Initial Guess

Renzo Miguel Caballero Rosas

April 17, 2020

Note: We compute the following results using **errorVsForecast.m** and **plot_epsilon.m**.

Boundedness of θ_t (we define θ_t^ε)

To guarantee an unique solution for the process X_t, we need boundedness in θ_t for $t \in [0, T]$. For our definition

$$\theta_t = \max\left(\theta_0, \frac{\alpha\theta_0 + |2\dot{p}_t|}{2\min(1-p_t, p_t)}\right), \tag{1}$$

we do not have a bound for θ_t if $p_t \to 0^+$ or $p_t \to 1^-$. However, if we ensure that $p_t \in [\varepsilon, 1-\varepsilon]$ for some $0 < \varepsilon < \frac{1}{2}$ and all $t \in [0, T]$, then $\theta_t < M(\varepsilon) < \infty$ for all $t \in [0, T]$. So, we define the corrected forecast

$$p_t^\varepsilon = \begin{cases} \varepsilon & \text{if} \quad p_t < \varepsilon \\ p_t & \text{if} \quad \varepsilon \leq p_t < 1-\varepsilon \\ 1-\varepsilon & \text{if} \quad p_t \geq 1-\varepsilon, \end{cases} \tag{2}$$

and the corrected (and bounded) drift coefficient

$$\theta_t^\varepsilon = \max\left(\theta_0, \frac{\alpha\theta_0 + 2|\dot{p}_t^\varepsilon|}{2\min(1-p_t^\varepsilon, p_t^\varepsilon)}\right). \tag{3}$$

Boundedness of θ_t (we define θ_t^ε)

Using our new definition θ_t^ε in the drift of our SDE:

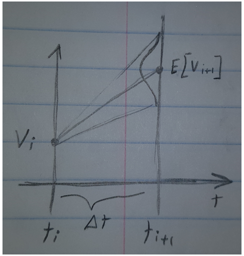
- We guarantee existence and uniqueness of the solution, and that $X_t \in]0,1[$ for all $t \in [0,1]$.
- As $X_t \in]0,1[$, the Lamperti transform is well defined for all $t \in [0,T]$ and we can use Itô's lemma to compute the SDE for the transformed process.

Least square minimization (LSM)

We consider the transition $\Delta V_i = V_{i+1} - V_i$ with $\Delta t = t_{i+1} - t_i$. $(V_{i+1}|V_i)$ is a random variable which conditional mean can be approximated by the solution of the system

$$\begin{cases} \mathrm{d}\mathbb{E}[V] & = -\theta_t^\varepsilon \mathbb{E}[V]\,\mathrm{d}t \\ \mathbb{E}\left[V(t_i)\right] & = V_i, \end{cases}$$

evaluated in t_{i+1} (i.e., $\mathbb{E}\left[V(t_{i+1})\right]$). Then, the random variable $(V_{i+1} - \mathbb{E}\left[V(t_{i+1})\right])$ has approximately zero mean.



If we assume that $\theta_t^\varepsilon = c \in \mathbb{R}^+$ for all $t \in [t_i, t_{i+1}]$, then $\mathbb{E}\left[V(t_{i+1})\right] = V_i e^{-c\Delta t}$. If we have a total of n transitions, we can write the regression problem for the conditional mean with L^2 loss function as

$$c^* \approx \arg\min_{c\geq 0} \left[\sum_{i=1}^{n} \left(V_{i+1} - \mathbb{E}\left[V(t_{i+1})\right] \right)^2 \right] = \arg\min_{c\geq 0} \left[\sum_{i=1}^{n} \left(V_{i+1} - V_i e^{-c\Delta t} \right)^2 \right]. \qquad (4)$$

Least square minimization (we are assuming $\theta_t^{\varepsilon} = c \in \mathbb{R}^+$)

We take the first order approximation w.r.t. c

$$e^{-c\Delta t} = 1 - c\Delta t + \mathscr{O}\left((c\Delta t)^2\right),$$

and introduce it in equation (4). We get

$$c^* \approx \arg\min_{c \geq 0} \underbrace{\left[\sum_{i=1}^{n} \left(V_{i+1} - V_i(1 - c\Delta t)\right)^2\right]}_{:=f(c)}. \tag{5}$$

As $f(c)$ is convex in c, solving (5) (finding c^*) is equivalent to solving

$$\frac{\partial f}{\partial c}(c^{**}) = 0, \tag{6}$$

and choosing $c^* = \max\{0, c^{**}\}$.

Least square minimization (we are assuming $\theta_t^\varepsilon = c \in \mathbb{R}^+$)

$$
\begin{aligned}
\frac{\partial f}{\partial c} &= \sum_{i=1}^{n} 2(-V_i)(-\Delta t)(V_{i+1} - V_i(1-\theta_0 \Delta t)) \\
&= \sum_{i=1}^{n} 2V_i \Delta t(V_{i+1} - V_i(1-c\Delta t)) \\
&= \sum_{i=1}^{n} 2V_{i+1}V_i\Delta t - 2V_i^2\Delta t + 2V_i^2\Delta t^2 c.
\end{aligned}
$$

Then, c^{**} satisfies

$$
c^{**} \approx \frac{\sum_{i=1}^{n} V_i(V_i - V_{i+1})}{\Delta t \cdot \sum_{i=1}^{n}(V_i)^2}. \tag{7}
$$

Notice that c^{**} has dimension $\textbf{time}^{-1}$.

Quadratic variation

We approximate the SDE by its E-M scheme. In particular, we approximate the Itô quadratic variation with the discrete one:

- Itô process quadratic variation: $[V]_t = \int_0^t \sigma_s^2 \, \mathrm{d}s$.
- Discrete process quadratic variation: $[V]_t = \sum_{0<s\leq t} (\Delta V_s)^2$.

Then, considering Δt the time between measurements, we approximate:

$$\theta_0^* \alpha^* \approx \frac{\sum_{i=1}^n (\Delta V_i)^2}{2 \Delta t \sum_{i=1}^n (V_i + p_i)(1 - V_i - p_i)}. \tag{8}$$

Estimation of $(\theta_0, \alpha, \varepsilon)$

In slide (2) we defined the parameter ε which has as only condition $0 < \varepsilon < \frac{1}{2}$. We want to estimate the set of parameters $(\theta_0, \alpha, \varepsilon)$, we call to the estimations $(\theta_0^*, \alpha^*, \varepsilon^*)$.

Recall from slide (4) that, to use the LSM estimation, we assume $\theta_t^\varepsilon = c \in \mathbb{R}^+$. Recall also our definition for θ_t^ε (equation (3)):

$$\theta_t^\varepsilon = \max\left(\theta_0, \frac{\alpha\theta_0 + 2|\dot{p}_t^\varepsilon|}{2\min(1 - p_t^\varepsilon, p_t^\varepsilon)}\right).$$

Fixed ε, we call $\mathscr{V} = \{\Delta V_i^\varepsilon\}_{i=1}^n$ to the set of n error transitions where for each measurement X_i, we have that $V_i = X_i - p_i^\varepsilon$. $\mathscr{P} = \{p_i^\varepsilon\}_{i=1}^n$ is the corresponding set of forecasts.
Now, if we also fix θ_0 and α, we can define the set of indexes
$\mathbf{I} = \{i \in \{1, \dots, n\} : \text{ the LSM estimation will estimate } \theta_0\}$ and
$\mathbf{J} = \left\{j \in \{1, \dots, n\} : \text{ the LSM estimation will estimate } \frac{\theta_0\alpha}{\varepsilon}\right\}$.

Estimation of $(\theta_0, \alpha, \varepsilon)$

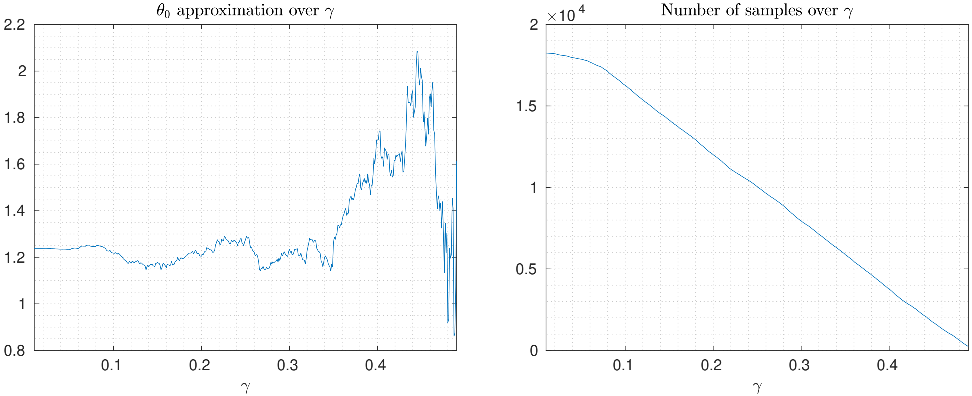
From the definition of θ_t^ε: We have that for $\varepsilon << 1$, and $p_t = \varepsilon$ or $p_t = 1-\varepsilon$, the approximation $\theta_t^\varepsilon \approx \frac{\theta_0 \alpha}{\varepsilon}$ holds. Then, for ε small enough, $\mathbf{J}$ can be approximated by

$$\mathbf{J} \approx \tilde{\mathbf{J}} = \{j \in \{1, \ldots, n\} : p_j^\varepsilon \in \{\varepsilon, 1-\varepsilon\}\}.$$

Now, coming back to the definition of θ_t^ε, we have that it is more likely that $\theta_t^\varepsilon = \theta_0$ if $p_t^\varepsilon \approx \frac{1}{2}$. Then, we can approximate $\mathbf{I}$ by

$$\mathbf{I} \approx \tilde{\mathbf{I}} = \left\{i \in \{1, \ldots, n\} : p_i \in (\gamma, 1-\gamma)\right\}, \ \gamma \approx \frac{1}{2}, \ \gamma < \frac{1}{2}.$$

Estimation of $(\theta_0, \alpha, \varepsilon)$: θ_0^*



In slide (9), we showed that for $\gamma \approx \frac{1}{2}$, the LSM estimation using indetex from $\tilde{\mathbf{I}}$ is an estimator for θ_0. On the left, we can see θ_0^* as a function of γ. The choose $\theta_0^* \approx 1.25$ seems reasonable from the plot.

Estimation of $(\theta_0, \alpha, \varepsilon)$: $\theta_0^* \alpha^*$ and α^*

Recall the quadratic variation estimator (8) from slide (7). It uses all the transitions $\mathscr{V}$ and estimated the product $\theta_0 \alpha$.

We get $\theta_0^* \alpha^* \approx 0.10$, from where we calculate $\alpha^* \approx 0.08$.

Estimation of $(\theta_0, \alpha, \varepsilon)$: ε^*

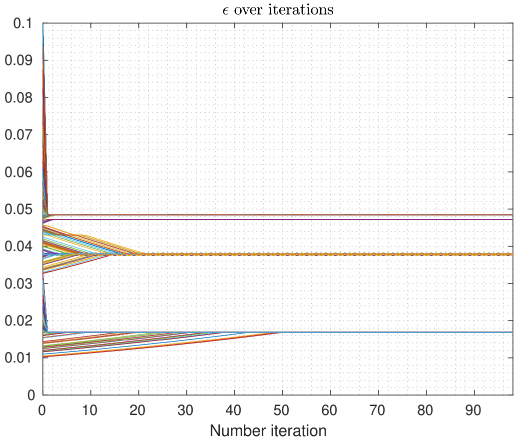
As we have an approximated value for $\theta_0\alpha$, if we can estimate $\frac{\theta_0\alpha}{\varepsilon}$, then we can estimate ε. In slide (9), we showed that for $\varepsilon << 1$, the LSM estimation using indexes from $\tilde{\mathbf{J}}$ is an estimator for $\frac{\theta_0\alpha}{\varepsilon} =: k$.

However, which values of ε satisfies $\varepsilon << 1$? We choose an initial guess for ε (call it ε_0), and iterating we aim to converge to some local minimum. We proceed with the following steps:

- We sample ε_0 from $\mathscr{U}[0.01, 0.1]$ and load $\varepsilon \leftarrow \varepsilon_0$.
- We create $\tilde{\mathbf{J}}$ and use the LSM estimation to find k. If $k < \theta_0^*$, then the assumption $\theta_t^\varepsilon = c \in \mathbb{R}^+$ is wrong and we reduce the value of ε, i.e., $\varepsilon \leftarrow \varepsilon * 0.999$. If $k \geq \theta_0^*$, we load $\varepsilon \leftarrow \frac{\theta_0^* \alpha^*}{k}$ (we allow a maximum relative change of 1%). We repeat this step *100 times*.
- We repeat steps 1 and 2 *500 times*. After, we plot the results.

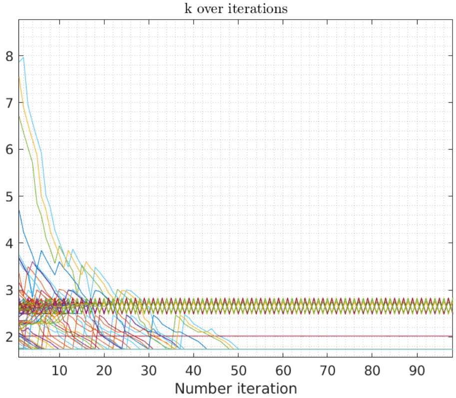
Estimation of $(\theta_0, \alpha, \varepsilon)$: ε^*

We can see that we converge to four different local minimums.



Estimation of $(\theta_0, \alpha, \varepsilon)$: ε^*

This is the value of k over iterations.
When $k < \theta_0^*$, we assume that the estimation and repeat the iteration step.

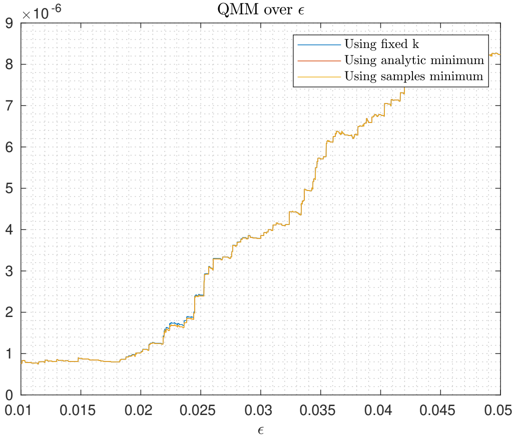


Estimation of $(\theta_0, \alpha, \varepsilon)$: ε^*

In this plot, we can see the minimum of the normalized value function (4) over ε. We normalize it with respect to the number of elements of $\tilde{\mathbf{J}}$.

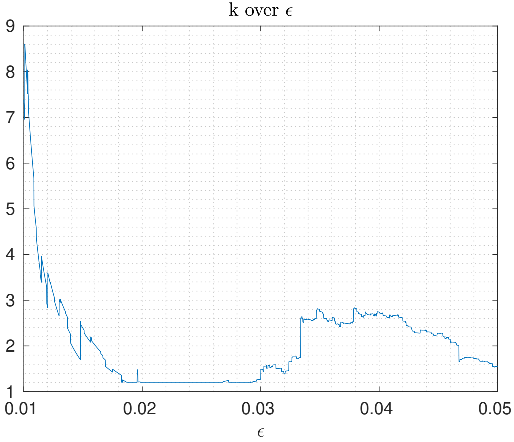
- Using fixed k: As $k < \theta_0^*$ has not sense, when we get that condition we set $k \leftarrow \theta_0^*$, and evaluate the value function in k.
- Analytical optimal value function.
- Numerical optimal value function.

We can see that all plots almost coincide. We choose $\varepsilon^* \approx 0.018$.



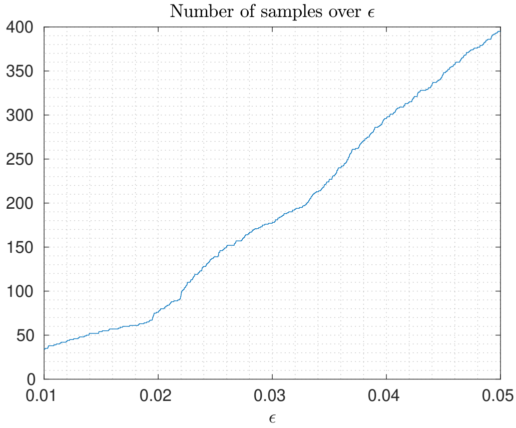
Estimation of $(\theta_0, \alpha, \varepsilon)$: ε^*

When we get $k < \theta_0^*$, we set $k \leftarrow \theta_0^*$.

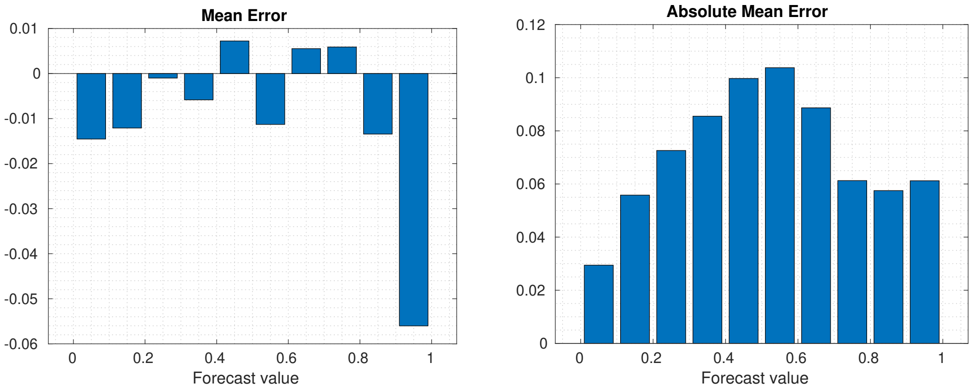


Estimation of $(\theta_0, \alpha, \varepsilon)$: ε^*

Number of elements in $\tilde{\mathbf{J}}$ as a function of ε.

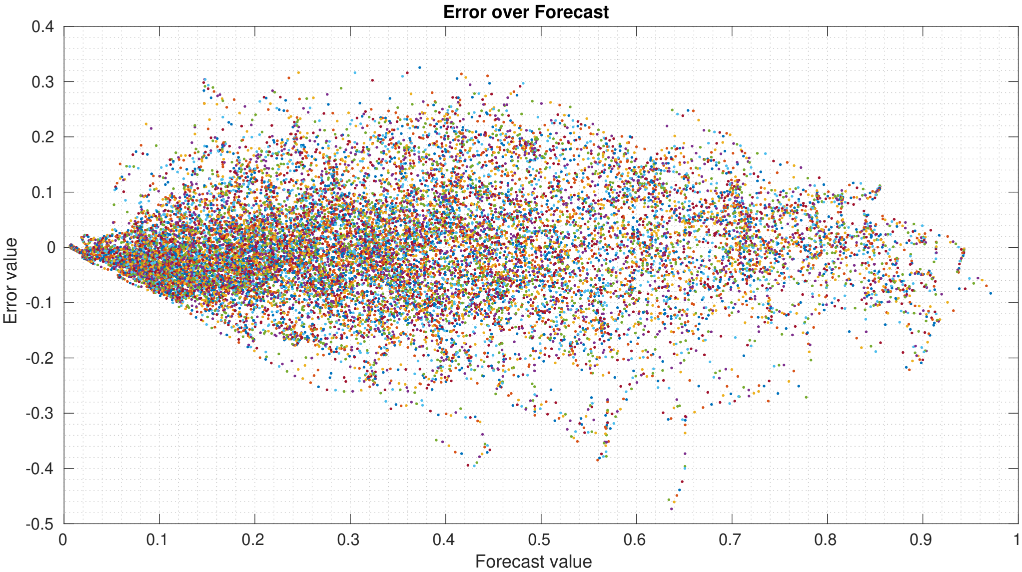


Extra: Interesting data processing



What we are seeing is the **mean error** and **mean absolute error** as a function of the forecast. This is, for each interval with length 0.1 (i.e., [0,0.1), [0.1,0.2), etc.), we average all the errors corresponding to measurement where the forecast was in that intervals, and after we average over the number of elements in each interval. In some future, we can use this information to construct an even more realistic model.

Extra: Error Vs. Forecast for all training days



How do we ensure $\dot{p}_i^\varepsilon = 0$ in $\tilde{\mathbf{J}}$?

To guarantee that $\dot{p}_{t_i}^\delta = 0$, we need that $p_{t_i} = p_{t_{i+1}}$. Then, if we have $n+1$ truncated consecutive forecasts (i.e., $p_{t_i}, \dots, p_{t_{i+n}}$), we can use the n transitions ΔV_{t_i} because $\dot{p}_{t_i}^\delta = \dots = \dot{p}_{t_{i+n-1}}^\delta = 0$, but $\dot{p}_{t_{i+n}}^\delta \neq 0$.