{x}_1 y_1$, which will be indeterminate[1] for $k > 1$.

3 Information weighting

The weight given to a sample observation determines the amount of information from that particular observation that is incorporated into the parameters' estimates. The non-recursive estimator above allows the calculation of such weights for any arbitrary sequence of learning gains. Also notice that the weights, $\omega_{t,i}$, defined in equation (6), are already in relative terms, though this equivalence does not hold under diffuse initials.

3.1 Decreasing-gain

The decreasing-gain (DG) specification has been prominent in the adaptive learning literature since the seminal contributions of Bray (1982); Marcet and Sargent (1989). Under $\gamma_t^{dg} = 1/(t+1)$, the weights are given by

$$\omega_{t,i}^{dg} = \begin{cases} \prod_{j=1}^{t}\left(\frac{j}{j+1}\right) = \frac{1}{2}\frac{2}{3}\cdots\frac{t-1}{t}\frac{t}{t+1} = \frac{1}{t+1} & for\, i = 0, \\ \frac{1}{i+1}\prod_{j=i+1}^{t}\left(\frac{j}{j+1}\right) = \frac{1}{i+1}\frac{i+1}{i+2}\cdots\frac{t}{t+1} = \frac{1}{t+1} & for\, 0 < i < t, \\ \frac{1}{t+1} & for\, i = t, \end{cases}$$

and every observation receives an equal weight that is decreasing with the sample size, which makes the DG-RLS particularly interesting for the analysis of learning convergence towards equilibrium.

OLS equivalence breakup The DG-RLS is often motivated as representative of real-time econometricians due to its resemblance with the Ordinary Least Squares (OLS) estimator from econometrics,

$$\hat{\phi}_t^{ols} = \left[\sum_{i=1}^{t}\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{t}\mathbf{x}_i y_i\right].$$

[1]More precisely, $\mathbf{x}_1\mathbf{x}_1'$ must have full rank for a non-degenerate inverse to exist; to guarantee the existence of this inverse, $\gamma_t = 1$ must also be ruled out.

However, as equation (5) now makes clear, such an equivalence would require disregarding the learning initials.

3.2 Constant-gain

The constant-gain specification (CG) became popular after Sargent (1999) for its improved capability of tracking the evolution of time-varying environments. This specification has also been under the spotlight of recent research for its potential of generating escape dynamics (Williams, 2019) and asymptotically stable distributions of beliefs (Galimberti, 2019). Under $\gamma_t^{cg} = \bar{\gamma}$, the weights are given by

$$\omega_{t,i}^{cg} = \begin{cases} \prod_{j=1}^{t}(1-\bar{\gamma}) = (1-\bar{\gamma})^t & for\ i = 0, \\ \bar{\gamma}\prod_{j=i+1}^{t}(1-\bar{\gamma}) = \bar{\gamma}(1-\bar{\gamma})^{t-i} & for\ 0 < i \leq t. \end{cases}$$

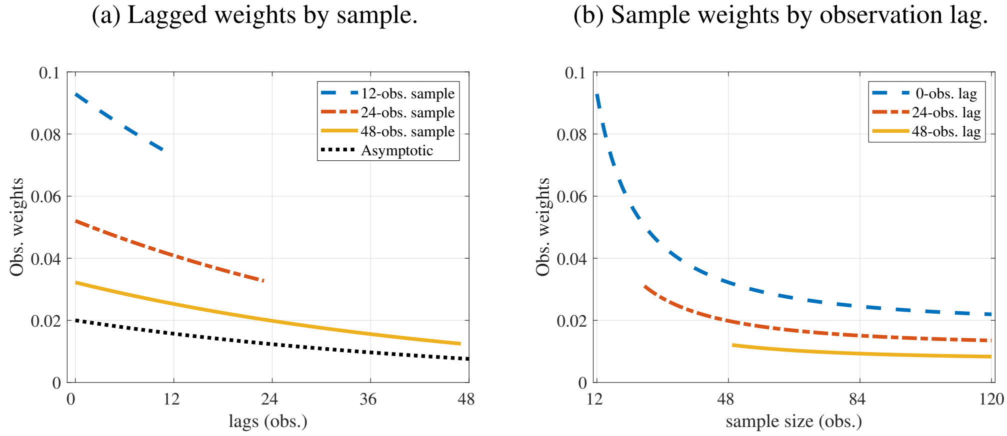
Hence, the weights given to past information under CG-RLS decrease with the observation lag ($l = t - i$), a property that makes this mechanism particularly well suited for modeling the behavioral assumption that agents give higher emphasis to more recent observations than to those farther into the past.

Finite sample distortion under diffuse initials The profile of weights given to past observations by the CG-RLS becomes time-varying under diffuse initials. Particularly,

$$\omega_{t,t-l}^{dcg} = \frac{\bar{\gamma}(1-\bar{\gamma})^l}{1-(1-\bar{\gamma})^t}, \tag{7}$$

is declining with both the observation lag and the sample size, distorting the behavioral interpretation of CG in small samples: letting t stand for age, equation (7) implies that a younger agent would assign a higher weight to any given observation than an older one experiencing the same observation. These effects are also illustrated in Figure 1..

Figure 1: Constant-gain weights under diffuse initials.



Notes: Weights calculated using equation (7) with $\bar{\gamma} = 0.02$.

Persistently influential initials The duration of the effects of the initials under CG-RLS within finite samples will depend on the gain magnitude. In fact, the number of observations for which the initial will have a greater weight than the whole sample of data, i^*, can be calculated by solving:

$$\sum_{j=1}^{i^*} \omega_{t,j}^{cg} = \omega_{t,0}^{cg},$$

$$i^* = \frac{\log\left(1/2\right)}{\log\left(1-\bar{\gamma}\right)},$$

which is decreasing with $\bar{\gamma}$. Strikingly, for a gain value of $\bar{\gamma} = 0.02$, which is typically found in applications with quarterly macroeconomic data (see Berardi and Galimberti, 2017a), $i^* \simeq 34$, or about $8\,1/2$ years of quarterly data for the CG-RLS to assign a higher weight to the sample of observations than the weight given to the learning initials in the PLM estimates.

3.3 Generalized decreasing-gain

A recent strand of the literature has revived the decreasing-gain specification with the proposal of so-called learning from experience, where agents expectations are modeled at cohort-level (Malmendier and Nagel, 2016, MN); this approach naturally leads to the emergence of dispersed beliefs, depending on the age structure of the population, which MN find empirical support in consumers survey data. Their modeling of learning is based on a generalization of decreasing-gain (GDG), where[2]

$$\gamma_t^{gdg} = \frac{\theta}{t+\theta}, \tag{8}$$

with $\theta > 0$; in fact, notice that the original DG-RLS is obtained when $\theta = 1$. Otherwise, it is interesting to cast the GDG-RLS weights in lag recursive form, starting from $\omega_{t,t}^{gdg} = \theta/t+\theta$ and expanding to

$$\begin{aligned}
\omega_{t,t-1}^{gdg} &= \omega_{t,t}^{gdg}\left(\frac{t}{t-1+\theta}\right), \\
\omega_{t,t-2}^{gdg} &= \omega_{t,t-1}^{gdg}\left(\frac{t-1}{t-2+\theta}\right), \\
&\vdots \\
\omega_{t,t-l}^{gdg} &= \omega_{t,t-l+1}^{gdg}\left(\frac{t-l+1}{t-l+\theta}\right),
\end{aligned}$$

which makes clear that the weight given to lagged observations within a sample of data decreases with the lag if $\theta > 1$, and increases if $\theta < 1$.

Age-dependent experiences The GDG specification also implies a time-varying profile of weights as the sample size grows. In behavioral terms this is what MN (p. 59) refer as "experiences earlier and later in life to have a different influence" on expectations. To see this consider how the weight given to a fixed observation lag evolves as the sample increases by one observation:

$$\begin{aligned}
\frac{\omega_{t,t-l}^{gdg}}{\omega_{t-1,t-1-l}^{gdg}} &= \frac{\frac{\theta}{t-l+\theta}\prod_{j=t-l+1}^{t}\left(\frac{j}{j+\theta}\right)}{\frac{\theta}{t-1-l+\theta}\prod_{j=t-l}^{t-1}\left(\frac{j}{j+\theta}\right)}, \\
&= \frac{t}{t-l}\left(\frac{t-1-l+\theta}{t+\theta}\right),
\end{aligned}$$

which is smaller than 1 if $\theta < t/l$ (always the case for $\theta \leq 1$, and for $l = 0$). Hence, the lagged observation weight decreases with the sample size as long as θ is small enough relative to the sample size/lag ratio.

---

[2] The exact gain specification in MN is

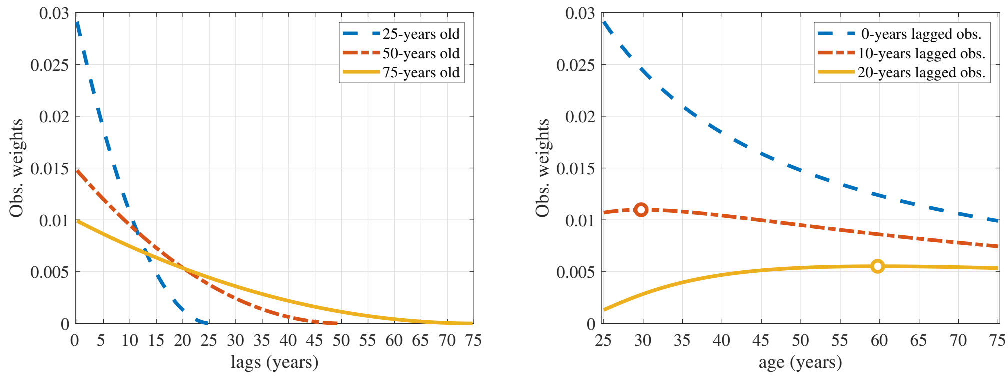
$$\gamma_{t,s}^{mn} = \begin{cases} \frac{\theta}{t-s} & if\, t-s \geq \theta, \\ 1 & if\, t-s < \theta, \end{cases}$$

where $t-s$ stands for the agent's age; however, this specification is problematic because it leads to degenerate inverses when $t-s \leq \theta$ (see also footnote (1)), an issue circumvented with a slight modification in equation (8).

Dormant memories When $\theta > 1$, the weight given to some lagged observations may increase as the sample size grows—although an individual's memories of past experiences tend to fade away, "dormant memories" may have a comeback as the individual ages. This effect is more clearly illustrated in Figure 2b: notice how the weights given to 20-years lagged observations increase during an individual's lifetime at the same time that the weights given to more recent observations are decreasing. Also notice that the higher θ the stronger the effects of dormant memories.

Figure 2: Generalized decreasing-gain weights.

(a) Lagged weights by age. (b) Lifetime weights by observation lag.



Notes: Weights calculated for quarterly data using eqs. (6)-(8) and $\theta = 3$. The circles in panel (b) indicate the turning points of the corresponding curves.

3.4 Time-varying gains

Time-varying gains offer an alternative to relax the determination of information weighting under learning. In the literature, this has been achieved either by merging the decreasing- and constant-gain specifications (Marcet and Nicolini, 2003; Milani, 2014), or by turning the determination of the gains endogenous with an additional adaptation mechanism (Kostyshyna, 2012; Berardi and Galimberti, 2017a). In both cases the learning gain is adjusted according to the recursive forecasting performance of the implied expectations, increasing/decreasing the gain (or switching from decreasing- to constant-gain and vice-versa) during periods of elevated/low forecasting errors. These time-varying approaches allow the modeling of behavioral shifts of attention that agents give to incoming data, which could be motivated as a concern with structural changes. One interesting implication of our results is that the learning gain does not need to increase for the weight given to the latest observation to rise; more formally, $\omega_{t,t} > \omega_{t,t-1}$ only requires that

$$\begin{aligned} \gamma_t &> \gamma_{t-1}(1-\gamma_t), \\ \gamma_t &> \frac{\gamma_{t-1}}{1+\gamma_{t-1}}. \end{aligned} \tag{9}$$

Gain-switching revisited Marcet and Nicolini (2003) propose an endogenous gain mechanism that can be represented in our framework as

$$\gamma_t^{gs} = \begin{cases} \frac{\bar{\gamma}}{1+s\bar{\gamma}} & if\, \mathscr{C} \\ \bar{\gamma} & otherwise, \end{cases} \tag{10}$$

where $s = t - t_s$ stands for the number of periods since the last time the constant-gain was used, and $\mathscr{C}$ for a regime-switching condition. Using equation (9) we find that a switch from the decreasing-gain, $\gamma_{t-1}^{gs} = \bar{\gamma}/(1+(s-1)\bar{\gamma})$, to the constant-gain regime, $\gamma_t^{gs} = \bar{\gamma}$, would always imply that the latest observation will be given a higher weight than the previous one—which is consistent with the structural change/tracking rationale given to this learning mechanism. When the switch goes on the other direction, from the constant-gain, $\gamma_{t-1}^{gs} = \bar{\gamma}$, to a decreasing-gain, $\gamma_t^{gs} = \bar{\gamma}/(1+\bar{\gamma})$, equation (9) turns into an equality, which means the new observation is given a weight that is equal to that given to the previous observation.

4 Concluding remarks

This note proposed a new and more general non-recursive representation of the recursive least squares algorithm, drawing renewed behavioral learning implications of alternative assumptions about the learning gains. One key finding is that, without a proper account for the learning initial, the estimation of models under the assumption of a constant gain over increasing samples of data would imply agents give a decreasing weight to more recent observations, distorting the relevance of learning in the determination of the latest economic developments being modeled. Another interesting finding is that the application of decreasing-gains to cohort-level data, an approach that has found empirical support from recent research on consumers' expectations survey data, can lead to the emergence of a U-shaped profile of weights, where data observed earlier in an agent's lifetime can have a comeback as dormant memories.

References

Berardi, M. and J. K. Galimberti (2013). A note on exact correspondences between adaptive learning algorithms and the kalman filter. *Economics Letters 118*(1), 139–142.

Berardi, M. and J. K. Galimberti (2014). A note on the representative adaptive learning algorithm. *Economics Letters 124*(1), 104 – 107.

Berardi, M. and J. K. Galimberti (2017a). Empirical calibration of adaptive learning. *Journal of Economic Behavior & Organization 144*, 219 – 237.

Berardi, M. and J. K. Galimberti (2017b). On the initialization of adaptive learning in macroeconomic models. *Journal of Economic Dynamics and Control 78*, 26 – 53.

Bray, M. (1982). Learning, estimation, and the stability of rational expectations. *Journal of Economic Theory 26*(2), 318–339.

Evans, G. W., S. Honkapohja, and N. Williams (2010). Generalized stochastic gradient learning. *International Economic Review 51*(1), 237–262.

Galimberti, J. K. (2019). An approximation of the distribution of learning estimates in macroeconomic models. *Journal of Economic Dynamics & Control 102*, 29–43.

Kostyshyna, O. (2012). Application of an adaptive step-size algorithm in models of hyperinflation. *Macroeconomic Dynamics 16*, 355–375.

Malmendier, U. and S. Nagel (2016). Learning from inflation experiences. *The Quarterly Journal of Economics 131*(1), 53–87.

Marcet, A. and J. P. Nicolini (2003). Recurrent hyperinflations and learning. *American Economic Review 93*(5), 1476–1498.

Marcet, A. and T. J. Sargent (1989). Convergence of least squares learning mechanisms in self-referential linear stochastic models. *Journal of Economic Theory 48*(2), 337–368.

Milani, F. (2014). Learning and time-varying macroeconomic volatility. *Journal of Economic Dynamics and Control 47*, 94–114.

Sargent, T. J. (1999). *The Conquest of American Inflation*. Princeton, NJ: Princeton University Press.

Williams, N. (2019). Escape dynamics in learning models. *The Review of Economic Studies 86*(2), 882–912.

A Online Appendix

A.1 Penalized WLS matrix derivation

It is sometimes useful to have estimators in matrix form. For the case with arbitrary initials, $\hat{\phi}_0$ and $\mathbf{R}_0$, the RLS is equivalent to the solution of a penalized regression problem given by

$$\hat{\phi}_t = \arg\min \left\{ \left(\mathbf{y}_t - \hat{\phi}_t'\mathbf{X}_t\right) \Omega_t \left(\mathbf{y}_t' - \mathbf{X}_t'\hat{\phi}_t\right) + \omega_{t,0} \left(\hat{\phi}_0' - \hat{\phi}_t'\right) \mathbf{R}_0 \left(\hat{\phi}_0 - \hat{\phi}_t\right) \right\},$$

where $\mathbf{X}_t = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t)$ is a $(k \times t)$ matrix collecting the regressors' time series, $\mathbf{y}_t = (y_1, y_2, \ldots, y_t)$, $\Omega_t = \mathrm{diag}\left(\omega_{t,1}, \omega_{t,2}, \ldots, \omega_{t,t}\right)$ is a weighting matrix with the sequence of weights in the main diagonal and zeros otherwise. The solution to this minimization problem is then given by

$$\hat{\phi}_t = \left(\mathbf{X}_t \Omega_t \mathbf{X}_t' + \omega_{t,0}\mathbf{R}_0\right)^{-1} \left(\mathbf{X}_t \Omega_t \mathbf{y}_t' + \omega_{t,0}\mathbf{R}_0\hat{\phi}_0\right).$$

A.2 Correspondence between penalized WLS and RLS

This appendix shows how the RLS of (2)-(3) can be derived from the penalized WLS formulation of (5) and (6) (using the paper's numbering), which are reproduced here for convenience:

$$\hat{\phi}_t = \hat{\phi}_{t-1} + \gamma_t \mathbf{R}_t^{-1}\mathbf{x}_t \left(y_t - \mathbf{x}_t'\hat{\phi}_{t-1}\right), \tag{2}$$

$$\mathbf{R}_t = \mathbf{R}_{t-1} + \gamma_t \left(\mathbf{x}_t\mathbf{x}_t' - \mathbf{R}_{t-1}\right), \tag{3}$$

$$\hat{\phi}_t = \left[\sum_{i=1}^{t} \omega_{t,i}\mathbf{x}_i\mathbf{x}_i' + \omega_{t,0}\mathbf{R}_0\right]^{-1} \left[\sum_{i=1}^{t} \omega_{t,i}\mathbf{x}_i y_i + \omega_{t,0}\mathbf{R}_0\hat{\phi}_0\right], \tag{5}$$

where

$$\omega_{t,i} = \begin{cases} \prod_{j=1}^{t}\left(1-\gamma_j\right) & for\ i = 0, \\ \gamma_i \prod_{j=i+1}^{t}\left(1-\gamma_j\right) & for\ 0 < i < t, \\ \gamma_t & for\ i = t. \end{cases} \tag{6}$$

First notice that iterating (3) recursively from $\mathbf{R}_0$ we have that

$$\mathbf{R}_t = \sum_{i=1}^{t} \omega_{t,i}\mathbf{x}_i\mathbf{x}_i' + \omega_{t,0}\mathbf{R}_0,$$

which is the inverse of the first term in (5), leading to

$$\hat{\phi}_t = \mathbf{R}_t^{-1}\left[\sum_{i=1}^{t} \omega_{t,i}\mathbf{x}_i y_i + \omega_{t,0}\mathbf{R}_0\hat{\phi}_0\right]. \tag{A.1}$$

For the second term notice that

$$\begin{aligned}\sum_{i=1}^{t}\omega_{t,i}\mathbf{x}_i y_i &= \sum_{i=1}^{t-1}\omega_{t,i}\mathbf{x}_i y_i + \gamma_t \mathbf{x}_t y_t,\\ &= (1-\gamma_t)\sum_{i=1}^{t-1}\omega_{t-1,i}\mathbf{x}_i y_i + \gamma_t \mathbf{x}_t y_t,\end{aligned}$$

and

$$\omega_{t,0}\mathbf{R}_0\hat{\phi}_0 = (1-\gamma_t)\,\omega_{t-1,0}\mathbf{R}_0\hat{\phi}_0,$$

where we use

$$\omega_{t,i} = (1-\gamma_t)\,\omega_{t-1,i},$$

which follows from (6). Hence, (A.1) is equivalent to

$$\hat{\phi}_t = \mathbf{R}_t^{-1}\left[\gamma_t \mathbf{x}_t y_t + (1-\gamma_t)\left(\sum_{i=1}^{t-1}\omega_{t-1,i}\mathbf{x}_i y_i + \omega_{t-1,0}\mathbf{R}_0\hat{\phi}_0\right)\right]. \tag{A.2}$$

Lagging (A.1) one period we find that

$$\mathbf{R}_{t-1}\hat{\phi}_{t-1} = \sum_{i=1}^{t-1}\omega_{t-1,i}\mathbf{x}_i y_i + \omega_{t-1,0}\mathbf{R}_0\hat{\phi}_0,$$

which can be substituted into (A.2) to yield

$$\hat{\phi}_t = \mathbf{R}_t^{-1}\left[\gamma_t \mathbf{x}_t y_t + (1-\gamma_t)\,\mathbf{R}_{t-1}\hat{\phi}_{t-1}\right]. \tag{A.3}$$

From (3) notice that

$$(1-\gamma_t)\,\mathbf{R}_{t-1} = \mathbf{R}_t - \gamma_t \mathbf{x}_t \mathbf{x}_t',$$

which substituted into (A.3) and after rearranging leads to

$$\begin{aligned}\hat{\phi}_t &= \mathbf{R}_t^{-1}\left[\gamma_t \mathbf{x}_t y_t + \left(\mathbf{R}_t - \gamma_t \mathbf{x}_t \mathbf{x}_t'\right)\hat{\phi}_{t-1}\right],\\ &= \gamma_t \mathbf{R}_t^{-1}\mathbf{x}_t y_t + \hat{\phi}_{t-1} - \gamma_t \mathbf{R}_t^{-1}\mathbf{x}_t \mathbf{x}_t'\hat{\phi}_{t-1},\\ &= \hat{\phi}_{t-1} + \gamma_t \mathbf{R}_t^{-1}\mathbf{x}_t\left(y_t - \mathbf{x}_t'\hat{\phi}_{t-1}\right),\end{aligned}$$

establishing the correspondence between the penalized WLS solution of (5) and the RLS of (2).

A.3 Absolute and relative weights

What in fact matters for estimation are the relative weights instead of their absolute values. To see that, consider the effects of multiplying the sequence of weights in the solution to the penalized WLS problem, $\{\omega_{t,0}, \omega_{t,1}, \ldots, \omega_{t,t}\}$ in equation (5), by a constant κ; clearly, such a re-scaling of the weights will have no effect over the resulting estimates because the constant factor entering in the numerator will cancel out with that entering in the denominator of the estimator. In order to calculate relative weights one needs to divide the absolute weights by their total. Letting W_t^n stand for the sum of weights starting from weight n up to weight t, from the definition of the absolute weights, (6), this sum of weights can be expanded according to

$$\begin{aligned} W_t^0 &= \sum_{i=0}^{t} \omega_{t,i}, \\ &= \prod_{j=1}^{t} (1-\gamma_j) + \sum_{i=1}^{t-1} \gamma_i \prod_{j=i+1}^{t} (1-\gamma_j) + \gamma_t. \end{aligned} \tag{A.4}$$

Expanding the first term of (A.4) we have that

$$\begin{aligned} \omega_{t,0} &= (1-\gamma_1)(1-\gamma_2)\ldots(1-\gamma_{t-1})(1-\gamma_t), \\ &= (1-\gamma_2)\ldots(1-\gamma_{t-1})(1-\gamma_t) - \gamma_1 \prod_{j=2}^{t} (1-\gamma_j), \\ &= 1-\gamma_t - \sum_{i=1}^{t-1} \gamma_i \prod_{j=i+1}^{t} (1-\gamma_j). \end{aligned} \tag{A.5}$$

Returning to (A.4) we then have

$$\begin{aligned} W_t^0 &= 1-\gamma_t - \sum_{i=1}^{t-1} \gamma_i \prod_{j=i+1}^{t} (1-\gamma_j) + \sum_{i=1}^{t-1} \gamma_i \prod_{j=i+1}^{t} (1-\gamma_j) + \gamma_t, \\ &= 1. \end{aligned}$$

Hence, in the context of the correspondence between the RLS and the penalized WLS outlined in this paper, the relative weights will be equal to their corresponding absolute weights.