دانشکده ریاضی و علوم کامپیوتر
دانشگاه صنعتی شریف

فیلتر کالمن (Kalman Filter)

گزارش مطالعه و پیاده‌سازی فیلتر کالمن

پروژه درس بهینه‌سازی در علوم داده

دکتر مجتبی تفاق، ترم دوم سال تحصیلی ۱۴۰۱

نگارندگان: محسن قدرتی، محمد صالح بهرامی

۱ مقدمه

۲ تعاریف اولیه

۱.۲ مدل‌های اتوریگرسیو

۲.۲ فیلتر

۳ معرفی مساله داده‌کاوی

۱.۳ صورت مساله و داده‌ها

۲.۳ تخمین حافظه موثر

۳.۳ تخمین سایز مناسب برای یادگیری دینامیک مساله

۴.۳ تهیه تعدادی ویژگی (کنترل دینامیک)

۴ فیلتر کالمن

۱.۴ تعریف

۲.۴ الگوریتم

۳.۴ تحلیل زمانی

۴.۴ پیاده‌سازی

۱.۴.۴ مدل ۱: دینامیک خطی بدون کنترل

فرض می‌کنیم دینامیک تغییرات قیمت به شکل زیر باشد:

$$\tilde{r}_n = f_{0,n}\tilde{r}_{n-1} + \sum_{i=1}^{m-1} f_{i,n} r_{i,n-1} + f_{m,n} + w_n, \qquad w_n \sim \mathcal{N}(0,\ \sigma_n^2). \tag{1}$$

به عبارت دیگر، فرض می‌کنیم اگر تغییرات قیمت از $i = 2, \ldots, m-1$ کندل گذشته تا کندل گذشته را بدانیم و تغییرات قیمت از کندل گذشته تا کندل فعلی را نیز بدانیم، بازدهی پیش رو ترکیبی خطی از این مقادیر به علاوه یک نویز با توزیع نرمال است. در این صورت برای بردار $\mathbf{x^{(n)}}$ فرم معادل زیر را خواهیم داشت:

$$\mathbf{x^{(n)}} = F_n \mathbf{x^{(n-1)}} + \mathbf{w^{(n)}}, \tag{2}$$

که در آن

$$\mathbf{x^{(n)}} := \begin{bmatrix} \tilde{r}_n \\ r_{1,n} \\ \vdots \\ r_{m-1,n} \\ 1 \end{bmatrix}, \quad F_n := \begin{bmatrix} f_{0,n} & f_{1,n} & \cdots & f_{m,n} \\ 1+\tilde{r}_{n-1} & 0 & \ddots & 0 \\ 0 & 1+\tilde{r}_{n-1} & 0 & \vdots \\ 0 & \ddots & & 0 \\ \vdots & 0 & 1+\tilde{r}_{n-1} & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}, \quad \mathbf{w^{(n)}} := \begin{bmatrix} w_n \\ -\tilde{r}_{n-1}^2 \\ \tilde{r}_{n-1} \\ \vdots \\ \tilde{r}_{n-1} \\ 0 \end{bmatrix}.$$

با توجه به این که در هر لحظه n، بردار $\mathbf{x^{(n-1)}}$ معلوم است و کواریانس بردار $\mathbf{w^{(n)}}$ تنها در درایه $1,1$ ناصفر و برابر با σ_n^2 است، به دنبال مقادیری از $f_{0,n}, \ldots, f_{m-1,n}$ می‌گردیم که σ_n^2 را کمینه کند. (در واقع خطای مدل را در حد امکان کاهش دهد.) هم‌چنین توجه می‌کنیم که اضافه کردن عدد ۱ به انتهای بردار $\mathbf{x^{(n)}}$ باعث می‌شود بتوان عرض از مبدا مدل ۱ را به $f_{m,n}$ منتقل کرد و لذا فرض صفر بودن میانگین نویز w_n برقرار است. اکنون سعی می‌کنیم با بازنویسی معادله بالا، فرم مجموع مربعاتی برای یافتن بهترین $f_{i,n}$ها ارایه کنیم.

اگر قرار دهیم $\theta_n = (f_{0,n}, \ldots, f_{m,n})^T$، به دنبال حل مساله بهینه‌سازی زیر هستیم:

$$minimize_{(\theta_n)} \quad |\tilde{r}_n - \langle \theta_n,\, \mathbf{x^{(n-1)}} \rangle|^2$$

اما با توجه به اینکه مساله بالا احتمالا بی‌نهایت جواب دارد و این موضوع باعث تغییرات سریع دینامیک مدل در هر لحظه می‌شود که خلاف طبیعت مدل است، و مهم‌تر آنکه انتظار داریم دینامیک مدل در طول یک پنجره کوتاه مدت (kتایی) پایدار باشد (یعنی $\theta_n \simeq \theta_{n-1} \simeq \cdots \simeq \theta_{n-(k-1)}$) و بتوانیم از اطلاعات این پنجره برای تقریب یک دینامیک کم‌نوسان بهره ببریم، مساله بهینه‌سازی فوق را به مساله زیر تعمیم می‌دهیم:

$$minimize_{(\theta_n)} \quad |\tilde{r}_{n-(k-1)} - \langle \theta_n,\, \mathbf{x^{(n-1-(k-1))}} \rangle|^2 + \cdots + |\tilde{r}_n - \langle \theta_n,\, \mathbf{x^{(n-1)}} \rangle|^2$$

که با در نظر گرفتن ماتریس $X^{(n)}$ به صورت

$$X^{(n)} = \begin{bmatrix} \mathbf{x^{(n-(k-1))}}^T \\ \vdots \\ \mathbf{x^{(n)}}^T \end{bmatrix} \quad \to \quad X^{(n)} := \begin{bmatrix} \mathbf{b_0^{(n)}} & \dots & \mathbf{b_m^{(n)}} \end{bmatrix}$$

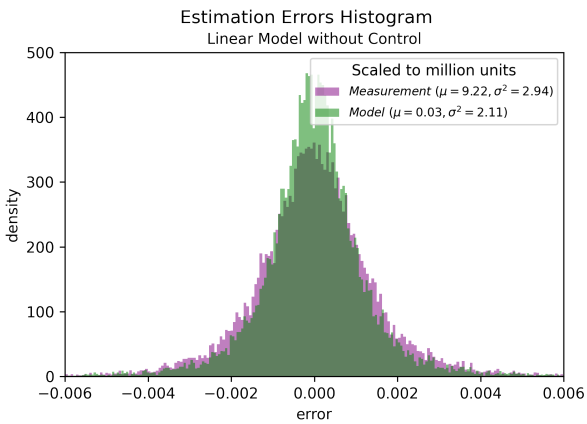
به فرم زیر قابل بیان است:

$$for\ given\ n: \quad minimize_{(\theta_n)} \; ||X^{(n-1)}\theta_n - \mathbf{b_0^{(n)}}||_2^2 \tag{3}$$

توجه به این نکته ضروری است که از بردار $X_0^{(n)}$، درایه اول مربوط به بازدهی آینده است و در درسترس نیست. لذا به جای حل مساله فوق برای n آن را برای $n-1$ حل می‌کنیم.

در ادامه به بررسی نتایج حاصل از حل مساله بهینه‌سازی ۳ به کمک روش کمترین مجموع مربعات و پیاده‌سازی فیلتر کالمن بر روی مدل ۱ می‌پردازیم.

مشاهده ۱. مطابق با نمودار زیر می‌توان فرض نرمال بودن نویز در اندازه‌گیری و مدل را تا حدودی تایید کرد. هم‌چنین توجه می‌کنیم که میانگین توزیع نویز، با دقت ۷ رقم اعشار صفر خواهد بود.



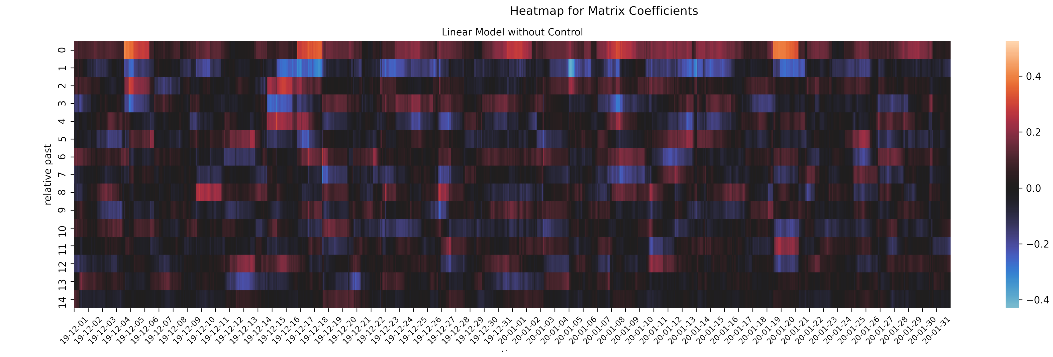
مشاهده ۲. همان‌طور که اشاره کردیم، فرض می‌کنیم دینامیک مدل دچار تغییرات سریع نمی‌شود. این نکته در نمودار زیر به صورت طیف‌های پیوسته قابل مشاهده است. هم‌چنین در نمودار زیر این نکته قابل توجه وجود دارد که ضرایب بزرگ‌تر و تاثیرگذارتر در مدل خطی پیشنهاد شده، به لحظات نزدیک‌تر به لحظه حال مرتبط می‌شوند. این نیز مشاهده‌ای است که انتظارش را داشتیم.

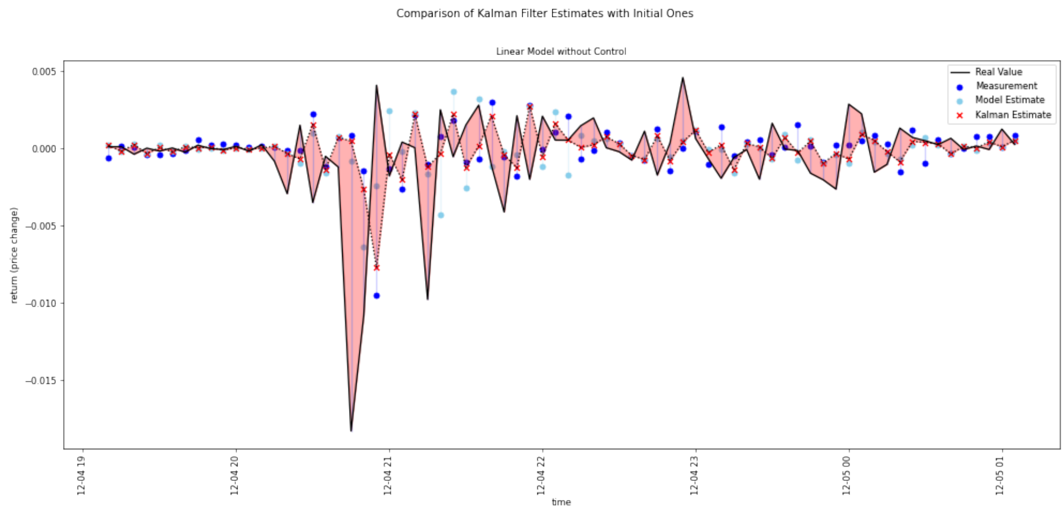
نتیجه ۳. در شکل می‌توان خروجی فیلتر کالمن را بر روی مدل خطی در هر لحظه به ازای یک پنجره خاص زمانی مشاهده کرد.

۲.۴.۴ مدل ۲: دینامیک خطی همراه با کنترل

در مدل زیر، فرض می‌کنیم تعدادی ویژگی (که در بخش قبل تهیه شده‌اند و در همان بخش، همبستگیشان با تابع هدف ما بررسی و تایید شده است) نیز علاوه بر بردار $\mathbf{x^{(n-1)}}$، در دینامیک تغییرات قیمت به شکل خطی موثر باشند. به عبارت دیگر دینامیک مساله را با مدل زیر تقریب می‌زنیم:

$$\tilde{r}_n = f_{0,n}\tilde{r}_{n-1} + \sum_{i=1}^{m-1} f_{i,n} r_{i,n-1} + f_{m,n} + \sum_{i=0}^{s-1} g_{i,n} u_{i,n} + w_n, \qquad w_n \sim \mathcal{N}(0,\, \sigma_n^2). \tag{4}$$





که با فرم زیر معادل است:

$$\mathbf{x^{(n)}} = F_n \mathbf{x^{(n-1)}} + G_n \mathbf{u^{(n)}} + \mathbf{w^{(n)}}, \tag{۵}$$

و در آن $\mathbf{u^{(n)}}$ بردار ویژگی‌ها در لحظه n است. در واقع:

$$G_n := \begin{bmatrix} g_{0,n} & \cdots & g_{s-1,n} \\ 0 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & 0 \end{bmatrix}, \quad \mathbf{u^{(n)}} := \begin{bmatrix} u_{0,n} \\ \vdots \\ u_{s-1,n} \end{bmatrix}.$$

با در نظر گرفتن ماتریس $X^{(n)}$ و بردار $\mathbf{b_0^{(n)}}$ مدل قبلی، و تعریف ماتریس U_n و بردار θ'_n به شکل زیر،

$$U_n := \begin{bmatrix} \mathbf{u^{(n-(k-1))}}^T \\ \vdots \\ \mathbf{u^{(n)}}^T \end{bmatrix}, \qquad \theta'_n := \begin{bmatrix} g_{0,n} \\ \vdots \\ g_{s-1,n} \end{bmatrix}$$

می‌توان به سادگی و مشابه با روند مدل ۱، به مساله کمترین مجموع مربعات زیر دست یافت:

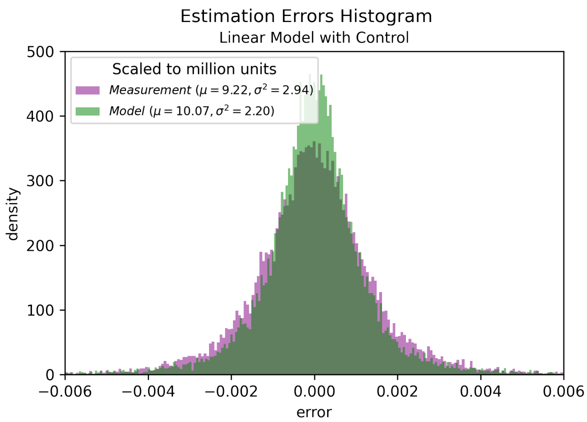
$$for\ given\ n: \quad minimize_{(\theta_n, \theta'_n)} \ ||X^{(n-1)}\theta_n + U_n \theta'_n - \mathbf{b_0^{(n)}}||_2^2 \tag{۶}$$

اکنون با ادغام دو ماتریس $X^{(n-1)}$ و U_n و ادغام دو بردار θ_n و θ'_n مساله بهینه‌سازی معادل زیر را داریم:

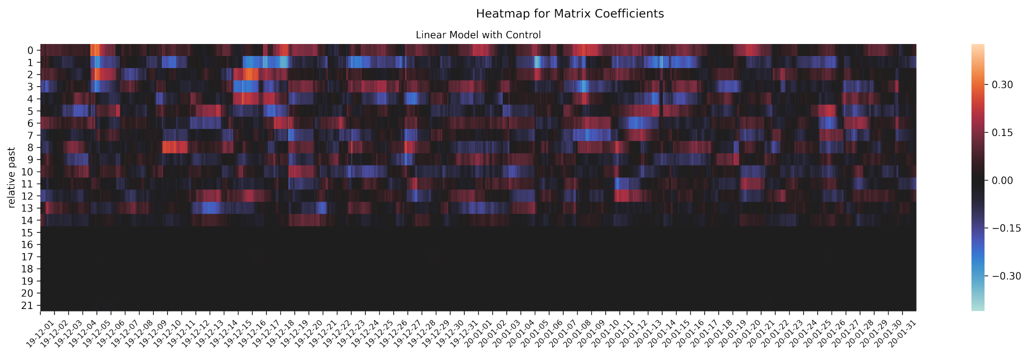
$$A_n := \begin{bmatrix} X^{(n-1)} & U_n \end{bmatrix}, \quad \theta^*_n := \begin{bmatrix} \theta_n \\ \theta'_n \end{bmatrix} \qquad \longrightarrow \qquad for\ given\ n: \quad minimize_{(\theta^*_n)} \ ||A_n \theta^*_n - \mathbf{b_0^{(n)}}||_2^2$$

در ادامه به بررسی نتایج حاصل از حل مساله بهینه‌سازی ۶ به کمک روش کمترین مجموع مربعات و پیاده‌سازی فیلتر کالمن بر روی مدل ۴ می‌پردازیم.

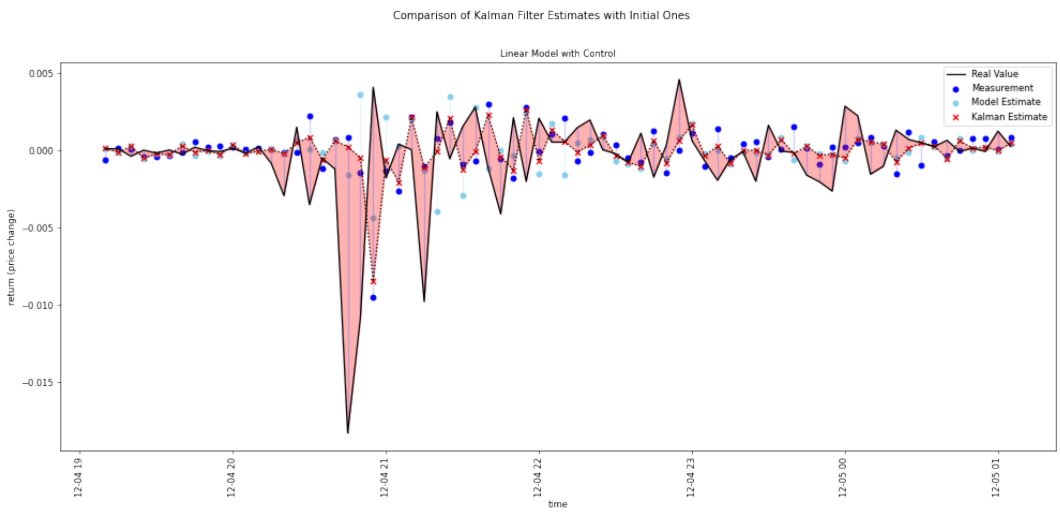
مشاهده ۴. مطابق با نمودار زیر می‌توان فرض نرمال بودن نویز در اندازه‌گیری و مدل را مجددا تا حدودی تایید کرد. هرچند همانطور که از نمودار پیداست، واریانس مدل کمی افزایش (جزیی) داشته و کم نشده است! در بخش‌های بعدی، زمانی که همه مدل‌های خطی را معرفی کردیم و به مقایسه عملکرد آن‌ها پرداختیم، دلیل عملیاتی این موضوع را توضیح خواهیم داد.



مشاهده ۵. نمودار زیر مشابه با نمودار مربوط به مدل ۱ قابل قبول است. نکته حایز اهمیت در این نمودار، تباهیده شدن ضرایب مربوط به ویژگی‌هاست که در واقع به دلیل انحراف معیار بالاتر ستون‌های مربوط به ویژگی‌ها نسبت به ستون‌های مربوط به بازدهی است و اطلاعاتی از دست نرفته است.



نتیجه ۶. در شکل زیر می‌توان خروجی فیلتر کالمن را بر روی مدل خطی در هر لحظه به ازای یک پنجره خاص زمانی مشاهده کرد.



۳.۴.۴ مدل ۳: دینامیک خطی حساس به زمان

منظور از مدل حساس به زمان، مدلی است که به نوعی به اطلاعات اخیرتر بیشتر توجه دارد (مثلا توسط یک ماتریس وزن‌دهی). همانطور که می‌توان حدس زد، چنین مدلی اساسا واریانس نویز بیشتری دارد. اما در عوض ممکن است به کمک فیلتر کالمن بتوان نویز آن را بهتر کنترل کرد، چرا که چنین مدلی در صورت خطی بودن واقعیت دینامیک مساله، سریع‌تر بروزرسانی می‌شود.

به دو روش می‌توان حساسیت به زمان را در مدل منعکس کرد. یکی حساسیت در یادگیری پارامترها، و دیگری حساسیت در توجه به اطلاعات سری زمانی.

روش اول معادل است با مساله بهینه‌سازی زیر:

$$for\ given\ n: \quad minimize_{(\theta_n^*)}\ ||D_1 A_n \theta_n^* - D_1 \mathbf{b_0^{(n)}}||_2^2, \quad D_1 = \frac{1}{\lambda} diag(\alpha^0,\ \dots,\ \alpha^{k-1}),\ \lambda = \sum_{i=0}^{k-1} \alpha^i \qquad (۷)$$

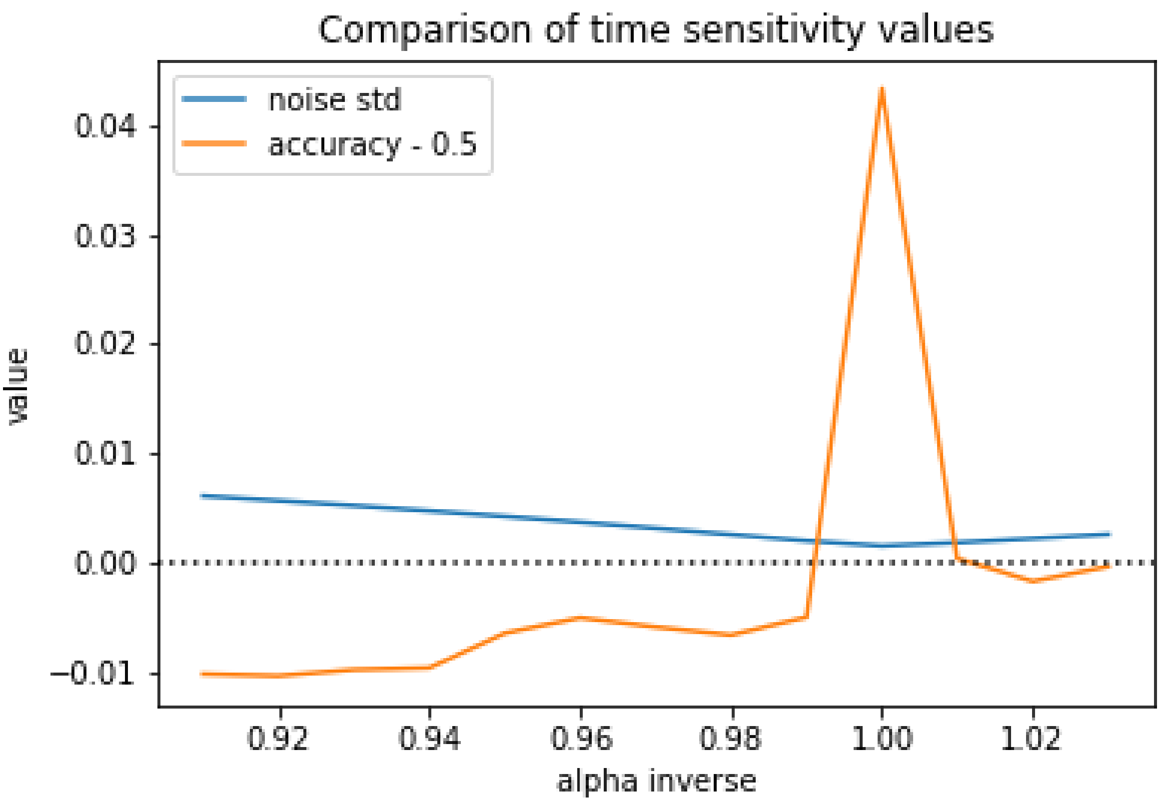
که در آن α پارامتر تاثیرگذاری است و مقداری حداقل برابر با ۱ دارد. پارامتر α در واقع بیان می‌کند کوچک کردن هر مقدار باقیمانده w_j در مساله مجموع کمترین مربعات، چه اندازه از کوچک کردن مقدار باقیمانده w_{j-1} مهم‌تر است.

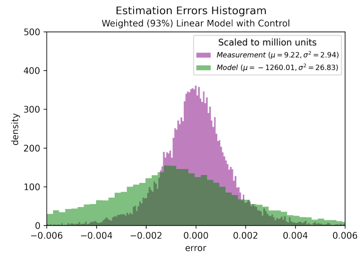
روش دوم معادل است با وزن‌دهی به درایه‌های بردار θ_n به گونه‌ای که هرچه اندیس i در $f_{i,n}$ افزایش می‌یابد، $|f_{i,n}|$ کاهش یابد. پس باید وزن‌های بزرگ‌تری به $f_{i,n}$های با اندیس بزرگ‌تر در مساله بهینه‌سازی ۶ بدهیم. یعنی مساله بهینه‌سازی زیر:

$$for\ given\ n: \quad minimize_{(\theta_n^*)}\ |\left[X^{(n-1)} D_2 \quad U_n\right] \theta_n^* - \mathbf{b_0^{(n)}}||_2^2, \quad D_2 = \frac{1}{\lambda} diag(\alpha^0,\ \dots,\ \alpha^m),\ \lambda = \sum_{i=0}^{m} \alpha^i \qquad (۸)$$

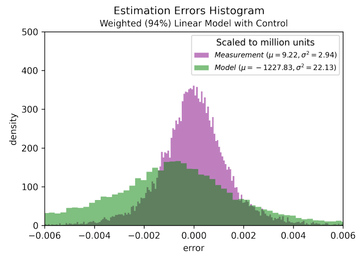
در ادامه تنها نتیجه حل مساله بهینه‌سازی ۷ آمده است:

مشاهده ۷. ابتدا به ازای مقادیر مختلف α با بررسی دقت پیش‌بینی (صرفا درستی علامت $\tilde{r}_n$) و انحراف معیار نویز مدل، بهترین α را انتخاب می‌کنیم:

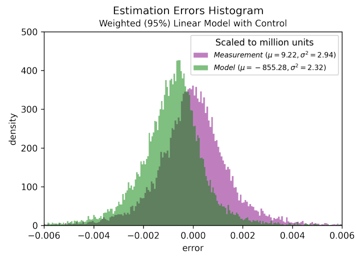




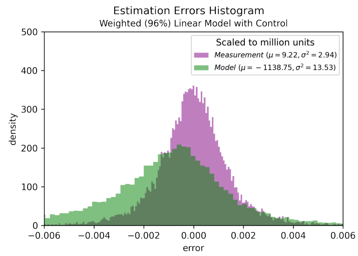
(آ) $\alpha^{-1} = ۰٫۹۳$



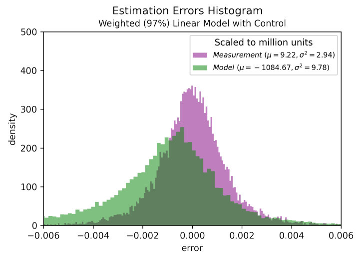
(ب) $\alpha^{-1} = ۰٫۹۴$



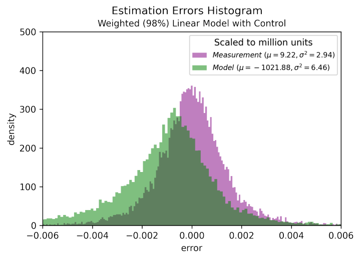
(ج) $\alpha^{-1} = ۰٫۹۵$



(د) $\alpha^{-1} = ۰٫۹۶$



(ه) $\alpha^{-1} = ۰٫۹۷$

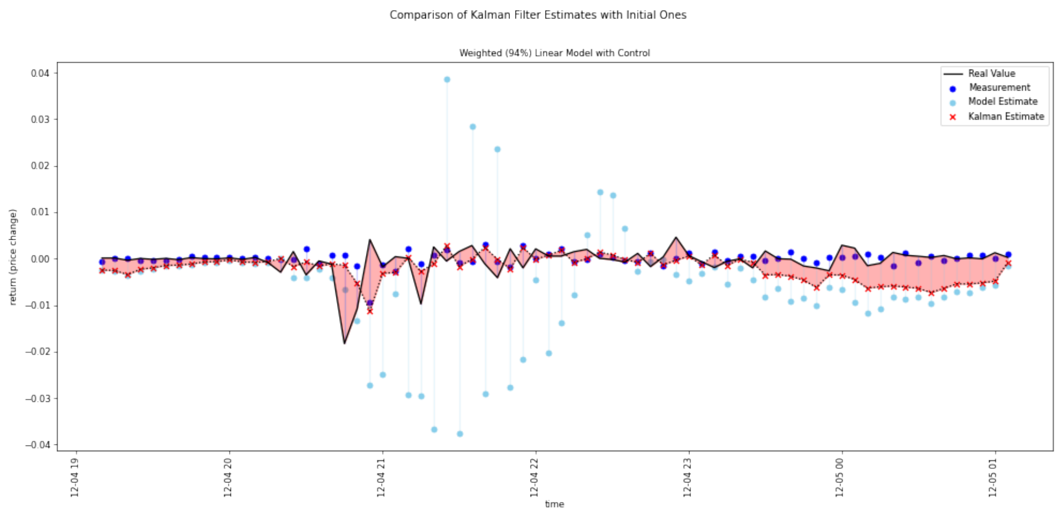


(و) $\alpha^{-1} = ۰٫۹۸$

بر اساس شکل در می‌یابیم نقاطی که در پیش‌بینی جهت $\tilde{r}_n$ بیشترین دقت را دارند (از تشخیص تصادفی فاصله دارند و در عین حال انحراف معیار نویزشان کم است) عبارت‌اند از ۰٫۹۴ و ۱. از طرفی کمترین انحراف معیار نویز، همانطور که پیش بینی شده بود، در حوالی ۱ رخ می‌دهد. پس اگر به دنبال مقداری غیر بدیهی (مخالف ۱) هستیم، $\alpha = ۰٫۹۴$ گزینه مناسبی است. (توجه می‌کنیم که نیمه عمر چنین انتخابی حدود ۱۲ است، یعنی حساسیت مدل به داده هر لحظه حدود دو برابر حساسیت آن به داده ۱۲ کندل قبل است!)

هم‌چنین در ادامه تصویری از هیستوگرام نویز تعدادی از انتخاب‌های α قابل مشاهده است.

نتیجه ۸. در شکل زیر می‌توان خروجی فیلتر کالمن را بر روی مدل خطی حساس به زمان (با $\alpha^{-1} = ۰٫۹۴$) در هر لحظه به ازای یک پنجره خاص زمانی مشاهده کرد.



۴.۴.۴ مدل ۴: دینامیک خطی با ماتریس چگال

در مدل‌های قبل، برای سهولت در محاسبه و سادگی مدل، فرض کردیم ماتریس‌های F_n و G_n تنک هستند. در این مدل، بدون تغییر فرض خطی بودن دینامیک مساله، فرض تنک بودن F_n و G_n را حذف می‌کنیم. در این صورت هرچند واریانس نویز تقریب آخرین بازدهی افزایش می‌یابد، اما با توجه به اینکه حالت تنک معادله خطی زیر، یک حالت خاص از معادله کلی است، انتظار داریم مساله بهینه‌سازی بتواند جواب‌های پایدارتری تولید کند. فایده دستگاه‌های خطی با واریانس نویز w_n بالاتر ولی پایداری بیشتر ان است که در بازارهای مالی به طور طبیعی مقدار بسیار زیادی نویز غیر قابل کنترل وجود دارد، و با پایدارتر کردن مدل، به ذات دینامیک مساله بهتر دسترسی داریم. (با فرض شبیه به خطی بودن این دینامیک

$$\mathbf{x}^{(\mathbf{n})} = F_n \mathbf{x}^{(\mathbf{n-1})} + G_n \mathbf{u}^{(\mathbf{n})} + \mathbf{w}^{(\mathbf{n})}$$

۵ فیلتر کالمن تعمیم‌یافته

۱.۵ تعریف

۲.۵ پیاده‌سازی

۱.۲.۵ مدل ۵: دینامیک درجه ۲ بدون کنترل

۲.۲.۵ مدل ۶: دینامیک درجه ۲ همراه با کنترل خطی

۳.۲.۵ مدل ۷: دینامیک شبکه عصبی

۴.۲.۵ مدل ۸: دینامیک مشتق‌ناپذیر

۶ فیلتر کالمن سریع

۷ فیلتر کالمن با چند مدل

۱.۷ جمع‌بندی مدل‌های ارایه شده

۲.۷ تعمیم فیلتر کالمن به چند مدل موازی

۱.۲.۷ تعمیم ضرایب کالمن

۲.۲.۷ ترکیب چند فیلتر

۸ ارجاع و منابع

[?]