Filtering bond and credit default swap markets

Peter Cotton

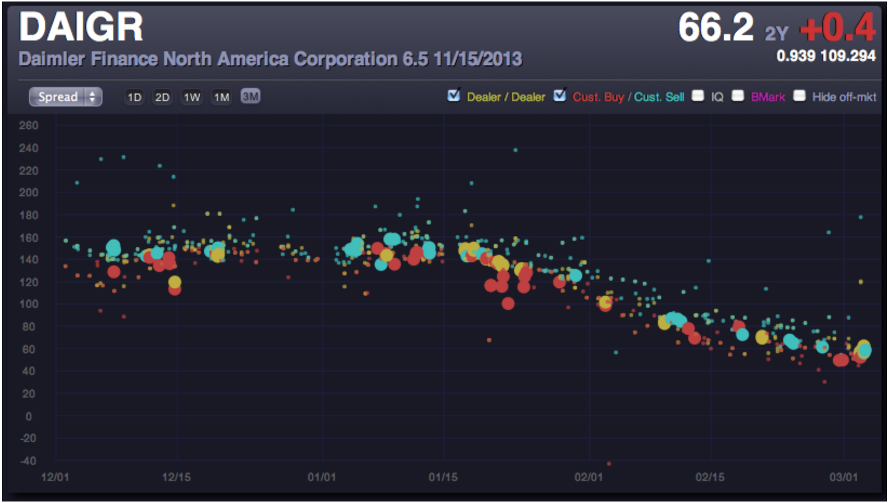
April 27, 2016

Overview

Fixing an inconsistency
Multivariate filters seemingly perform worse then univariate
Serial correlation distorts accuracy metrics
Vendor comparison

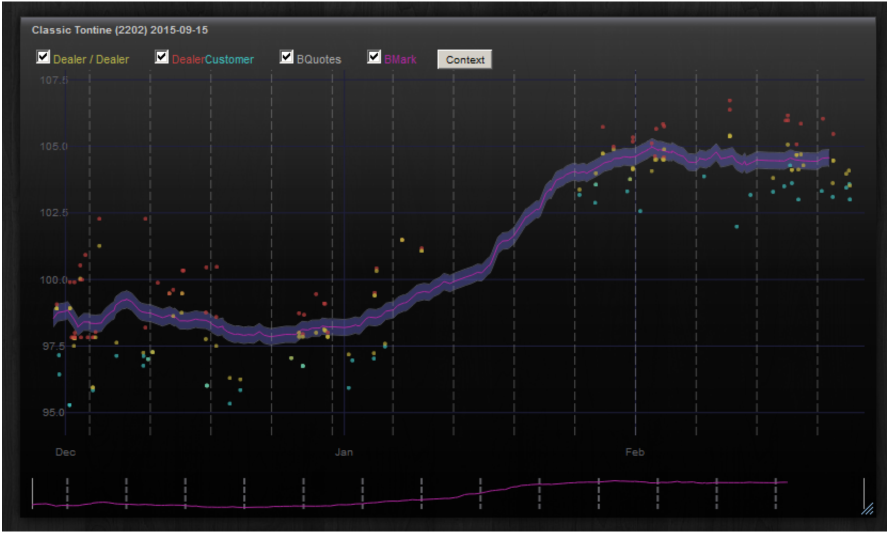
Filtering credit markets

A visual introduction

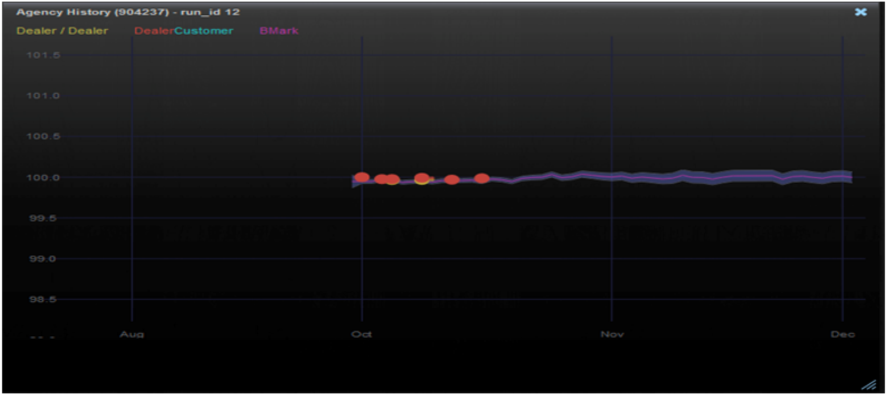

DAIGR
Daimler Finance North America Corporation 6.5 11/15/2013
66.2 2Y +0.4
0.939 109.294
Spread
1D 2D 1W 1M 3M
Dealer / Dealer
Cust. Buy / Cust. Sell
IQ
BMark
Hide off-mkt
260
240
220
200
180
160
140
120
100
80
60
40
20
0
-20
-40
12/01
12/15
01/01
01/15
02/01
02/15
03/01

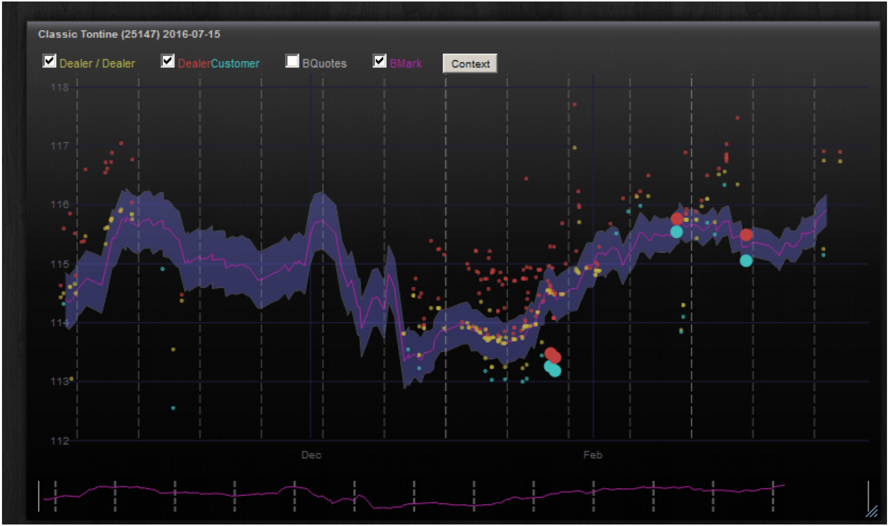

DAIGR
Daimler Finance North America Corporation 3 03/28/2016
113.9 5Y +2.6
1.982 103.950
Spread
1D 2D 1W 1M 3M
Dealer / Dealer
Cust. Buy / Cust. Sell
IQ
BMark
Hide off-mkt
215
210
205
200
195
190
185
180
175
170
165
160
155
150
12/01
12/15
01/01
01/15
02/01
02/15
03/01

Filtering credit markets



Filtering credit markets




Classic Tontine (25147) 2016-07-15
Dealer / Dealer
DealerCustomer
BQuotes
BMark
Context
118
117
116
115
114
113
112
Dec
Feb

Filtering credit markets

A unified state space for bond and CDS markets

Hazard, funding curves can be represented as hidden processes. From this basic state, bond and credit default swaps can be computed using the usual present value computations. Additional state, representing the width of the market at various maturities, can be used to provide a rough estimate of where trading will occur. The actual observed trades y_{t_i} are considered noisy measurements of a non-linear function of state x_{t_i}.

$$y_{t_i} = h\left(x_{t_i}\right) + \epsilon_i \tag{1}$$

where h incorporates credit default swap or bond pricing formulas, depending on the observation in question.

Kalman filtering

Linearize by differentiation, for now. Then consider a sequence of observations times $t_1, ..., t_k$ at which our latent vector process x is observed indirectly, via an observation equation

$$y_{t_i} = H_i x_{t_i} + \epsilon_i \tag{2}$$

We assume ϵ_i is mean zero multivariate gaussian with covariance R_i. For brevity we refer to y_{t_i} as y_i, x_{t_i} as x_i and so forth. We assume the evolution of x in between the times specified can be written

$$x_{i+1} = A_i x_i + u_i \tag{3}$$

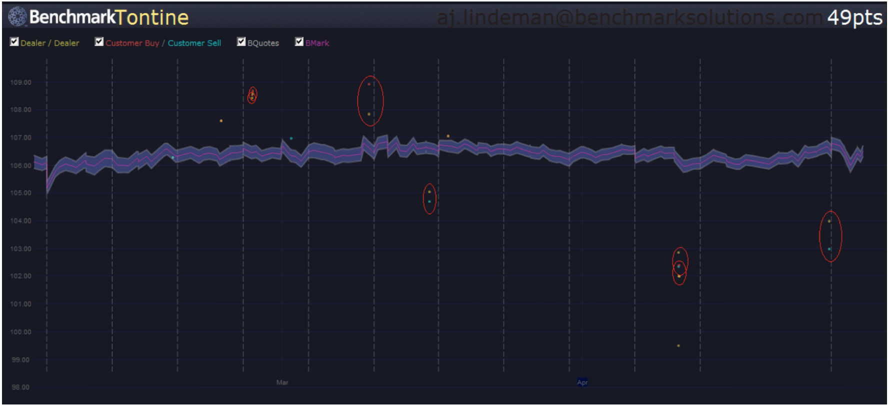
where u_i are also gaussian. In this linear gaussian system the recursive estimation of x_t is achieved by the well known Kalman filter. I think we knew that already!

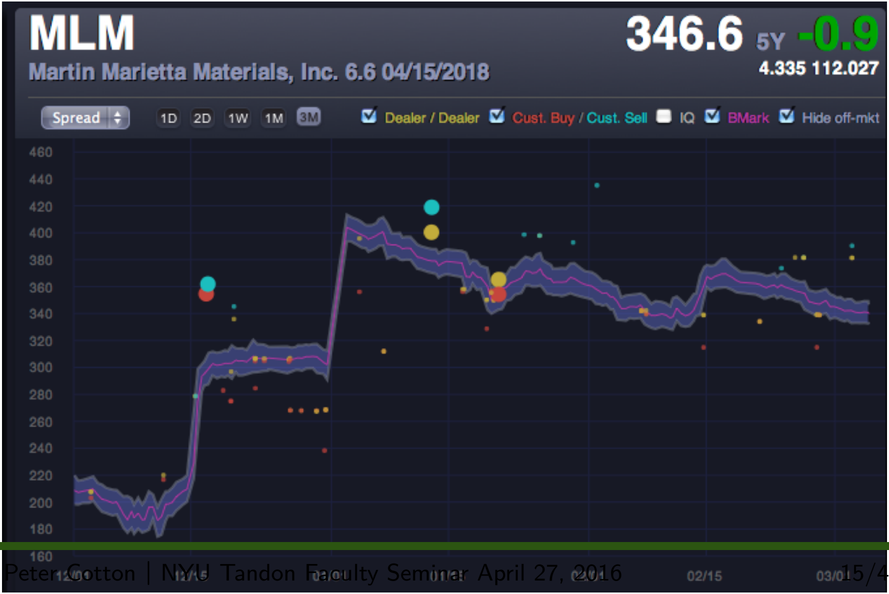
Issues

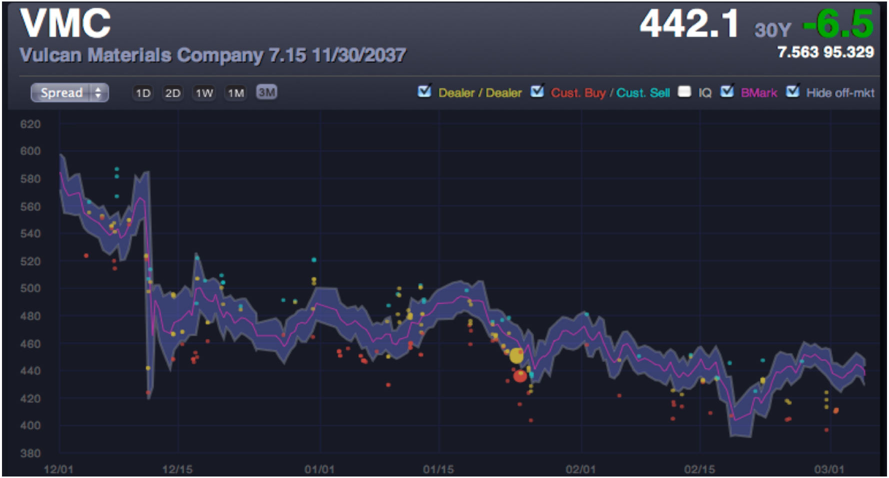
Stare at bond price time series for a while and you'll quickly discover that every assumption is violated

1. Brokered trades
2. Jumps
3. Lags
4. Off-market quotes
5. Funding peculiarities

Filtering credit markets

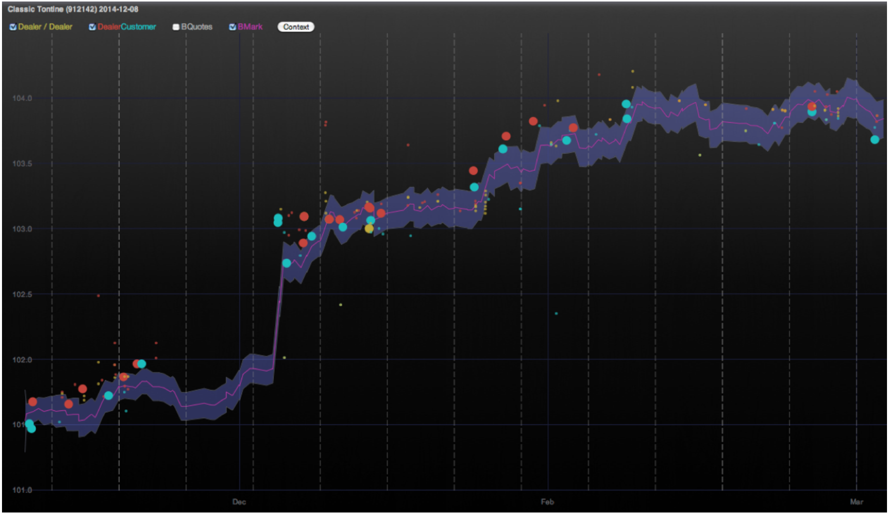



MLM
Martin Marietta Materials, Inc. 6.6 04/15/2018
346.6 5Y -0.9
4.335 112.027
Spread
1D 2D 1W 1M 3M
Dealer / Dealer
Cust. Buy / Cust. Sell
IQ
BMark
Hide off-mkt
460
440
420
400
380
360
340
320
300
280
260
240
220
200
180
160
02/15
03/01

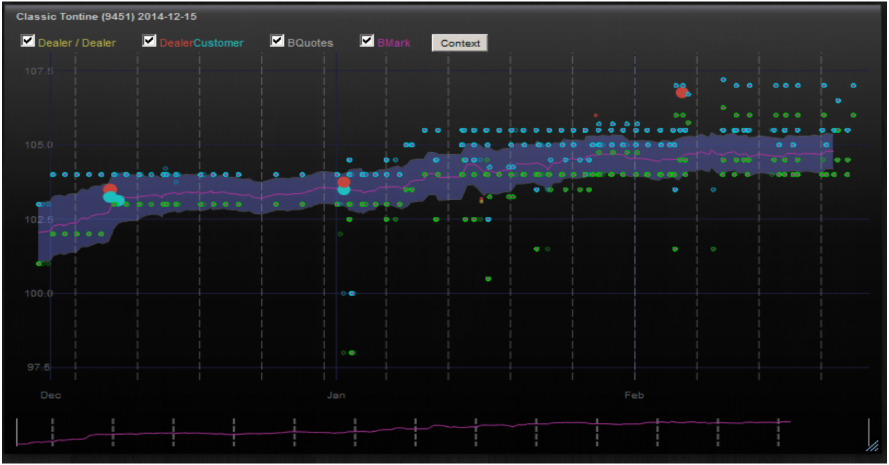


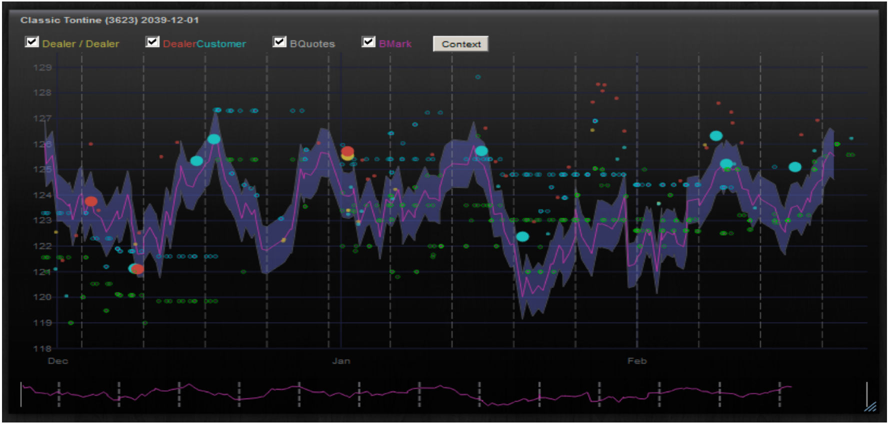
VMC
Vulcan Materials Company 7.15 11/30/2037
442.1 30Y -6.5
7.563 95.329
Spread
1D 2D 1W 1M 3M
Dealer / Dealer
Cust. Buy / Cust. Sell
IQ
BMark
Hide off-mkt
620
600
580
560
540
520
500
480
460
440
420
400
380
12/01
12/15
01/01
01/15
02/01
02/15
03/01

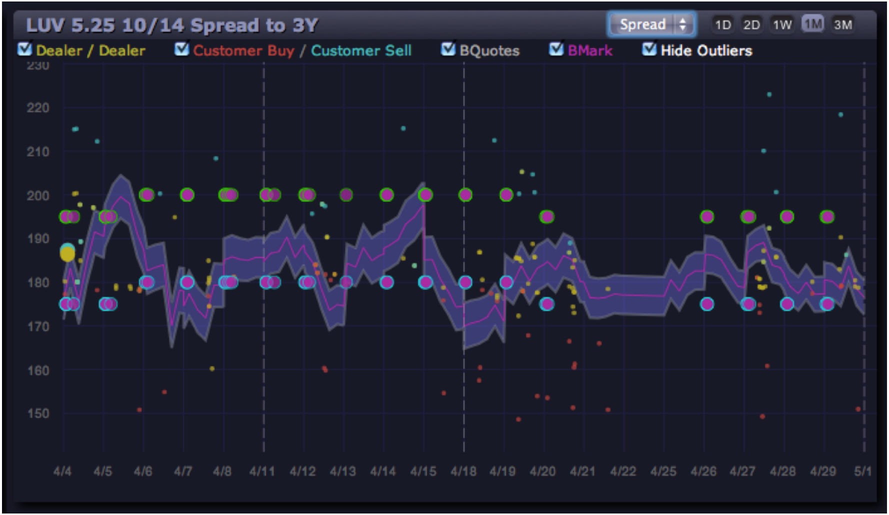
Filtering credit markets



Filtering credit markets




Classic Tontine (3623) 2039-12-01
Dealer / Dealer
DealerCustomer
BQuotes
BMark
Context
129
128
127
126
125
124
123
122
121
120
119
118
Dec
Jan
Feb


LUV 5.25 10/14 Spread to 3Y
Spread
1D 2D 1W 1M 3M
Dealer / Dealer
Customer Buy / Customer Sell
BQuotes
BMark
Hide Outliers
230
220
210
200
190
180
170
160
150
4/4 4/5 4/6 4/7 4/8 4/11 4/12 4/13 4/14 4/15 4/18 4/19 4/20 4/21 4/22 4/25 4/26 4/27 4/28 4/29 5/1

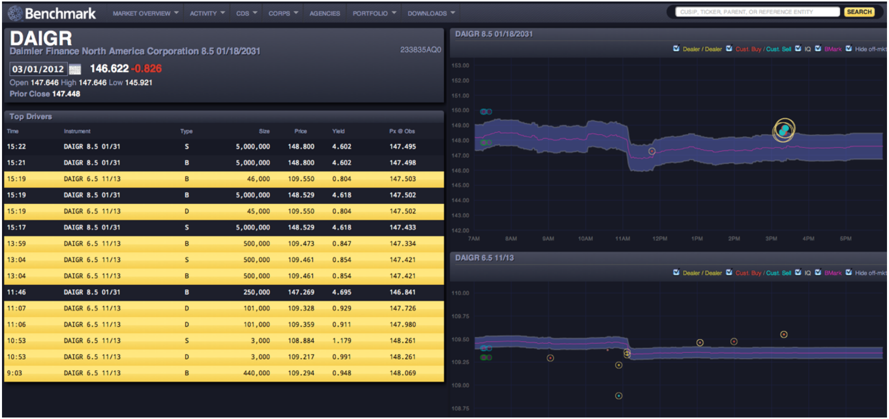
Filtering credit markets

Explaining and justifying pricing

The contemporaneous impact of an observation y_{k+1} is proportional to the Kalman gain. (I think we knew that already too). This creates the pricing narrative.

But most observations are not contemporaneous ...

Historical importance of observations



The derivatives of the Kalman filter estimate with respect to a **past** observation y_i is not something we see too often.

1. Re-represent the Kalman estimate in the form of a weighted least squares problem (c.f. Duncan Horn representation).

2. Compute sensitivities of the solution of the weighted least squares problem. Sometimes a little adjoint trick helps here.

Kalman as least squares on the current state

Sketch: We set up a least squares problem involving the current state x_k only. The solution to this problem is identical to the Kalman filter's current estimate. This establishes that the current estimate $\hat{y}_k$ is a simple linear function of the current state x_k, so we can compute the derivative of the current estimate with respect to all previous observations.

To be tidy, assume a gaussian prior on the initial state x_0. To avoid annoying special cases in what follows, we clean up the notation by indexing back to -1 as follows:

$$\begin{aligned} y_{-1} &= H_{-1}x_{-1} + \epsilon_{-1} \\ x_0 &= A_{-1}x_{-1} + u_{-1} \end{aligned}$$

and here H_{-1} and A_{-1} are identity matrices, $\epsilon{-1}$ is identically zero, y_{-1} is set equal to the mean of our prior and u_0 adopts its covariance.

With the boundary conditions cleaned up in this fashion we can invert the dynamical equations, assuming only that A's have left inverses A^{-1}, as follows:

$$x_j = A_j^{-1}\left(x_{j+1} - u_j\right) \tag{4}$$

and then re-arrange the observation equations so that the only value of x_i that appears is x_k.

The inversion looks like:

$$\begin{aligned}
y_k &= H_k x_k + \epsilon_k \\
y_{k-1} &= H_{k-1} x_{k-1} + \epsilon_{k-1} \\
&= H_{k-1}\left(A_{k-1}^{-1}\left(x_k - u_{k-1}\right)\right) + \epsilon_{k-1} \\
&= H_{k-1}A_{k-1}^{-1}x_k - H_{k-1}A_{k-1}^{-1}u_{k-1} + \epsilon_{k-1} \\
y_{k-2} &= H_{k-2} x_{k-2} + \epsilon_{k-2} \\
&= H_{k-2}\left(A_{k-2}^{-1}\left(x_{k-1} - u_{k-2}\right)\right) + \epsilon_{k-2} \\
&= H_{k-2}A_{k-2}^{-1}x_{k-1} - H_{k-2}A_{k-2}^{-1}u_{k-2} + \epsilon_{k-2} \\
&= H_{k-2}A_{k-2}^{-1}\left(A_{k-1}^{-1}\left(x_k - u_{k-1}\right)\right) - H_{k-2}A_{k-2}^{-1}u_{k-2} + \epsilon_{k-2} \\
&= H_{k-2}A_{k-2}^{-1}A_{k-1}^{-1}x_k - H_{k-2}A_{k-2}^{-1}A_{k-1}^{-1}u_{k-1} - H_{k-2}A_{k-2}^{-1}u_{k-2} + \epsilon_{k-2} \\
&\ldots
\end{aligned}$$

From which it is apparent that if we write $Y = (y_k, y_{k-1}, y_{k-2}, ..., y_{-1})$ then

$$Y = Gx_k + \eta \tag{5}$$

where G is the concatenation of the coefficients of x_k given above and η is the gaussian random variable equal to the sum of u_k's and ϵ_k's. Thus we have a simple least squares problem for the contemporaneous state x_k which is not dissimilar to the Duncan-Horn representation of the Kalman filter.

Sensitivities of the least square problem

Suppose x solves $Qx = b(y)$. We wish to compute the derivative of $g(x)$ w.r.t. y (because, to recap this will tell traders how important every historical observation is to the current price estimate whether or not the observation pertains to a bond in question).

In particular, if y is the observation and x the solution of a generalized least squares problem with error co-variance R we can cast it in this form by writing:

$$\begin{aligned} g(x) &= Hx \\ Q &= H^T R^{-1} H \\ b(y) &= H^T R^{-1} y \end{aligned}$$

Consider now

$$f(x, y) = 0 \tag{6}$$

where

$$f(x, y) = Qx - b(y) \tag{7}$$

We use derivatives of

$$\tilde{g} = g - \lambda^T f(x, y) \tag{8}$$

with respect to y as a means of computing derivatives of g with respect to y.

Note that

$$\frac{\partial \tilde{g}}{\partial y} = \frac{\partial g}{\partial x}\frac{\partial x}{\partial y} - \lambda^T \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial y} + \frac{\partial f}{\partial y} \right) \tag{9}$$

and this will simplify if we choose λ judiciously as a solution of

$$\frac{\partial g}{\partial x} = \lambda^T \frac{\partial f}{\partial x} \tag{10}$$

which is the adjoint equation. For then

$$\begin{aligned} \frac{\partial \tilde{g}}{\partial y} &= \frac{\partial g}{\partial x}\frac{\partial x}{\partial y} - \lambda^T \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial y} + \frac{\partial f}{\partial y} \right) && (11) \\ &= -\lambda^T \frac{\partial f}{\partial y} && (12) \\ &= \lambda^T \frac{\partial b}{\partial y} && (13) \end{aligned}$$

Now specializing to

$$g(x) = Hx \tag{14}$$

and $b(y)$ as above we can solve for this convenient choice of λ by writing

$$\begin{aligned} H &= \frac{\partial g}{\partial x} && (15) \\ &= \lambda^T \frac{\partial f}{\partial x} && (16) \\ &= \lambda^T Q && (17) \\ &= \lambda^T H^T R^{-1} H && (18) \end{aligned}$$

where the second equality is the adjoint equation. Thus we can compute derivatives of $\tilde{g}$ with respect to y, and thereby compute derivatives of g with respect to y which is what we set out to do.

Accuracy

At some point you have to explain the accuracy of your prices to clients. On the vendor side it is sometimes argued that customers are insensitive to pricing quality and the service is therefore sticky. This is party true but that argument presumes there will be no material change in market structure or competitive forces. In fact major buy side firms are gearing up to better quantify their transaction costs relative to peers. Others wish to use their inventory to generate alpha. And sell-side firms look for lower cost means of making markets and even assessing traders.

Unfortunately, accuracy is a subtle beast...

A simple target

The next inter-dealer trade. Issues:

1. Rare for many bonds
2. Noisy
3. Repeated
4. Paired
5. Serially correlated

A time, size and money-under-the-bridge weighted target

Fix some moment t at which a price is supplied by a vendor and consider the J subsequent inter-dealer trades:

$$FVT(t;J) = \frac{\sum_{j=1}^{J} p_j s_j e^{-(t_j - t)} e^{-M_j^-}}{\sum_{j=1}^{J} s_j e^{-(t_j - t)} e^{-M_j^-}} \tag{19}$$

where p_j, s_j and t_j are the price, size and time of subsequent inter-dealer trades with time measured in business days and

$$M_j^- = \frac{1}{c} \sum_{k=1}^{j-1} s_k \tag{20}$$

is the cumulative trading volume up to but not including the trade in question.

Fixing an inconsistency

The use of M_j^- is slightly unnatural because it means the target is not invariant to splitting of future trades. We can easily fix this, however, by integrating in money instead of time.

$$FVT'(t;J) = \frac{\int_{m=0}^{M_J^+} p(m) e^{-m} e^{-(t(m)-t)} dm}{\int_{m=0}^{M_J^+} e^{-m} e^{-(t(m)-t)} dm}$$

where $M_J^+ := M_{J+1}^-$ is the total amount of money under the bridge up to and including the J'th trade, $p(m)$ is the price when m dollars of trading has occurred, and $t(m) - t$ is the time we have progressed when m dollars of trading has occurred.

Multivariate filters seemingly perform worse then univariate

A more serious issue with accuracy measures for bond pricing relates to the inefficiency in the market. One needs to be careful not to reward univariate models (treating only the bond in question) for over-fitting. A stylized simulation to prove the point following West and Papanicoloau can run as follows. Disregard funding rates and use piece-wise hazards:

$$\lambda(t) = \lambda_i, t_{i-1} \leq t_i \tag{21}$$

so that yields

$$y_i = \int_0^{T_i} \lambda(t) dt \tag{22}$$

are clearly linear in the hazard rates.

Assume now that hazards are independent random walks

$$\lambda_k(t_d t) = \lambda_k(t) + N(0, dt\sigma_k^2) \tag{23}$$

and for simplicity set $\sigma_1 = 0.1$, $\sigma_2 = 0.05$ and $\sigma_3 = 0.01$. Assume bond prices are observed at times generated by a self-exciting Hawkes process:

$$dc_t = (c_t - c)e^{-\kappa t} + f dN_t \tag{24}$$

West and Papanicolaou studied various types of bonds based on these parameters. They then used both univariate and multivariate Kalman filters, and also a simple model where the estimate was simply the last trade observation.

	Last Trade	Univariate K.F.	Multivariate K.F.
Apparent error	0.020	0.0377	0.0383
Actual error	0.056	0.0512	0.0286

Table 1: Apparent error when measured against the next trade, and actual error against ground truth. In a simulation the model that simply uses the last trade of the bond in question appears to perform better than the multivariate Kalman filter applied to the entire curve - but that is an illusion. Numbers courtesy of Nick West and George Papanicolaou.

Serial correlation distorts accuracy metrics

A related problem occurs if we assume that trade observations are serially correlated errors with respect to some ground truth. Simple models can over-fit but appear to perform well in simple accuracy statistics.

With all that said ... here is a comparison between the current version of Benchmark, built by A.J. Linderman, Antoine Toussaint et al, and a prominent vendor offering real-time pricing. (We can't compare the original Benchmark pricing service as there were no other real-time services at the time).

Time since last trade	Market leader	BMRK	Count
5 - 20 mins	0.238	0.140	7,348
20 mins -2 hr	0.215	0.156	9,326
2hr - 4hr	0.250	0.174	726
4hr - 10 hr	0.217	0.159	5,931
10hr - 2 days	0.217	0.171	2,613
2 days - 1 week	0.234	0.189	4,738
			30,672

Vendor comparison

Questions?