1 Hazard model estimation for trade arrival process

We apply the kernel hazard model to assess the trade arrival process in a futures markets for the case of silver. This is conjectured to depend on the depth of silver as well as on the depth of gold (that is highly correlated with silver). We select intraday data of silver (SI_H_2013 contract) and gold (GC_G_2013 contract) futures for 2012-12-11. The contracts selected are liquid at the chosen day. The data is taken from RDTH database.

1.1 Data cleaning

To prepare the data for analysis, we apply a number of filters on it that are discussed as follows.

1. We restrict our attention to liquid hours of silver futures contract that typically correspond to open outcry trading hours. These are characterized by high levels of trading frequency as shown in figure 1. Thus, we select data in 07:25-12:25 interval.

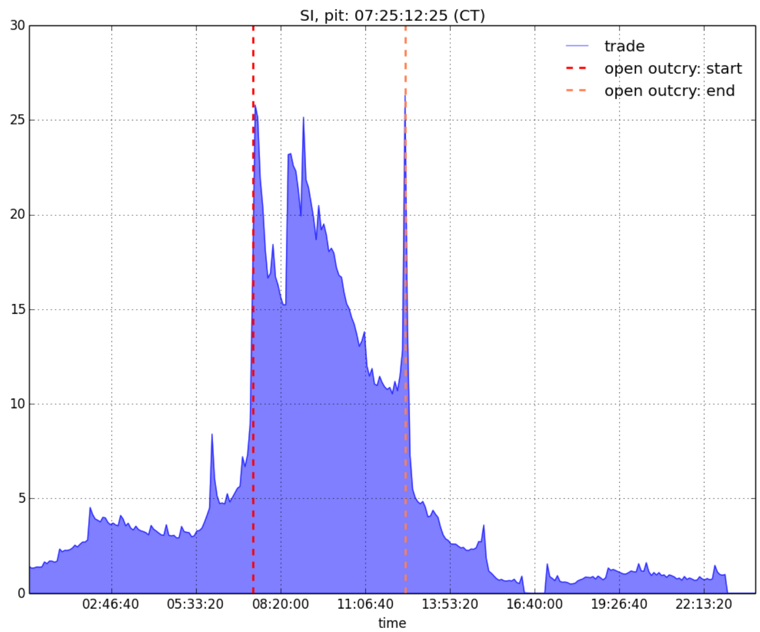

Figure 1: Trade frequency for 5-minute buckets for silver

2. Second, focusing on the trade data, we remove trades with non-positive prices.
3. We aggregate both trades and quotes occurring at the same time stamp. The median of the prices corresponding to these transactions is assigned as the current price for that aggregated trade. In terms of quotes, the sum of bid and offer sizes are selected.

1.2 Descriptive plots

We select depth of silver and gold as our individual and common covariates. Each depth series is calculated as

$$DPTH_t = \log(ASKSIZE_t) - \log(BIDSIZE_t)$$

We then normalize each of them by their standard deviation to make them on the same scale for the multivariate kernel computation.

1.2.1 Trades

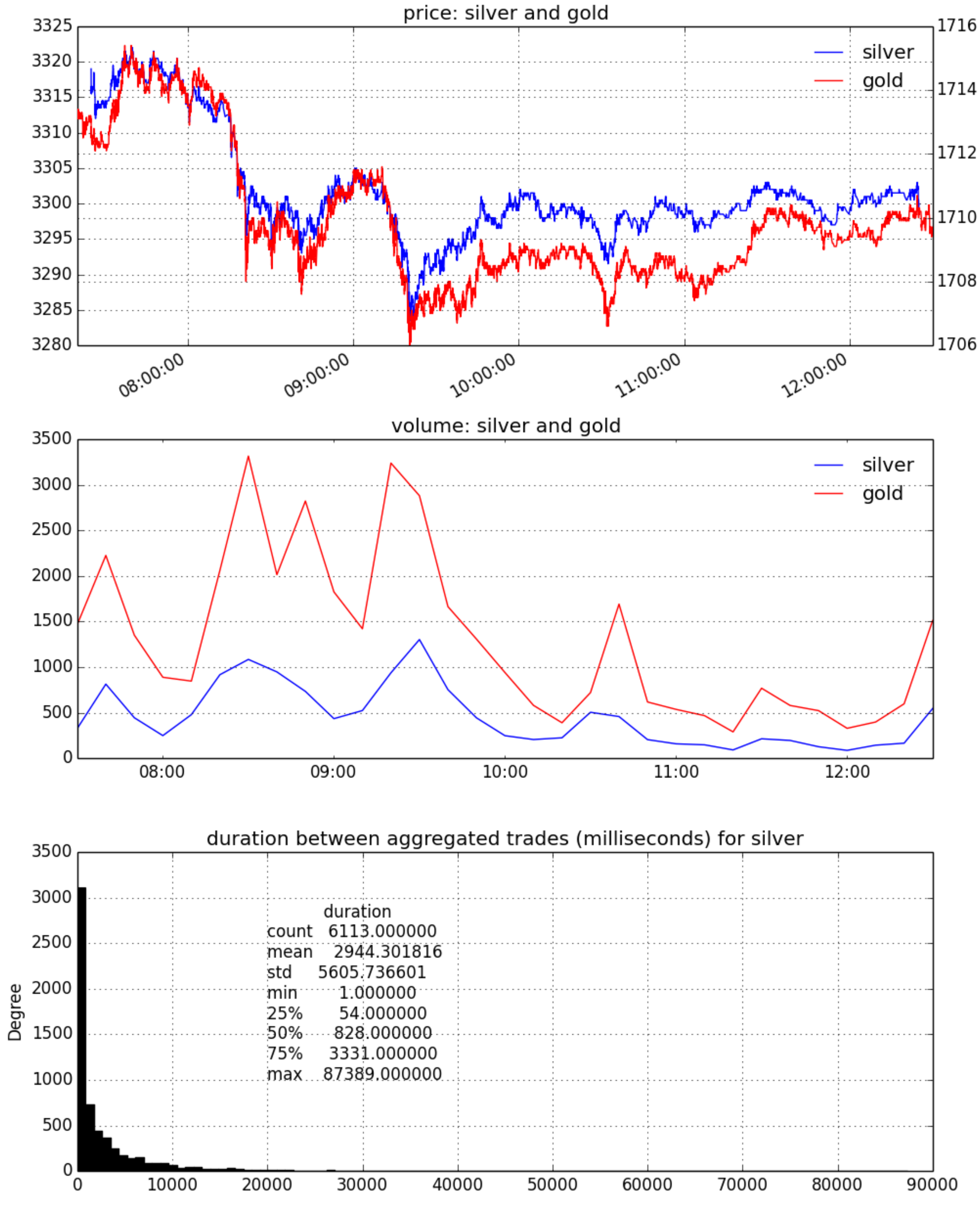

Figure 2: Silver (blue) and gold (red) futures contracts price, volume and duration (for silver only)

1.2.2 Quotes

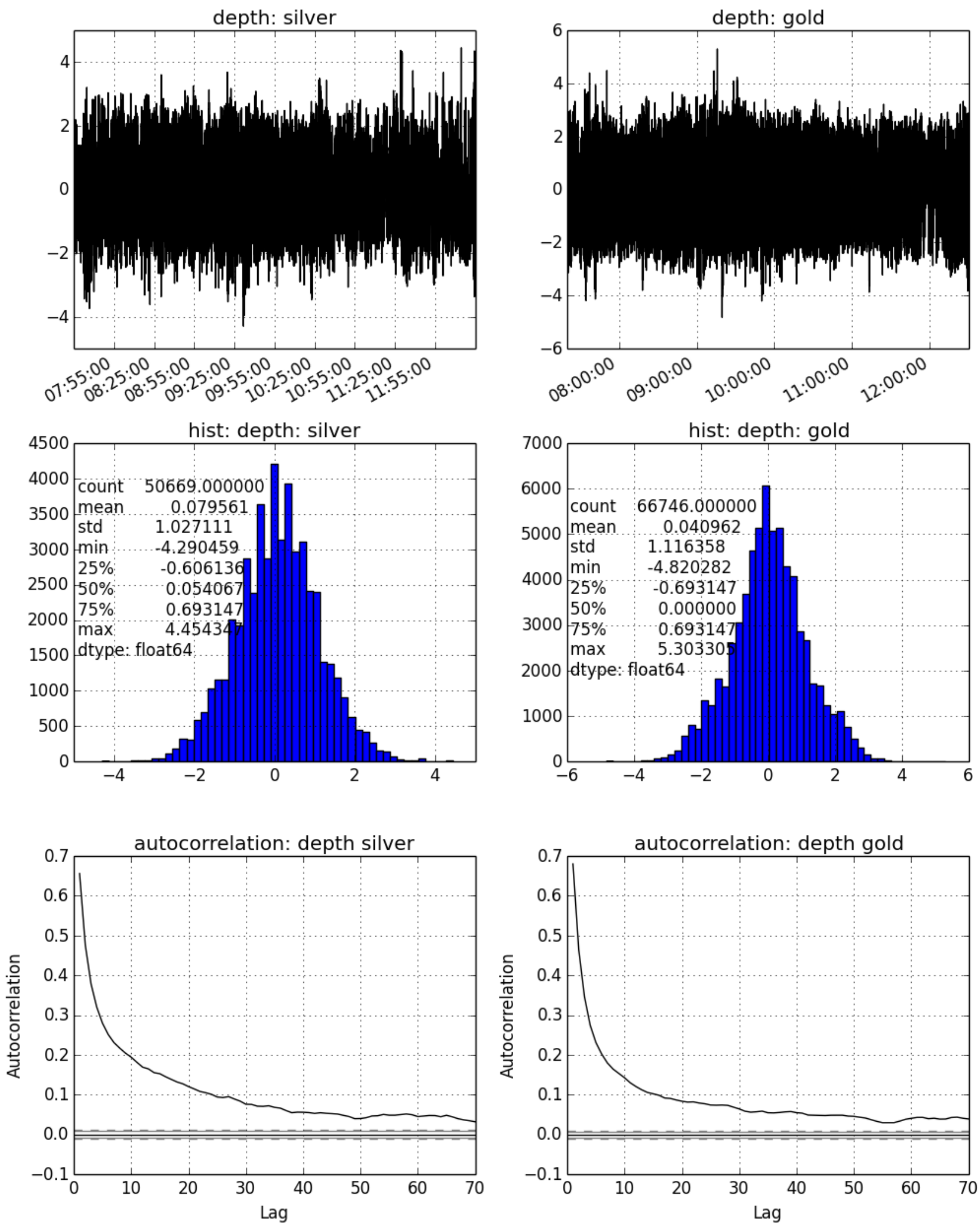

Figure 3: Silver and gold futures contracts depth plots, histograms and autocorrelations

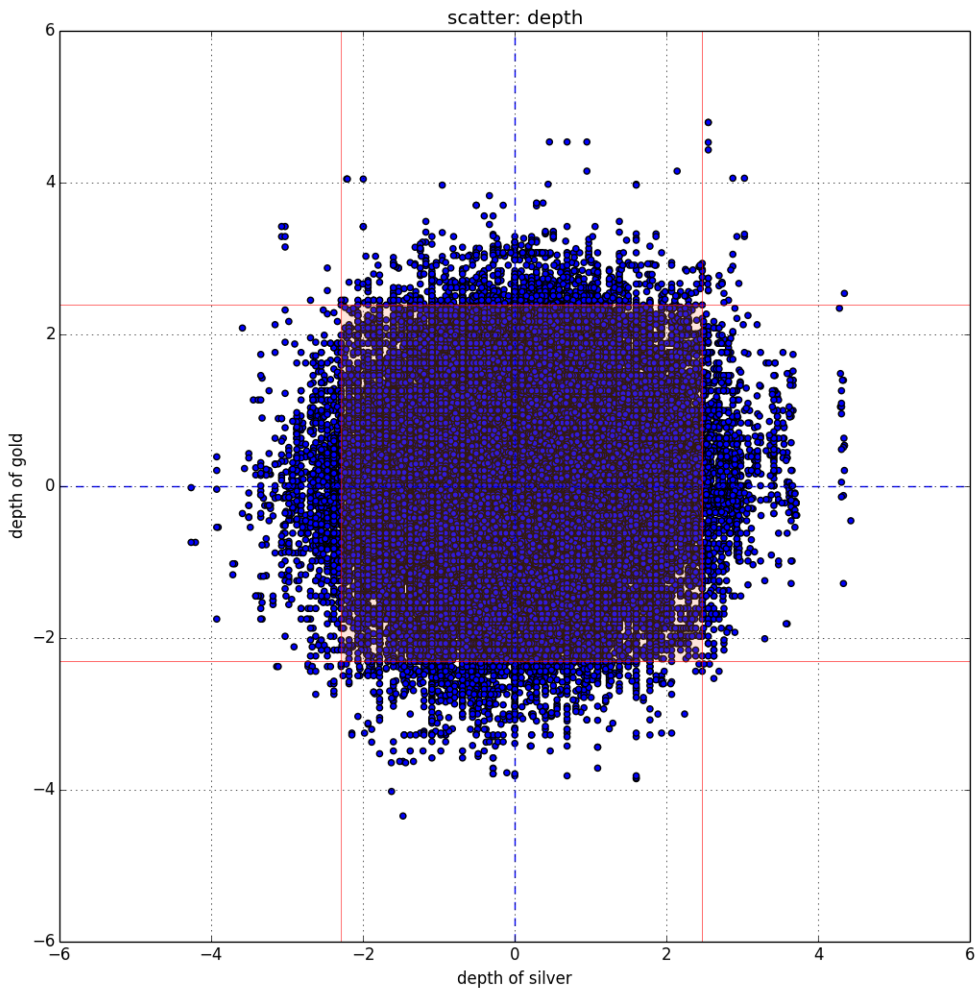

Figure 4: Scatter plot of depth of silver and gold (red lines are 0.01 and 0.99 quantile for each depth series)

1.3 Bandwidth selection: cross-validation

Using Gamiz et al. (2013) cross validation method we estimate the following cross-validation score for each bandwidth $\underline{b}$:

$$\hat{Q}_0\left(\underline{b}\right) = n^{-1}\left\{\sum_{i=1}^{n}\int\left[\hat{\alpha}_{K,\underline{b}}\left\{s, Z_i\left(s\right)\right\}\right]^2 Y_i\left(s\right)ds - 2\sum_{i=1}^{n}\int\hat{\alpha}_{K,\underline{b}}^{[i]}\left\{s, Z_i\left(s\right)\right\}dN_i\left(s\right)\right\}$$

In particular for each of the covariates (i.e. depth of silver and gold) we select a grid of 10 equally-sparsed points located between 0.01 and 0.99 quantiles.

1.3.1 Optimal bandwidth results

b=1.1

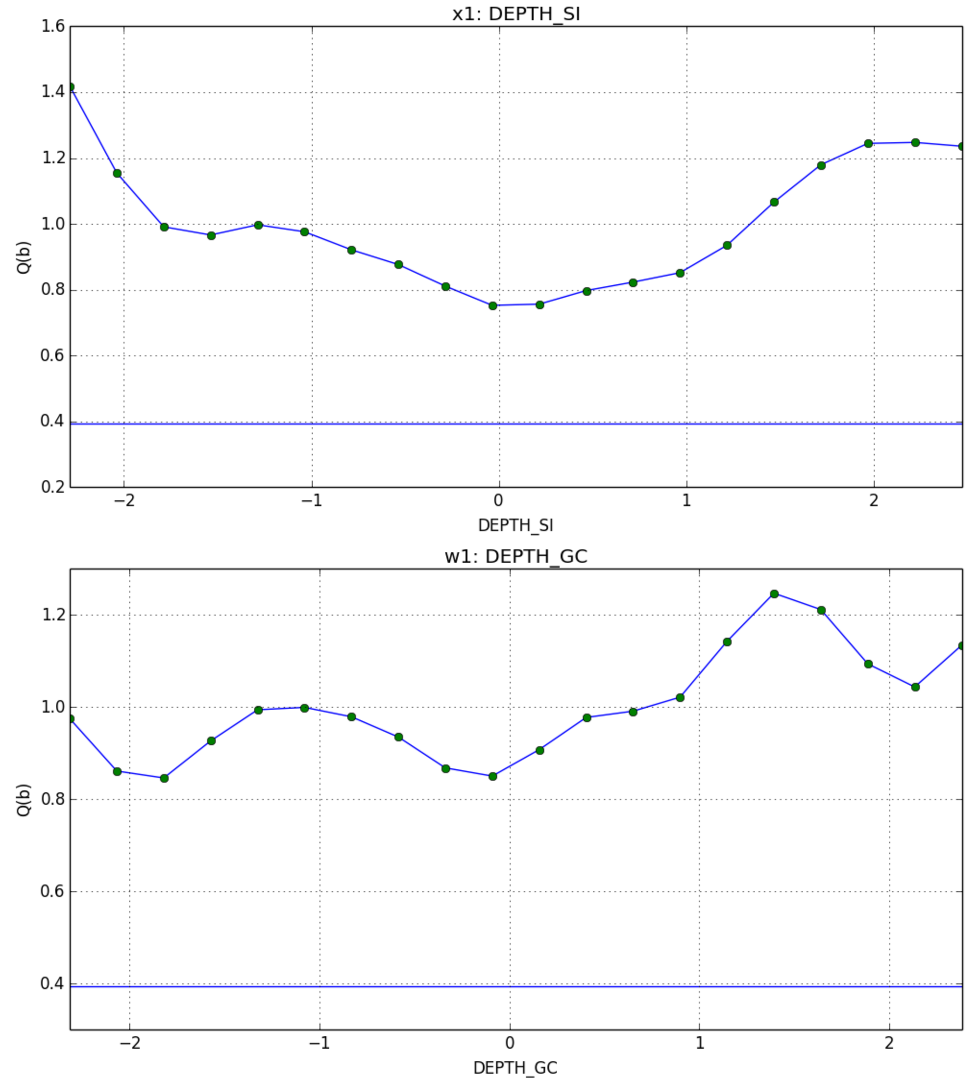

Figure 5: Hazard rate with optimal bandwidth

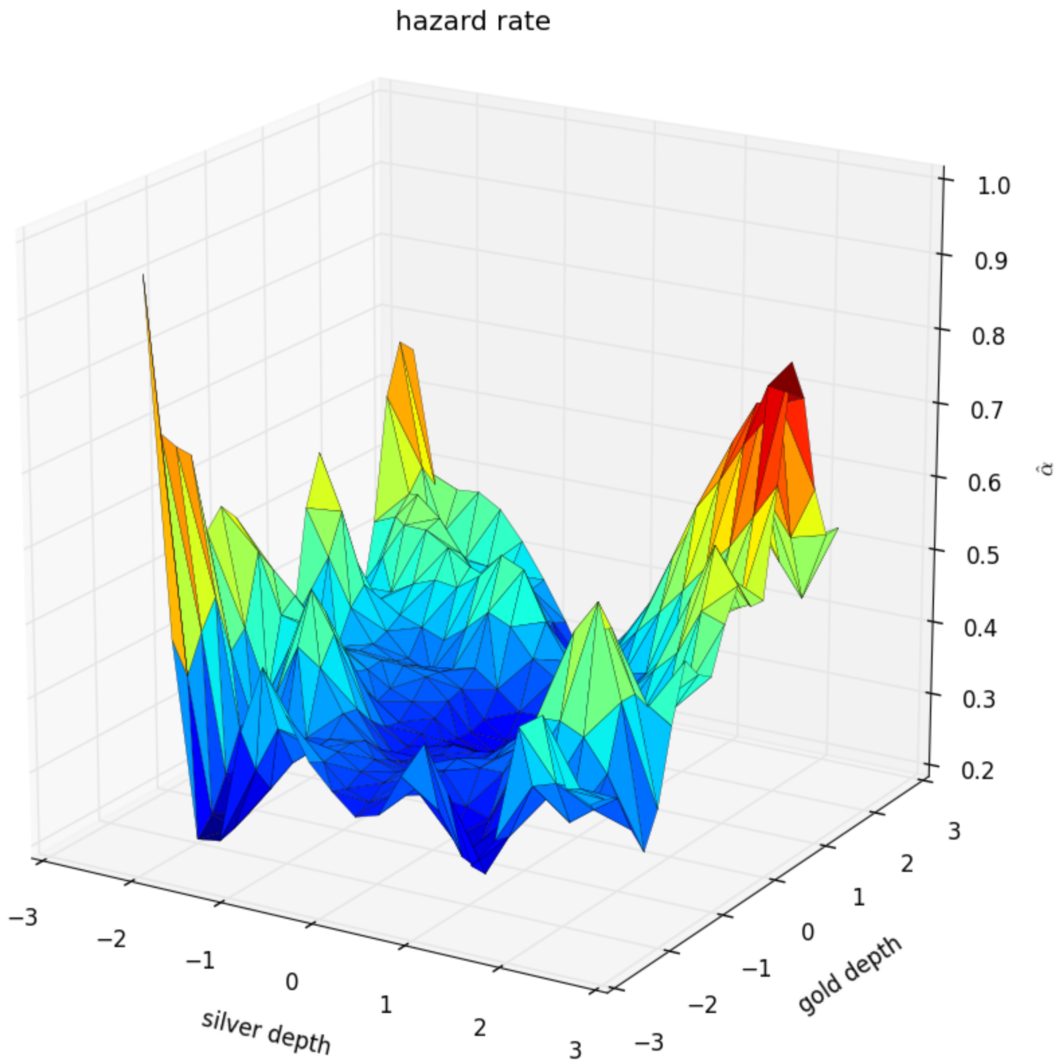

Figure 6: Hazard rate in 3 dimensions

1.3.2 Optimal bandwidth+30% results

b=1.1*1.3

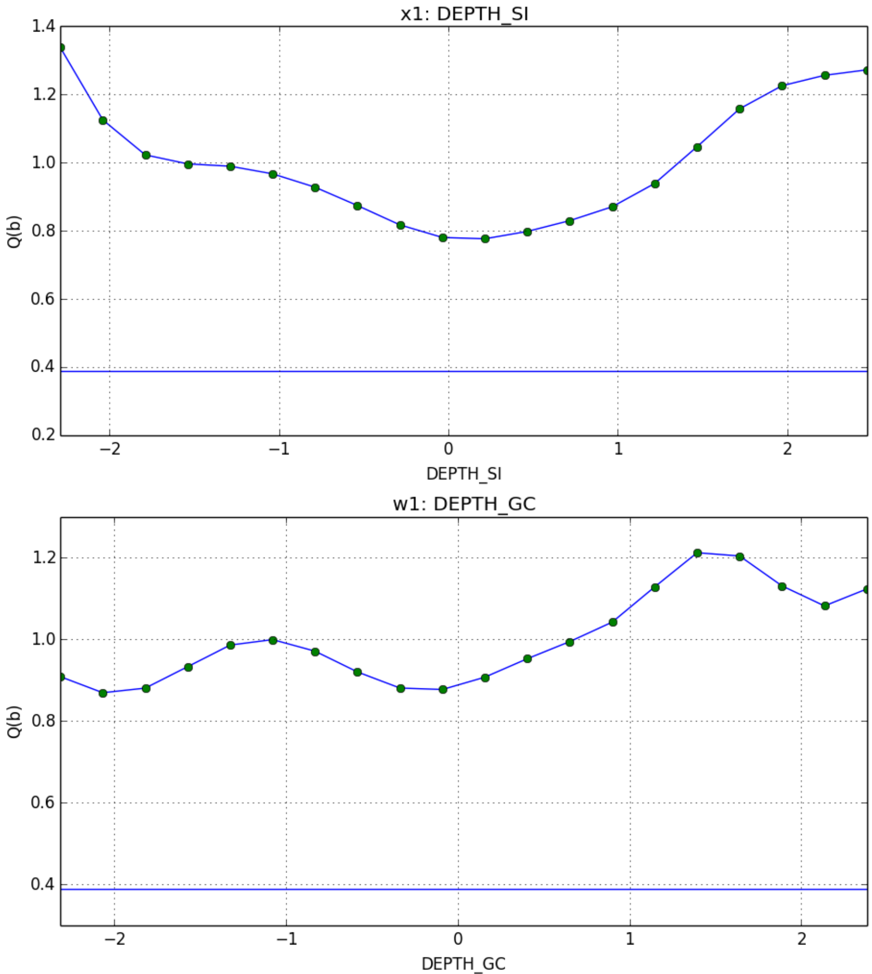

Figure 7: Hazard rate with optimal bandwidth + 30%

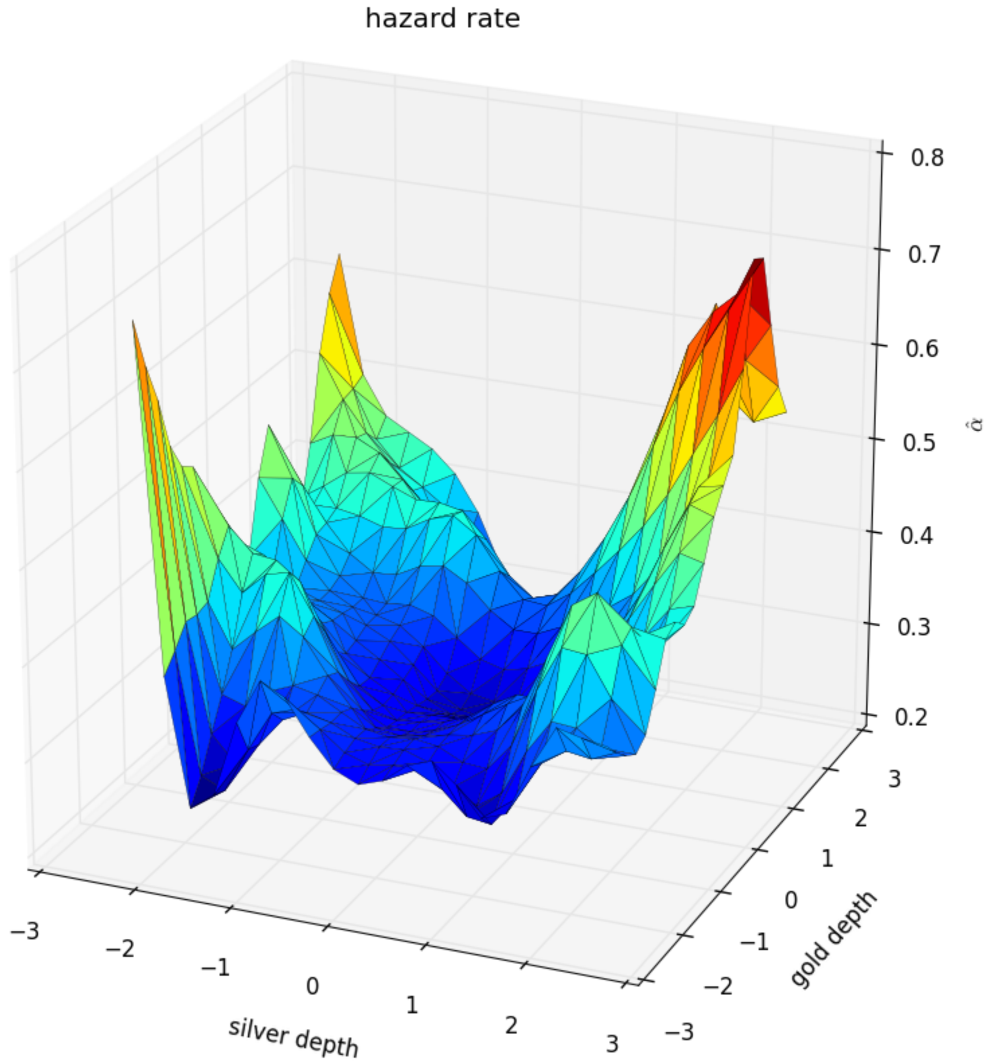

Figure 8: Hazard rate in 3 dimensions with optimal bandwidth +30%