

Similarity-Based Quantile Assessment

Gilbert W. Bassett Jr.*

August 2011

October 2011 Revised

Abstract: Gilboa, Schmeidler, and coauthors pose the problem of assessing an unknown real-valued outcome y_0 based on its associated known x_0 and a database of observations $\{(y_i, x_i), i = 1, \dots, n\}$. The assessment is a weighted combination of the y_i 's with weights depending on a similarity function $s(x_i, x_0) > 0$ that measure the assessor's notion of the similarity between x_0 and the x_i in the database. In this paper the approach is used to assess the quantiles of y_0 . Rather than being assessed directly the quantiles are derived from the similarity-based empirical distribution. Features of the resulting assessed quantiles are illustrated. Assessment of the value-at-risk for IBM conditional on the market return provides an empirical application of the approach.

*University of Illinois at Chicago
gib@uic.edu

Similarity-Based Quantile Assessment

1. Introduction

The similarity-based approach for assessing an outcome has been considered in papers by Gilboa, Schmeidler, and coauthors; see references. As an example, a realtor wants to assess the expected price of a house with characteristics x_0 ; see, for example, Gayer, Gilboa, and Lieberman (2007). The realtor can consult a database of recent sales prices and house characteristics. A similarity assessment might be used because no houses in the database match the characteristics x_0 . Even if there are such houses the realtor might want to utilize sales information for houses with characteristics similar to, but not identical with, x_0 . The similarity assessment for the expected sales price is a weighted average of the sales data in

the database-- $\sum_{i=1}^n \lambda(x_i, x_0) y_i$ --where the (convex) weights are given by

$$\lambda(x_i, x_0) = \frac{s(x_i, x_0)}{\sum_{i=1}^n s(x_i, x_0)}, \quad s(x, x_0) > 0.$$

The similarity approach was proposed as a representation of how a person forms an opinion about y_0 . The intent was to describe how assessments are made and how they are understood by oneself and explained to others. (“I think the house will sell for \$200K because similar houses have sold for that amount”)¹. Rather than an estimator in a formal statistical model the assessment derives from intuition that y_0 with its x_0 will be similar to the y_i 's that have already been observed with a similarity depending on the x_i 's similarity to x_0 . The approach subsumes questions about how the components of x_0 relevant for an assessment of y_0 are determined. If there is an underlying structural model for $y_0(x_0)$, the assessment might be viewed as the output of the associated reduced form.

¹ Good explanations/arguments have been recently proposed as the motivation for human reasoning; see Mercier and Sperber (2011). Reasoning evolved not so much as a way to get the correct answer, as to convince and enlist cooperation among other (social) humans. Axiomatic formulations are compelling because they make convincing arguments.

Gilboa and Schmeidler (2003) provide an axiomatic rationale for the similarity approach. The axioms establish a framework for the analogical reasoning underlying the similarity intuition. The main axiom is a compelling consistency requirement about how assessments should work when databases are combined or concatenated. The axiom says that an assessment derived from the combination of databases should be a convex combination of the assessments for the databases considered separately. You cannot combine databases and get an assessment outside the range of its constituents. If the assessment was \$200K with database A and \$180K for database B, you cannot get to \$210K by combining A and B. It is clear that the similarity formula satisfies this convexity requirement. Gilboa and Schmeidler showed that (together with other conditions²) the converse is also true; convex assessments imply the existence of a similarity function such that assessments agree with the similarity formula.

Following the initial proposal was the recognition that similarity assessment was formally the same as nonparametric regression smoothing. In the regression situation y_0 is the noisy outcome of an unknown continuous function $m(x_0)$. An estimate for $m(x_0)$ is a weighted average of y_i 's where weights depend on the proximity of x_i to x_0 . The bias from using observations $x_i \neq x_0$ can be offset by increasing the effective number of observations used for the estimate. The resulting estimate as a function of x_0 produces a graph, $\{(y_0, \hat{m}(x_0)) \mid x_0 \in R^d\}$, that is a smoothed fit to the (y_i, x_i) in the database. While the motivation is different, the underlying technology for similarity assessment and regression smoothing is essentially the same; similarity assessment is regression smoothing with a kernel given by the similarity function.³

² The other main requirement, which is assumed below, is that the similarity function does not depend on the ordering or the number of observations in the database. Situations in which order does matter (such as when more recent observations carry greater similarity information) are proposed to be handled by including time as an additional component of x_0 . For discussion of the pros and cons of the similarity approach see section 4 of Gilboa and Schmeidler (2010). Also, see the approach in Gayer(2010) where the similarity function depends on the number of observations.

³ See Gilboa and Schmeidler (2003); Gilboa, Lieberman, and Schmeidler (2006), (2008), (2010), (2011).

Another variation on the original similarity proposal was its extension to other features of y_0 . The initial proposal called for an assessment of the expected value of y_0 conditional on its x_0 thus resulting in the similarity-based weighted average of the y_i 's. This was followed by similarity-based proposals for assessing probabilities. The assessment formula was the same except that the y_i 's are replaced with the (one-zero) indicator function-- $I[y_i = z]$ --so that the assessments become similarity-based weighted frequencies.⁴

In this paper the probability assessment is modified to assess the cumulative distribution function (cdf) of y_0 . The similarity cdf is defined by replacing the $I[y_i = z]$ used for probabilities with $I[y_i < z]$. This yields the similarity based assessment for the cdf of y_0 conditional on x_0 , given the database A, which will be denoted, $F_A(z | x_0)$. This cdf encompasses the other assessment features in that $E(y_0)$ (the expected value of the random variable with cdf $F_A(z | x_0)$) is the similarity assessment for the average, and the probabilities/frequencies implicit in $F_A(z | x_0)$ are the same as the similarity-based probabilities.⁵

The similarity based θ^{th} conditional (on x_0) quantile given the database A is then given by the inverse of the cdf:

$$Q_A(\theta | x_0) = \min\{z : \theta \leq F_A(z | x_0)\}$$

Among the noteworthy features of this similarity based θ^{th} quantile is that it is not a similarity-based weighted function (average or otherwise) of θ^{th} quantiles⁶. Also since it is derived from a well-defined cdf, the resulting quantile function remains well defined throughout its domain. As such it does have the “crossing” issues that are typically associated with estimates based directly on

⁴ See Billot et al. (2005). Eichberger and Guerdjikova (2009), Gayer (2010); Gilboa, Lieberman, and Schmeidler (2010)

⁵ It will be assumed that the same similarity function is used for the different features y_0 .

⁶ For procedures based on spline fitting of regression quantiles see, Koenker, Ng and Portnoy(1994).

quantiles⁷. This and other properties are discussed and illustrated in the examples below.

The distinction between conditional expectations and conditional quantiles is emphasized in the quantile regression literature⁸. Relationships between explanatory variables x and the conditional expectation $E(y/x)$ can be very different than between x and $Q(\theta | x)$. For a linear model specification, the behavior of $E(y | x\beta)$ and its least squares estimate $E(y | x\hat{\beta})$ can be different than $Q(\theta | x\beta(\theta))$ and its regression quantile estimate $Q(\theta | x\hat{\beta}(\theta))$. These distinctions apply also to the smoothed versions of conditional expected values and quantiles.

Comparison of *linear* OLS and regression quantiles with the nonparametric similarity assessments are included in the examples and illustrations in the next sections.

The final section of the paper presents a simple finance application of quantile assessment. The value-at-risk, VaR_p , is a measure of portfolio or security risk; it is the p^{th} quantile of the asset's return distribution. For a given probability p it specifies how much (at least) will be lost. A hedge fund risk manager for example wants to assess value-at-risk given the characteristics x_0 of a portfolio or asset, analogous to the house characteristics in the realtor problem. With the simple market model for the return of an individual stock, such as IBM, conditioning is on the market return, and $VaR_{.1}(x_0)$ is the .1 quantile of today's IBM return distribution given a market return of x_0 . The last section illustrates quantile assessment for IBM's VaR based on a database of daily return data.

2. Preliminaries

Databases are represented by capitals, A or B: $A = \{(y_i^A, x_i^A), i = 1, \dots, n_A\}$. The assessment for the cumulative distribution function (cdf), of y_0 ("at" x_0 -- $y_0(x_0)$) based on the database A is given by:

⁷ See, for example, He(1997).

⁸ See Koenker (2005), Koenker and Hallock (2001), Koenker and Bassett (1978).

$$F_A(z | x_0) = S_A^{-1}(x_0) \sum_{i=1}^{n_A} s(x_i^A, x_0) I[y_i^A < z]$$

where $I[y_i^A < z]$ is the 1 or 0 indicator function, and $S_A(x_0) = \sum_{i=1}^{n_A} s(x_i, x_0)$. The resulting cdf has jumps at each of the y_i where the jump heights depend on the similarity function.

The assessment for the associated quantile function is the inverse of the cdf,

$$Q_A(\theta | x_0) = \min\{z : \theta \leq F_A(z | x_0)\}.$$

Being the inverse, the quantile function has jumps at values corresponding to the jump heights of the cdf, and the values taken by $Q_A(\theta | x_0)$, $\theta \in (0, 1)$ are the y_i in the database. This makes the computation of the quantile function straightforward; see Appendix.

Sometimes it is useful to consider the data generating process that produces a $y_0(x_0)$ that has the similarity-based cdf. This can be represented by, $y_0 = Q_A(U | x_0)$ where U is a uniform random variable on $[0, 1]$

The similarity assessment will sometimes be compared to the usual empirical cdf of all the $y_i \in A$ (which does not depend on x_0 or x_i). This is denoted by $F_A^e(z)$ where, $F_A^e(z) = n_A^{-1} \sum_{i=1}^{n_A} I[y_i^A < z]$.

In this paper attention is focused on the one-dimensional case where $x_0 \in R$ and $x_i \in R$. This allows visualization of properties and consideration of how quantile assessment works in simple cases.

The similarity function is assumed to not depend on the ordering or number of observations in the database. Being a similarity measure it is typically assumed positive throughout its range, a maximum at $x_i = x_0$, and a nonincreasing function of the distance between x_i and x_0 . Positive throughout its range means never dividing by zero in the definition of the cdf, and that dissimilarity cannot be

infinitely large⁹. Permutation invariance means that for designs in which there are J subsets with each x_i in a subset equal to a common value, all the observations in the subset will have the same similarity and the similarity-based cdf for the subset will be the same as the usual empirical cdf. These simple J -subset designs are used in the illustrations to motivate features of quantile assessment.

The similarity function used for the illustrations is, $s(x, x_0) = e^{-\alpha|x-x_0|}$. The $\alpha \geq 0$ parameter controls the degree of similarity between x and x_0 ¹⁰. When α is small there is little variation in $s(x, x_0)$ as a function of x and all the data has about the same similarity. Similarity begins to matter more and more as α increases.

The similarity cdf $F_A(z | x_0)$ can be usefully interpreted as interpolating between two extreme cases. In the first, similarity is constant, $s(x, x_0) = c$ for all x . This corresponds to $\alpha = 0$ for the above s -function. In this case every observation is equisimilar to x_0 , and $F_A(z | x_0) = F_A^e(z)$, the empirical cdf of all the $y_i \in A$, for any x_0 .

The other extreme is the limiting case in which all the x_i are infinitely dissimilar to x_0 except for the observation(s) closest to x_0 . For the above s -function this corresponds to $\alpha \rightarrow \infty$. Let $A^*(x_0)$ denote the subset of the database whose x_i are closest to x_0 , $A^*(x_0) = \{(x_i, y_i) | \min_{i=1, \dots, n} (|x_i - x_0|) = |x_i - x_0|\}$. Note that (i) $A^*(x_0)$ might have only a single observations closest to x_0 , or (ii) multiple observations as when there are multiple x_i s with the same value, or (iii) multiple observations as when x_0 is the midpoint of two adjacent x_i s. In these cases the similarity based cdf is the empirical cdf of the $y_i \in A^*(x_0)$, $F_A(z | x_0) = F_{A^*}^e(z)$. If

⁹ The similarity considered here depends on the absolute distance between x_i and x_0 . A relative measure (for example based on nearest neighbors) is not considered.

¹⁰ Axioms for similarity functions $s(x, x_0) = e^{-d(x, x_0)}$ where $d(x, x_0)$ is a distance function are presented in Billot et al. (2008).

there is only one closest y_{i^*} observation, the cdf collapses to the degenerate cdf,

$$F_A(z | x_0) = F_{A^*}^e(z) = I(y_{i^*} < z).$$

The combination of databases A and B produces a similarity cdf that is the convex combination of A and B:

$$F_{A\&B}(z | x_0) = \lambda F_A(z | x_0) + (1 - \lambda) F_B(z | x_0)$$

where $\lambda = \lambda(x_0, A, B) = \frac{S_A(x_0)}{S_A(x_0) + S_B(x_0)}$. The dgp that produces the

y_0 for the combination is, $y_0 = Q_{DB}(U | x_0)$, where DB is a random database, equal to A with probability λ and B with probability $1 - \lambda$.

Combining databases is the reverse of partitioning a given database into subsets. The cdf for A is therefore a convex combination of a subset partition, where the convex weights are determined by the similarity function.

The definition of the similarity quantile function in terms of the similarity cdf means that the simple convex formulas for combining cdfs do not work for quantiles. The similarity based θ^{th} quantile of A&B is not a similarity-based weighted function (average or otherwise) of the θ^{th} quantiles of A and B. This well recognized feature of the usual U quantiles means, for example, that the median (or average) of subsets of medians is not the median. The subset medians and averages define different estimates¹¹. This and other features of quantile assessment are discussed below.

The similarity assessment viewed as a function of x_0 generates a nonparametric, nonlinear, quantile fit to the x/y data. These can be compared to regression quantile and ordinary least squares estimates in which the conditional model is linear in x . Given independent variables, $x_{ik}, k = 1, \dots, K$, the (linear) regression quantiles are defined by

¹¹ For the median of subset medians (the remedian) see Rousseeuw and Bassett(1990). The average of subset medians (and other quantiles) is not the same as the median (or quantiles), but with iid sampling the first order asymptotic distribution is the same as the median (and other quantiles); see Bassett, Tam and Knight (2002). See Knight and Bassett (2002) where the second order and other properties of the average of subset quantiles are considered.

$$\min_{b_0, b_1, \dots, b_k} \left\{ \sum_{i=1}^n \rho(y_i - v_i : \theta) \mid v_i = b_0 + \sum_{k=1}^K x_{ik} b_k, i = 1, \dots, n \right\}$$

where $\rho(y_i - v_i : \theta) = \theta |y_i - v_i|^+ + (1 - \theta) |y_i - v_i|^-$ and $|v|^+ = |v|$ if $v > 0$ and 0 otherwise, and $|v|^- = -|v|$. The linear model least squares estimate is defined by $\rho(y_i - v_i) = (y_i - v_i)^2$.

While not immediately apparent, the relationship between the “ ρ ” in the quantile regression problem and the indicator function in the empirical cdf (that implicitly defines quantiles), is the connection between integrals and their

derivatives. A cdf is the derivative of its integral, $G(v) = \int_{-\infty}^v F(t) dt$, where G is

necessarily convex (because F is nondecreasing). The minimum of

$H(v : \theta) = G(v) - v\theta$ therefore is a v^* such that $0 \in \partial H(v^* : \theta)$ or

$0 \in \partial G(v^*) - \theta = F(v^*) - \theta$, which says that v^* is a θ^{th} quantile.

The relevance for of this for ρ is that its derivative, viz., subdifferential $\partial \rho(y_i - v_i : \theta)$ is: (i) $\theta - I(y_i < v)$ when $y_i \neq v_i$ and (ii) the set, $\{[-(1 - \theta), \theta]\}$ if $y_i = v_i$. The condition for a minimum when there is only b_0 is

$0 \in \sum_{i=1}^n \partial \rho(y_i - b : \theta)$ which says b is a θ^{th} quantile of the $y_i \in A$.

3. Examples and Illustrations

Example 0.

Suppose all the x_i in the database are equal to k^A , $A = \{(y_i^A, x_i^A) \mid x_i^A = k^A\}$.

In this case $S_A(x_0) = s(k^A, x_0)n_A$, so that the weights are all the same,

$\frac{s(k^A, x_0)}{S_A(x_0)} = \frac{1}{n}$ and the assessed distribution is just the usual empirical cdf

$F_A(z \mid x_0) = F_A^e(z)$ --for any x_0 , and does not depend on s.¹²

¹² The average similarity however decreases as the distance between the $x_i = k^A$ in the database increases from the target x_0 .

Next let the database B be like A except that all its x_i are at $k^B = 1$ where $k^A = 0$. For $x_0 = (1 - \lambda)$, $0 < \lambda < 1$, the assessed cdf is seen to be the convex combination of the usual empirical cdfs at A and B:

$$F_{A\&B}(z | x_0) = \delta F_A^e(z) + (1 - \delta) F_B^e(z) \text{ where,}$$

$$\delta(\lambda : \alpha) = \frac{n^A e^{-\alpha(1-\lambda)}}{n^A e^{-\alpha(1-\lambda)} + n^B e^{-\alpha\lambda}} = \frac{1}{1 + (n^B / n^A) e^{-\alpha(2\lambda-1)}}. \text{ (Conversely, the convex}$$

weights $\delta(\lambda)$, $0 < \lambda < 1$, allows the α -parameter to be identified for an assessor whose similarity function is $s(x, x_0) = e^{-\alpha|x-x_0|}$).

As noted above, when $\alpha = 0$ all the observations are equally similar to x_0 and the similarity assessment is the empirical distribution of all the y_i in the database for all x_0 . When $\lambda = 1/2$ the target x_0 is equidistant from the two distinct x_i in the database and the similarity assessment is again the empirical cdf of all the y_i in the database. Alternatively, as $\alpha \rightarrow \infty$ and $\lambda < 1/2$, the similarity cdf is, $F_{A\&B}(z | x_0) = F_B^e(z)$ whereas, when $\lambda > 1/2$, $F_{A\&B}(z | x_0) = F_A^e(z)$. That is, in the limiting cases, the similarity-based cdf is the cdf of the database whose x_i is closest to x_0 .

Example 1. Let A be the database of the previous example with $k^A = 1$, and let B and C be similar except that all their observations are at $k^B = 2$ and $k^C = 6$. We want to assess the cdf at $x_0 = 3$, given the similarity function, $s(x, x_0) = e^{-|x-x_0|}$. The data for the example is shown in Table 1a, and the plot of the data is shown in Figure 1a.

In this case the assessment is a weighted average of the usual empirical cdfs where the weights depend on the similarity function and the number of observations in each of the databases:

$$F_{A\&B\&C}(z | x_0 = 3) = \lambda_1 F_A^e(z) + \lambda_2 F_B^e(z) + \lambda_3 F_C^e(z).$$

$$(\lambda_1 = n^A s(1, 3) / S(x_0 = 3) = .245, \lambda_2 = n^B s(2, 3) / S(x_0 = 3) = .665,$$

$$\lambda_3 = n^C s(6, 3) / S(x_0 = 3) = .090).$$

Figure 1b shows the empirical cdfs at $x=1$, $x=2$ and $x=6$, as well as their convex combination that yields the assessment for $F_{A\&B\&C}(z|x_0=3)$.

Example 2. This is similar to the previous example except that there are additional y_i observations in each of A, B and C databases. We want to see how the simple cdfs at k^A, k^B, k^C , combine for (i) different α values and (ii) as a function of x_0 .

Figures 2a, 2b, and 2c show assessments for three different α values, (.1,1,10). For each α , the figures depict selected quantiles as a function of x_0 .

When α is small the s-function evaluates all of observations as having about the same similarity and hence there is little variation in $F(z|x_0)$ as a function of x_0 . This is evidenced in Figure 2a where the quantile lines are nearly horizontal.

As α increases the s-function puts increasing weight on x_i near x_0 . As a result, the quantile lines begin to more closely interpolate the observations in each of A, B, and C; see Figures 2b, 2c, 2d.

When α goes to infinity, see Figure 2d, $F(z|x_0)$ converges to the ordinary empirical cdf for the database closest to x_0 (if x_0 is the midpoint of two nearest x_i then $F(z|x_0)$ is the average of the adjacent simple cdfs--as in the above 2-sample case).

These nonparametric quantile “fits” are contrasted with linear OLS and regression quantiles in Figure 2e.

A well known property of least squares for the k-sample design is that the estimate based on all the data is the same as regression through the subsample means (weighted by the number of observations in each subsample). The subsample means are sufficient statistics for all the data when estimated by least squares. In the ultra simple case where the database only includes y_i this corresponds to the average of subset averages being the same as the average. In Bassett, Tam, and Knight (2002) the regression quantile and least squares estimates were compared for a k-sample design and it was noted that the

subsample quantiles are not sufficient for regression quantiles. For example, in Figure 2e, the .5 regression quantile is not the same as the L1 (or L2) estimate through the subsample medians. It was further noted that, analogous to the similarity based quantile, the first order condition for regression quantiles involves the empirical cdf at each of the subsamples.

Example 4. The previous illustrations show assessments when there is a non trivial empirical cdf reference at each of the subsamples. Example 4 shows how the assessments work for a more general design. Figure 3 shows a more general configuration of data. Parts a, b and c show how the quantile fits vary with x_0 and α . When α is small, the observations are all similar and the cdf and quantiles vary only slightly with respect to x_0 . As α increases the data near x_0 receive more weight and the fits begin to more closely interpolate the observations, as in Figures 3b and 3c.¹³

4. IBM Value-at-Risk conditional on the Market return.

This section illustrates the similarity-based assessments in the context of a simple market model for the return of IBM. Let y_0 denote today's return of IBM. We want to assess the quantile function of y_0 given today's market return x_0 , as measured by the New Stock Exchange Composite (NYX) index. The assessment is based on three years of daily data for IBM and the NYX index, May 27, 2008 to June 10, 2011.

Figures 4a, b, c and d show the similarity quantile fits as a function of the NYX for various α values. The fits are overlaid on the daily returns for the 762 observations in the database.

As observed previously, when α is small the similarity function makes little distinction between x_i and x_0 , and the fits correspond to the empirical cdf and quantile function using all of IBM's daily returns. As α increases the quantile fits

¹³ The appendix shows how the fits for a general design are, approximately, a convex combination of simple empirical cdfs for a partition of the database into subsets with similar x data.

begin to more closely interpolate the data and reveal the dependence of the quantiles on the market return. For example, Figure 4b shows how the quantiles depend on the NYX. The increasing quantile spread as NYX moves away from the origin means increasing IBM return variability as a function of the magnitude of the market return.

The quantiles can be used to measure value-at-risk conditional on the market return. Consider, for example, the .1 quantile assessment. This is IBM's value-at-risk at probability=.1 as a function of the NYX. Reading from Figure 4b shows $Q(.1|x=.01)=-0.85\%$, which means that if the market goes up by .01=1% there is a .1 probability that IBM will fall by 0.85% or more. Alternatively, $Q(.1|x=-.01)=-1.75\%$ means that if the market goes down by 1% there is .1 chance that IBM will fall by 1.75% or more. Notice that the difference between the market return and the .1 quantile in the first case is, $[.01-(-.0085)]=.0185$ whereas in the later case the difference is, $[-.01-(-.0175)]=.075$. The value at risk for IBM relative to the change in the market is greater when the market goes up 1% than when the market goes down by 1%.

These nonparametric similarity quantiles are compared to linear regression quantile and least squares estimates in Figure 4e. The model specifies IBM's return according to whether the market is up ($x_i > 0$) or down ($x_i < 0$).¹⁴ These linear fits like the similarity fits in Figures 5b and 5c show the tendency for the spread of the return distribution to increase as the magnitude of the market return increases.

[Additional results (not shown) were assessed for the similarity function $s(m_t, m_0) = \exp(-\alpha_1 |m_t - m_0| - \alpha_2 t)$, which allows for distinctions between market returns that are more or less recent. Figure 4f shows (for the specified α values, the similarity between m_t and $m_0=0$ as a function the time between m_t and the

¹⁴ This up-down market model was used in Bassett, Gerber, and Rocco (2004) for quantile portfolio attribution. Portfolio attribution using quantile regression was considered in Bassett and Chen (2001).

current time. The assessments are available on request and are similar to those reported below.]

Value-at- Risk over time

The value-at-risk assessment above is for a single day based on the previous three years of daily returns. This is now extended to historical value-at-risk assessments for the past 20 years. The data is IBM and NYX daily returns from January 1, 1990 to June 10, 2011. Assessments for value-at-risk (.1) are produced for each day based on a rolling window of the previous three years of daily data. The quantile assessments are shown for the similarity function with $\alpha=100$.

Figure 5a shows the .1 quantile fits for each day given three alternative market returns: NYX=0, NYX=+2% and NYX=-2%. The figure shows that VaR(x) has not been constant over the 20 years. For each of the NYX returns, the figure shows VaR declining to a minimum during 2002-2004, then increasing to the beginning of 2008 when it began to decline again.

The figure also shows that $VaR_p(x_0)$ trends have varied as a function of NYX. For example, the increasing risk until 2002 to 2004 was most pronounced for NYX=-2%.

Figure 5b compares the similarity quantile assessment with a linear model/least squares benchmark. The market model for IBM daily returns is, $IBM_t = a_t + b_t M_t + \varepsilon_t$. OLS estimates for a_t, b_t and the standard deviation of the error, σ_t were produced for the same rolling windows as used for the above quantile estimates.

The figure shows the OLS vs. similarity assessment for NYX=-1%. The line $a_t - b_t 1\%$ in the figure shows the time varying estimates, $a_t - b_t 1\%$. The horizontal line at -0.01 provides a benchmark showing the expected return when $x=-.01=1\%$ and $a_t = 0, b_t = 1$. Comparison of this benchmark against $a_t - b_t 1\%$ shows how *expected* returns have changed over time. Most of the variation around the benchmark is due to beta, so that estimates below the line indicate a

beta greater than 1 (IBM decreases more than 1% when the market decreases by 1%), and conversely.

The “ $a - b1\% - 1.29\sigma$ ” line adjusts the expected return for the .1 quantile. The standard deviation was estimated for the rolling three year window and then multiplied by -1.29--the .1 quantile of the standard normal distribution. This gives the least squares estimate for the .1 value-at-risk conditional on a market return of -1%.

The figure shows that the ups and downs in the least squares based VaR mimic the quantile assessment. The least squares values however are seen to be consistently lower than the similarity based assessment. This means that-- conditional on $NYX = -.01$ --the risk of IBM as measured by the least squares VaR is greater than when based on quantile assessment.

The same comparison is shown in Figure 5c except that conditioning is on $NYX = +.01$ —the market up +1%. The benchmark line for expected returns is now indicated at .01. The main difference from the previous figure is that the similarity based risk is now greater than the risk based on least squares.

The fact that the value-at-risk comparisons can be sometimes greater and sometimes less when comparing least squares and quantile assessment is due differences in the way the two methods condition on x_0 . With least squares the same estimated variability (standard deviation) is used for every x_0 . As a result, the difference in the least squares value at risk when $NYX = +1\%$ and when $NYX = -1\%$ is, $b_t(1\% - (-1\%)) = b_t 2\%$ (which is the difference in the expected returns). Like its quantile regression counterpart, the quantile-based assessment in contrast, depends on x_0 . The assessment for $x_0 = -1\%$ comes mainly from the variability in IBM on days when the NYX return is around -1%. Likewise, the risk assessment conditional on $x_0 = +1\%$ is due to the behavior of IBM on days when NYX is around +1%. With its greater flexibility, it may be expected that the similarity-based risk measure will be better calibrated to subsequent returns, a topic awaiting future research.

Appendix Computing Similarity Based Quantiles

Let the data base be, $\{(y_i, x_i), i = 1, \dots, n\}$, $y_i \in R$ and $x_i \in R^d$. Let the y_i be ordered (allowing ties) from smallest to largest, $y_1 \leq y_2, \dots, \leq y_n$. For a given similarity function $s : R^d, R^d \rightarrow R_{++} = (0, \infty)$ consider the normalized partial sums,

$$S_i(x_0) = \frac{\sum_{j=1}^i s(x_j, x_0)}{\sum_{i=1}^n s(x_i, x_0)}, i = 1, \dots, n$$

and set $S_0(x_0) = 0$. (Note that since $s(x_i, x_0) > 0$, the $S_i(x_0), i = 1, \dots, n$, are increasing, $S_i(x_0) < S_j(x_0), i < j$).

The points, $\{(y_i, S_i(x_0)) | i = 1, \dots, n\}$ essentially determine the graphs of the jump functions F and Q . To allow for the possibility that the y_i are not distinct consider the modified set, $\{(y_i, S_i^*(x_0)) | S_i^*(x_0) = \max\{S_j(x_0) : y_j = y_i\}, i = 1, \dots, n\}$. It follows directly from the definition that the cdf, F , has jumps at each y_i where,

$$F(y_i | x_0) = S_{i+1}^*(x_0), i = 1, \dots, n.$$

Similarly, the quantile function, Q , has jumps at each $S_i^*(x_0)$ where it takes the value y_i :

$$Q(S_i^*(x_0)) = y_i, i = 1, \dots, n.$$

As a function of x_0 , $F(z | x_0)$ and $Q(\theta | x_0)$ can be evaluated at $x_0 = x_i, i = 1, \dots, n$. Additionally, the cdf and quantile function can be evaluated at the midpoint of adjacent x_i values.

The R-code for computing similarity based quantiles was prepared by Yu Chen and is available at: <http://tigger.uic.edu/~gib/>.

References

- Bassett, Gilbert W., Mo-Yin S. Tam and Keith Knight (2002), “Quantile Models and Estimators for Data Analysis”, *Metrika*, v.55, 17-26
- Bassett, Gilbert W. and Hsiu-lang Chen (2001). “Portfolio Style: Return-based Attribution Using Quantile Regression”, *Empirical Economics*, 26, 293-305.
- Bassett, Gilbert W., Geoffrey Gerber and Pasquale Rocco (2004). “Robust Strategies for Quantitative Investment Management”, *Statistics for Industry and Technology*, Birkhauser Verlag, Basel/Switzerland, p.27-37.
- Billot, Antoine, Itzhak Gilboa, Dov Samet, and David Schmeidler (2008). “Axiomatization of an Exponential Similarity Function”, *Mathematical Social Sciences*, 55, 107-115.
- Billot, Antoine, Itzhak Gilboa, Dov Samet, and David Schmeidler (2005). “Probabilities as Similarity-Weighted Frequencies”. *Econometrica*, 73, No. 4 p. 1125–1136.
- Eichberger, Jürgen and Ani Guerdjikova, Case-Based Belief Formation Under Ambiguity, Paper 2009.
- Gayer, Gabrielle (2010). Perception of Probabilities in situations of risk: A case Based Approach”, *Games and Economic Behavior*, 68, 130-143.
- Gayer, Gabrielle , Itzhak Gilboa, and Offer Lieberman (2007). “Rule-Based and Case-Based Reasoning in Housing Prices”, *BE Journal of Theoretical Economics*. 7, Article 10
- Gilboa, Itzhak, Offer Lieberman, and David Schmeidler (2006). “Empirical Similarity”, *Review of Economics and Statistics*, 88, 433-4444.
- Gilboa, Itzhak, Offer Lieberman, and David Schmeidler (2010). On the Definition of Objective Probabilities by Empirical Similarity, *Synthese*, 172: 79-95.
- Gilboa, Itzhak, Offer Lieberman, and David Schmeidler (2011). “A Similarity-Based Approach to Prediction”. *Journal of Econometrics* 162 (2011) 124–131.
- Gilboa Itzhak, Offer Lieberman, and David Schmeidler (2008). “Axiomatization of an Exponential Similarity Function”, *Mathematical Social Sciences*, 55, 107–115.
- Gilboa, Itzhak and David Schmeidler (2003). “Inductive Inference: An Axiomatic Approach”, *Econometrica* 71, 1–26.

- Gilboa, Itzhak and David Schmeidler (2010). "Case Based Predictions: Introduction". Preprint for the volume, *Case Based Prediction*, World Scientific Publishers.
- He, X. (1997). "Quantile Curves Without Crossing", *Am. Statistician* 51, 186–92.
- Knight, Keith and Gilbert W. Bassett Jr. (2002). "Second Order Improvements of Sample Quantiles using Subsamples",
<http://www.utstat.utoronto.ca/keith/papers/subsample.pdf>
- Koenker, Roger and Kevin F. Hallock, 2001. "Quantile Regression" *Journal of Economic Perspectives*, American Economic Association, vol. 15(4), pages 143-156, Fall.
- Koenker, Roger and Gilbert W. Bassett Jr. (1978). "Regression Quantiles", *Econometrica*, 46, 1, 33-50.
- Koenker, R. (2005). *Quantile Regression*. Cambridge: Cambridge University Press.
- Koenker, R., P. Ng, and S. Portnoy (1994). "Quantile Smoothing Splines", *Biometrika* 81, 673–80.
- Knight, Keith and Gilbert W. Bassett Jr. (2002). "Second Order Improvements of Sample Quantiles using Subsamples",
<http://www.utstat.utoronto.ca/keith/papers/subsample.pdf>
- Mercier, Hugo and Dan Sperber (2011). "Why do Humans Reason? Arguments for and Argumentative Theory", *Behavioral and Brain Science*, 34, 57-111.
- Rousseeuw, Peter J. and Gilbert W. Bassett Jr. (1990). "The Remedial: A Robust Averaging Method for Large Data Sets", *Journal of the American Statistical Association*, 85, 4009, 97-104.

Table 1a
 Assessment data for $\hat{F}(z : x_0 = 3)$

Data Base		
A: x=1	B: x=2	C: x=6
1	5	4
2	6	5
3	8	7
6	9	8
8	10	9
9	12	
	15	
	17	

Similarity function: $s(x_i, x_0) = e^{-|x_i - x_0|}$

Figure 1a
 Assessment data for $\hat{F}(z : x_0 = 3)$

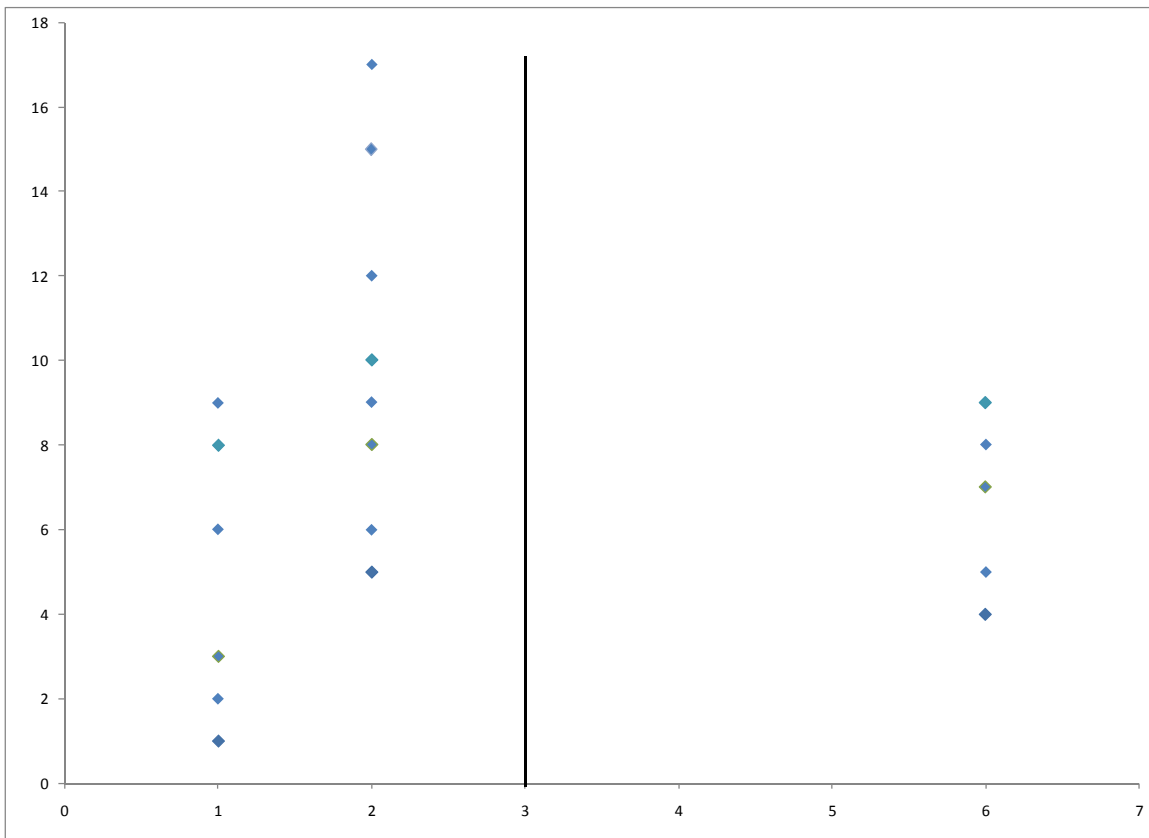


Figure 1b
 Combining $\hat{F}(z: x_i)$ into the assessment for $\hat{F}(z: x_0 = 3)$
 Similarity: $s(x_i, x_0) = e^{-|x_i - x_0|}$

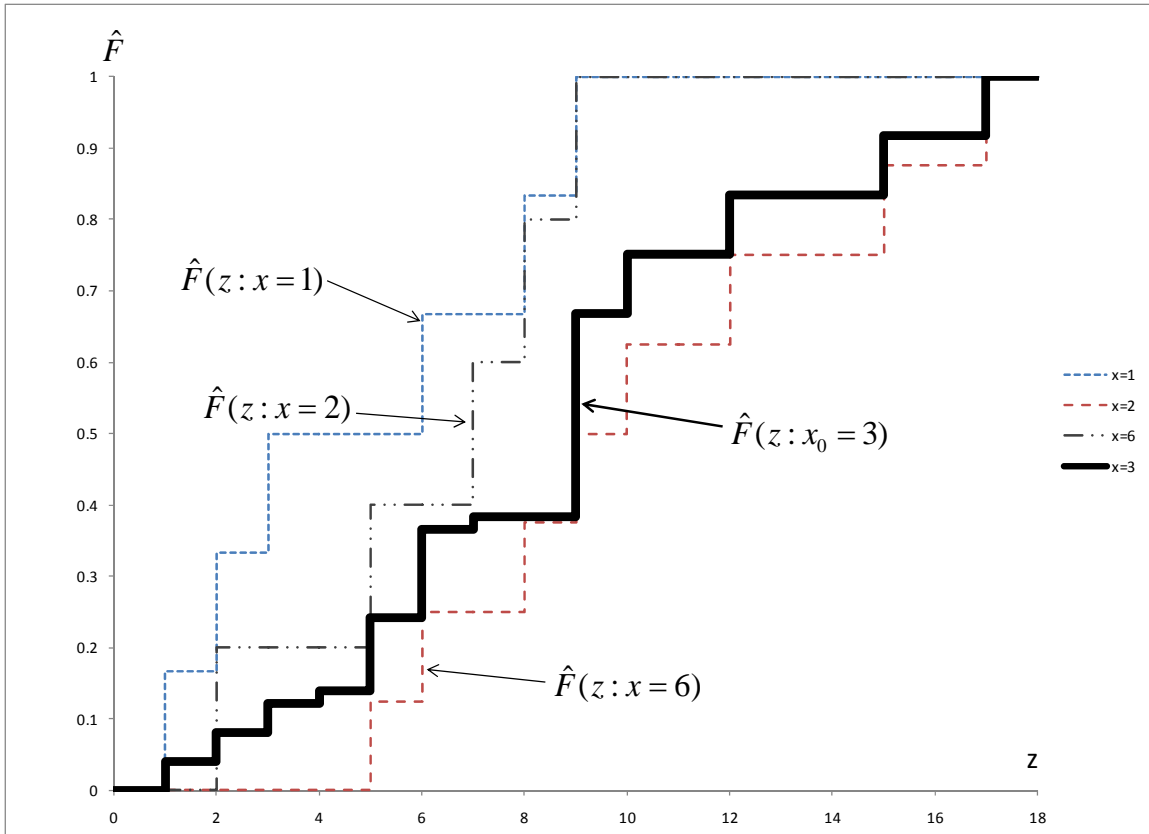


Figure 2a
 Similarity Based Quantiles
 Alpha=.1

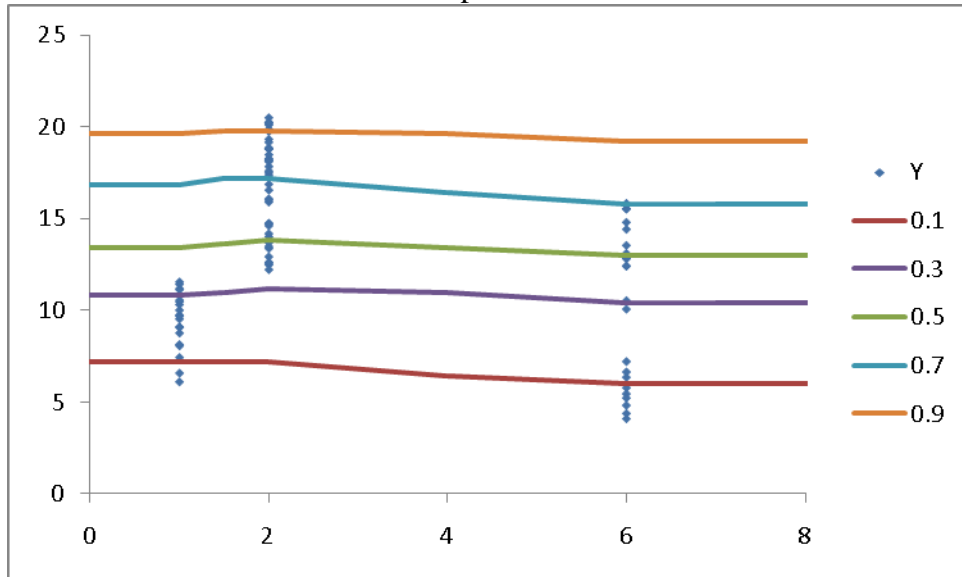


Figure 2b
 Similarity Based Quantiles
 Alpha=1

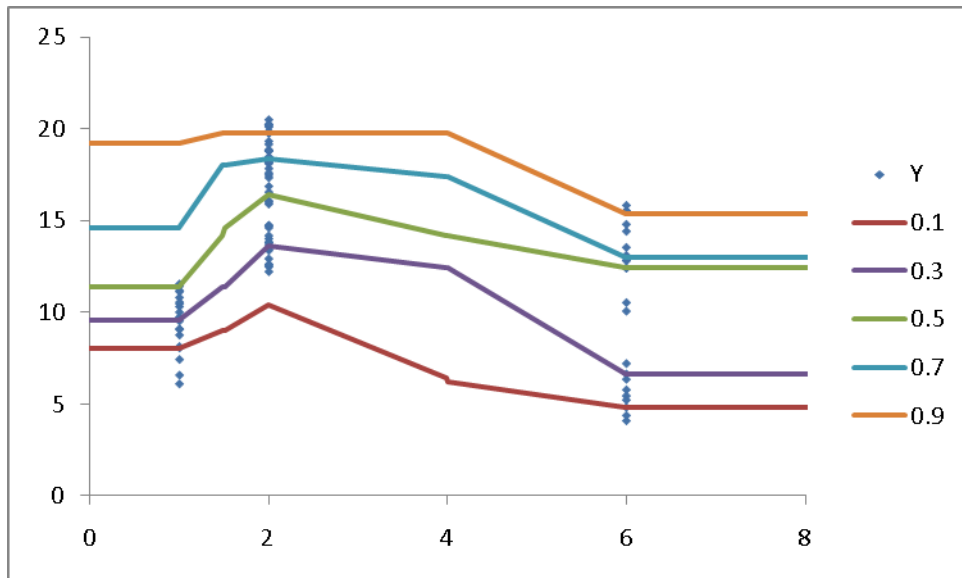


Figure 2c
Similarity Based Quantiles
Alpha=10

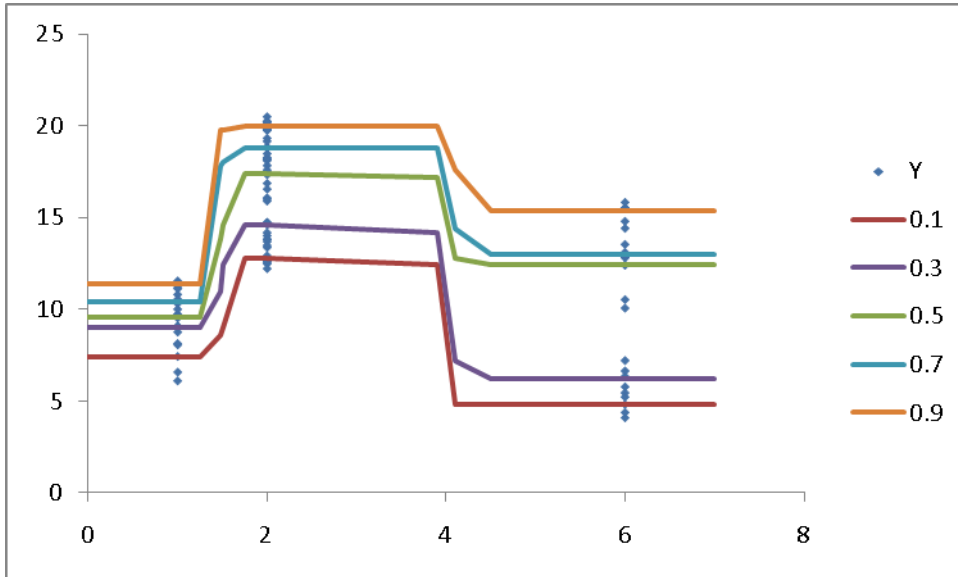


Figure 2d
Similarity Based Quantiles
Alpha=100

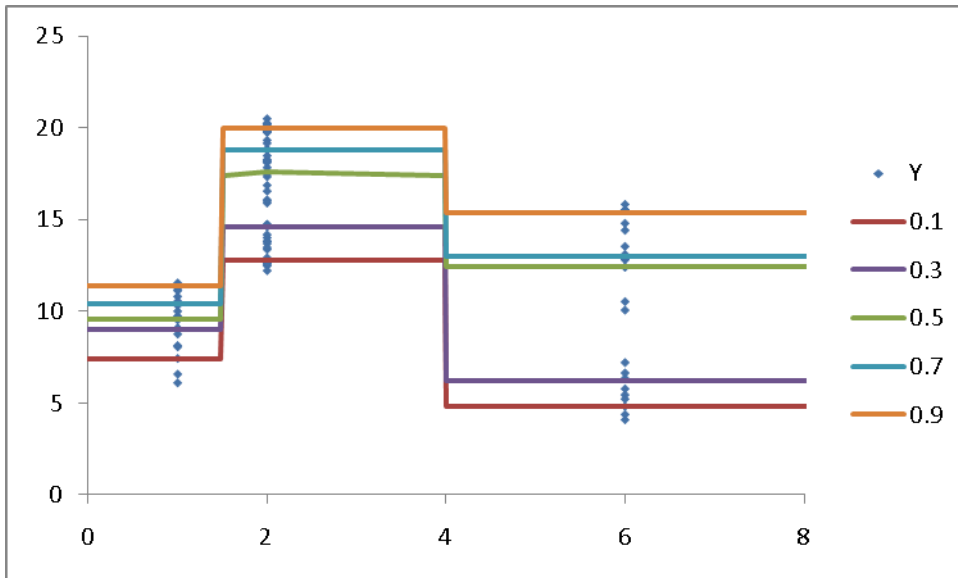


Figure 2e
Linear Model
OLS and Regression Quantiles

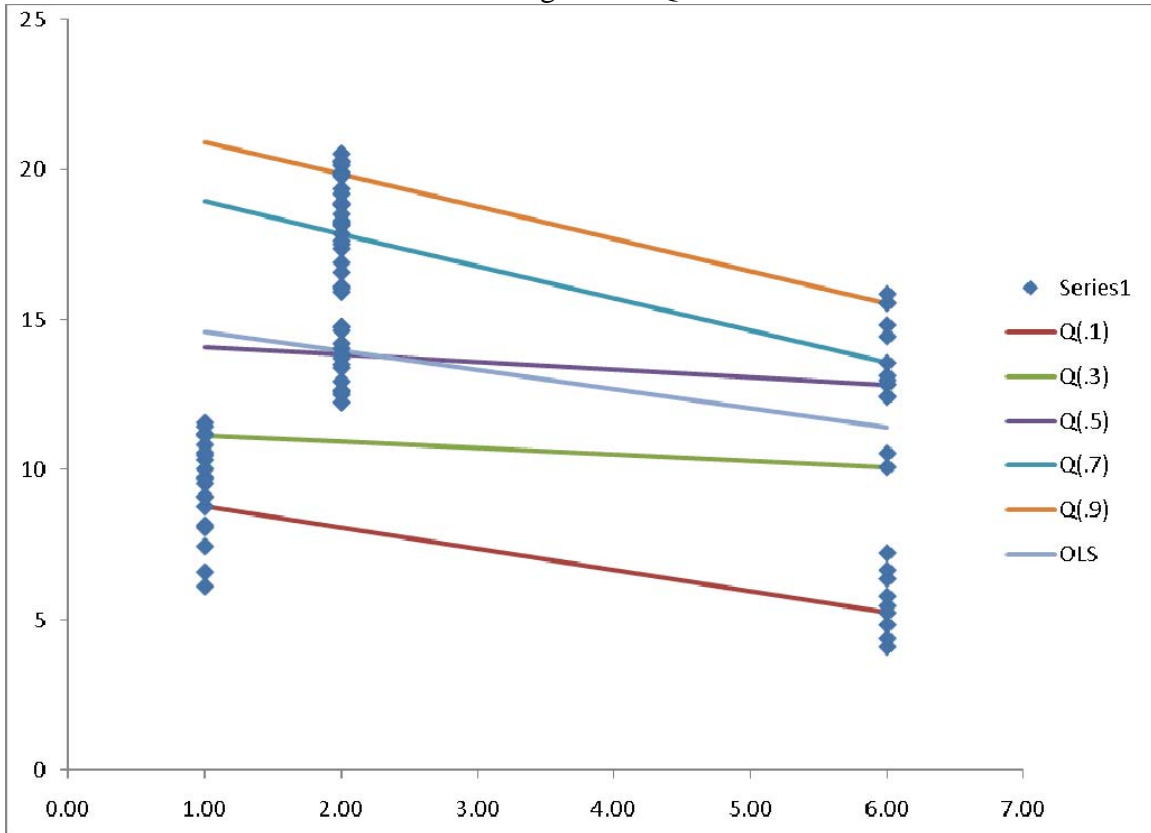


Figure 3a
Similarity Based Quantiles
 $\alpha = 1$

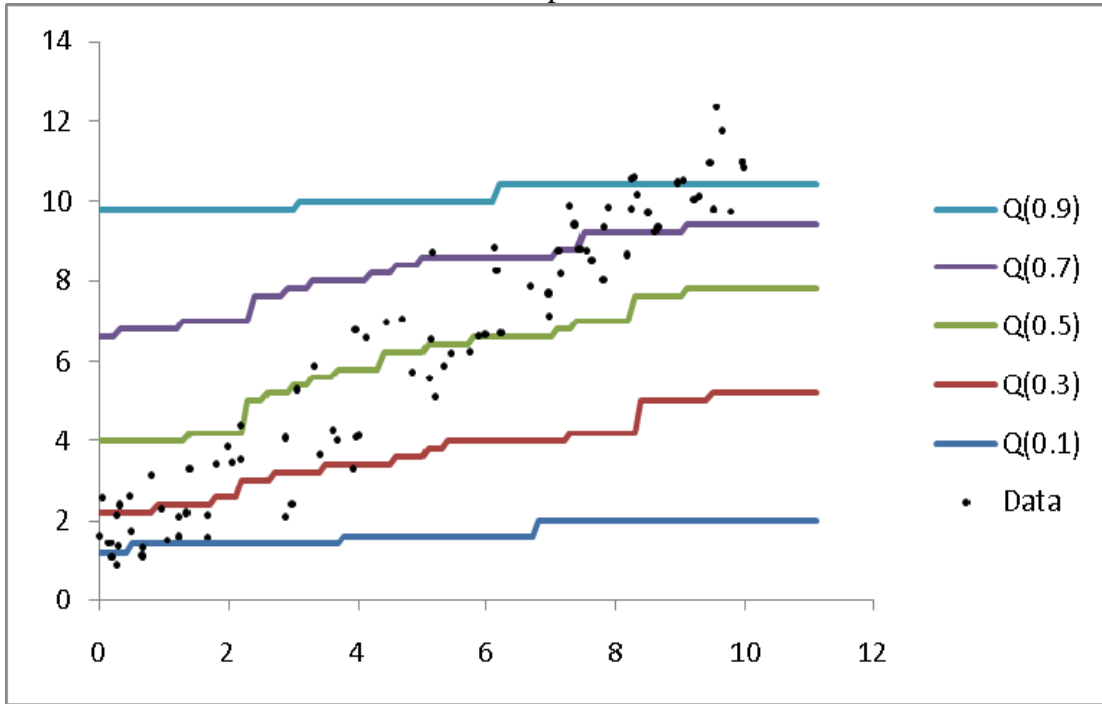


Figure 3b
Similarity Based Quantiles
 $\alpha = 0.5$

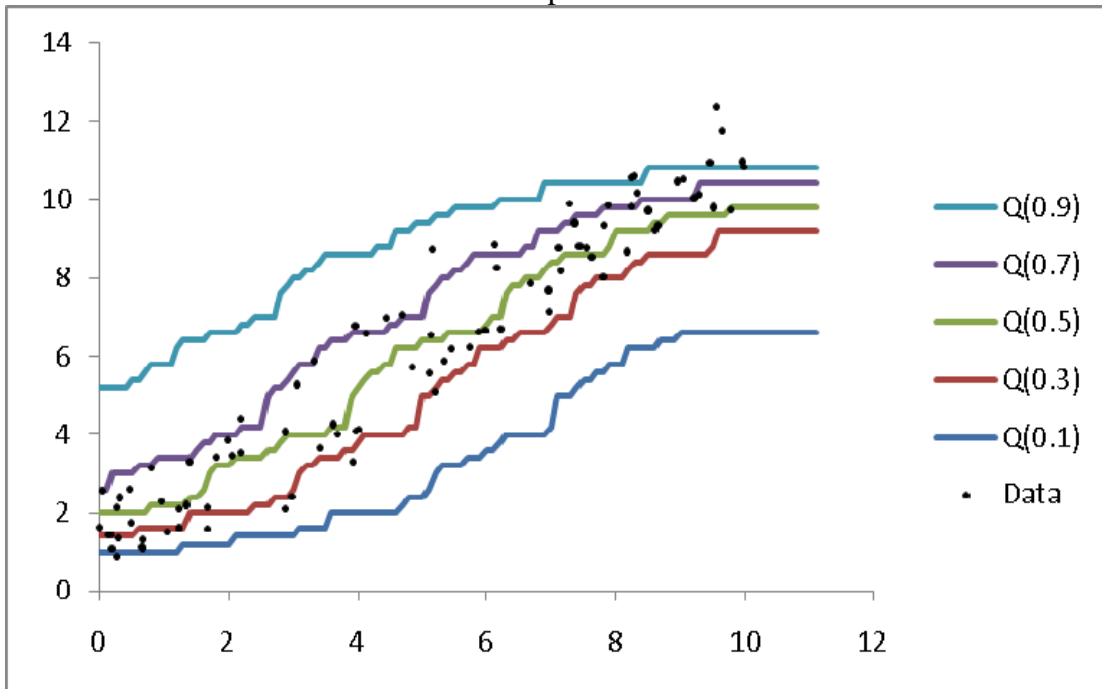


Figure 3c
Similarity Based Quantiles
Alpha = 1

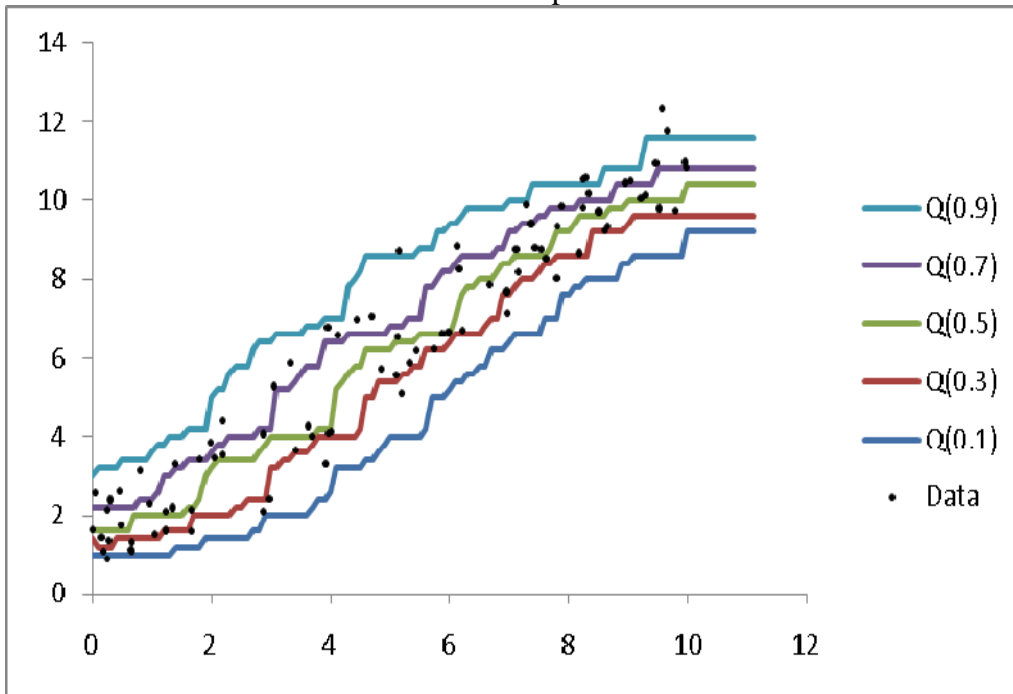


Figure 3d
Similarity Based Quantiles
Alpha = 2

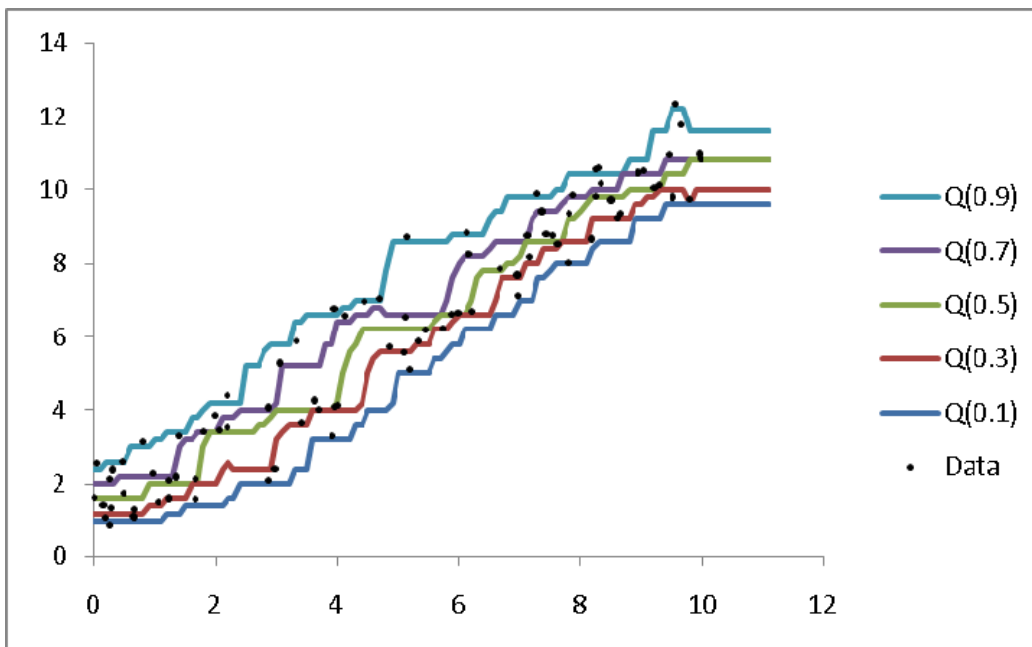
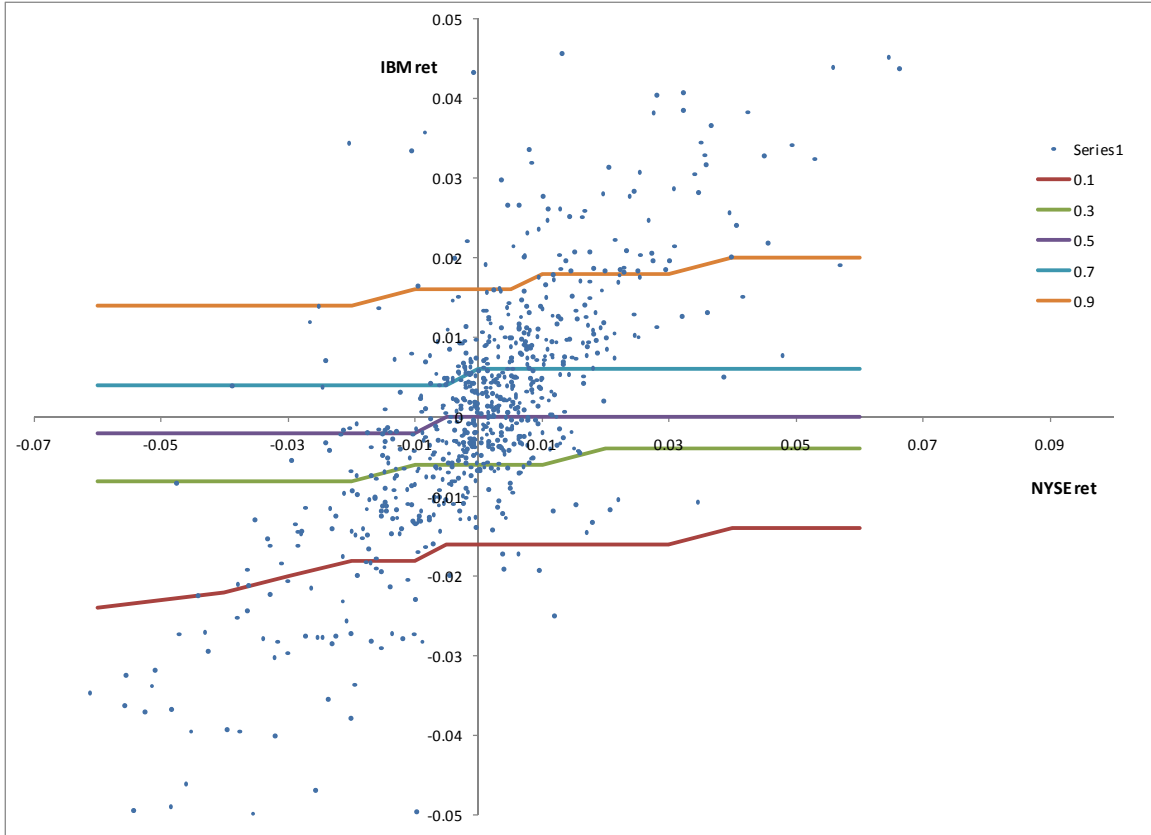


Figure 4a
 IBM Daily Return vs Market Daily Return
 Various Quantiles
 Alpha = 10



Notes: Data: IBM (y_t) and NYSE Composite (x_t) daily returns, May 27, 2008 to June

10, 2011; ($t=1, \dots, 768$). $F_t(z | x_0) = S^{-1}(x_0) \sum_{t=1}^{768} s(x_t, x_0) I[y_t < z]$ for

$z \in [-.05, -.048, -.046, \dots, .096, .098, .10]$,

$x_0 \in [-.05, -.04, -.03, \dots, .03, .04, .05]$. $Q(\theta | x_0) = \min\{z : \theta \leq F_A(z | x_0)\}$,

$\theta \in [.1, .3, .5, .7, .9]$

Figure 4b
IBM Daily Return vs Market Daily Return
Various Quantiles
Alpha = 100 (See notes with 4a)

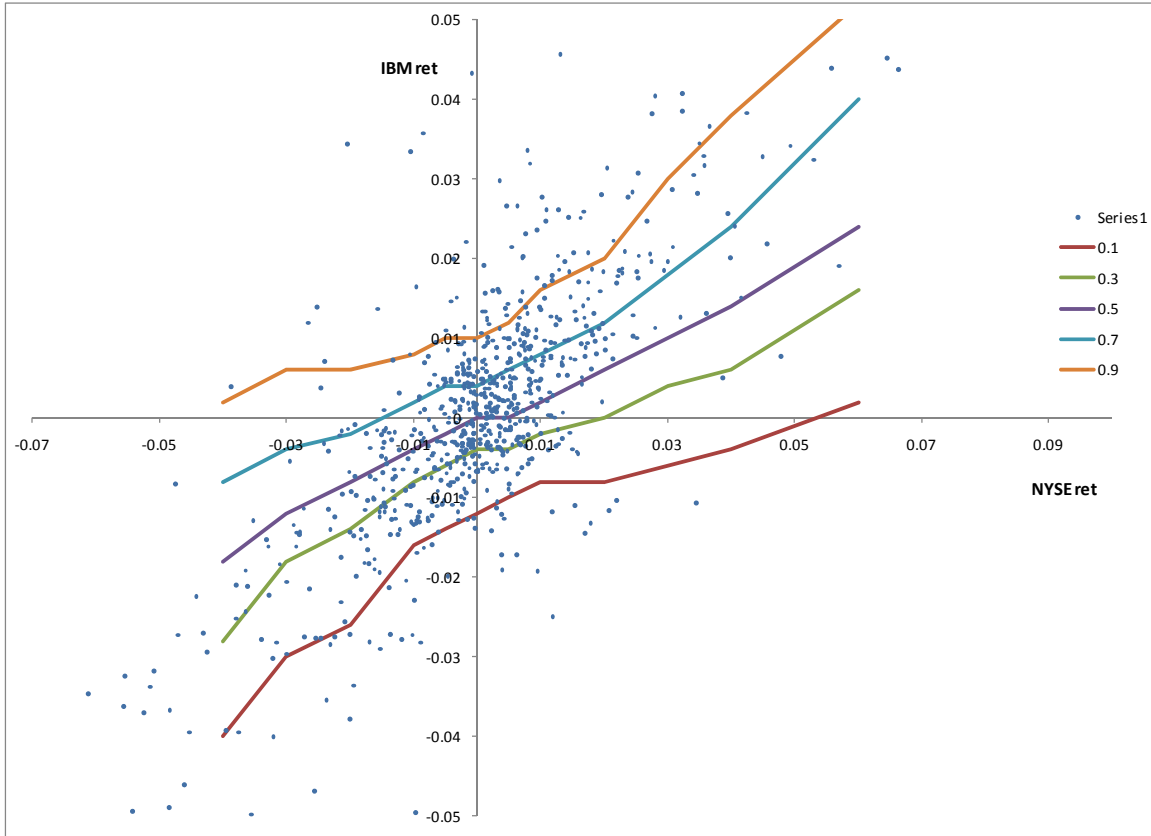


Figure 4c
IBM Daily Return vs Market Daily Return
Various Quantiles
Alpha = 500(See notes with 4a)

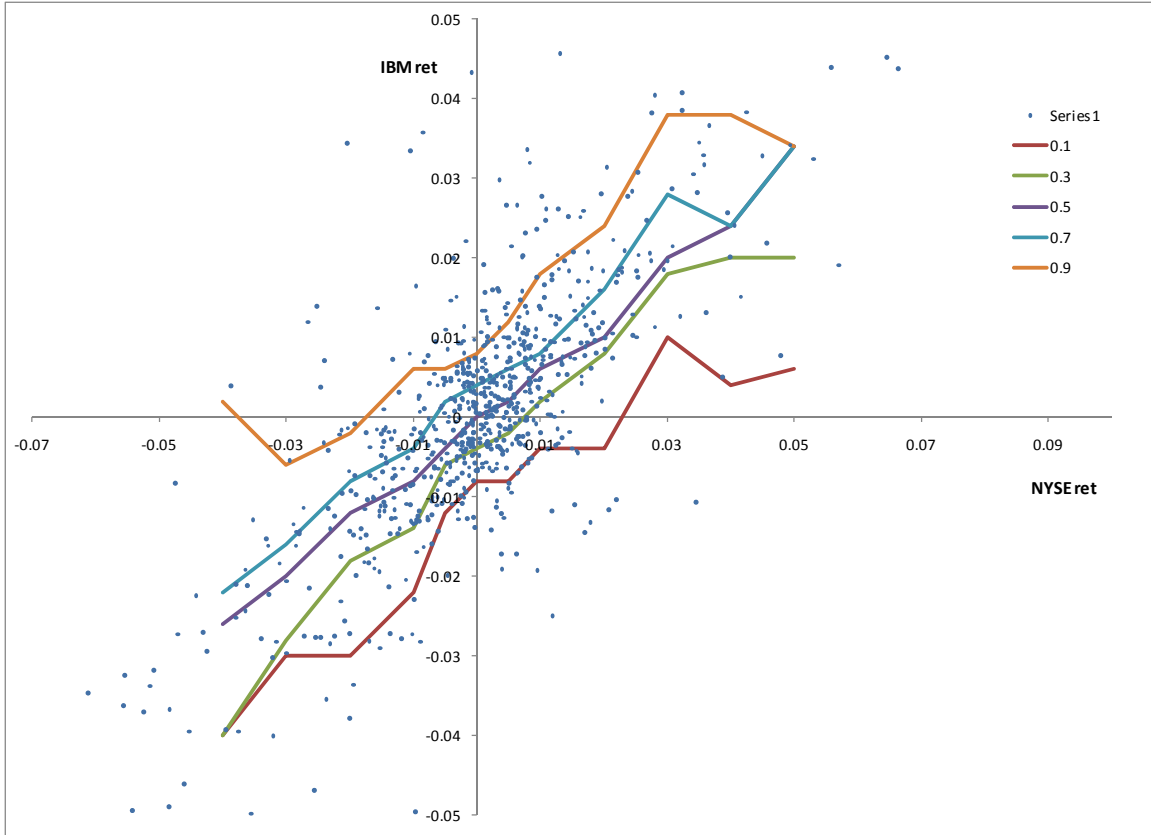


Figure 4d
IBM Daily Return vs Market Daily Return
Various Quantiles
Alpha = 1000(See notes with 4a)

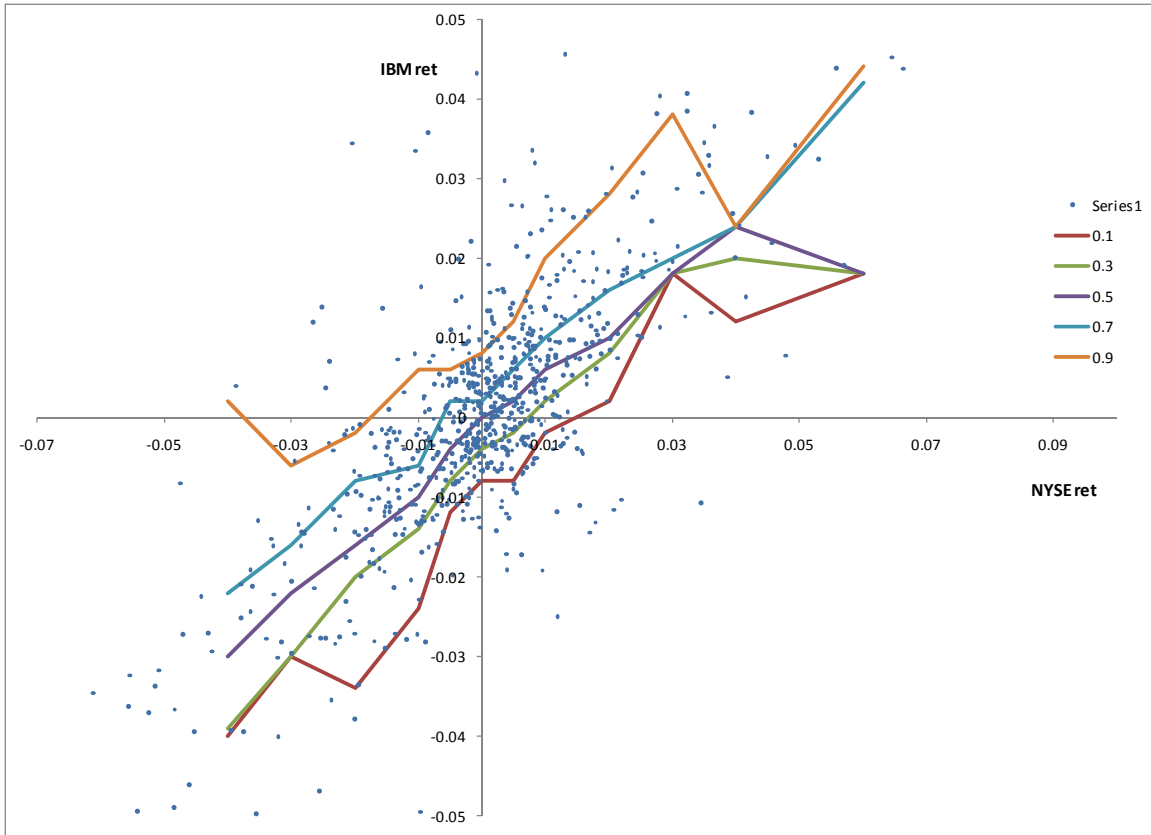


Figure 4e
IBM Daily Return vs Market Daily Return
Quantile Regression and OLS

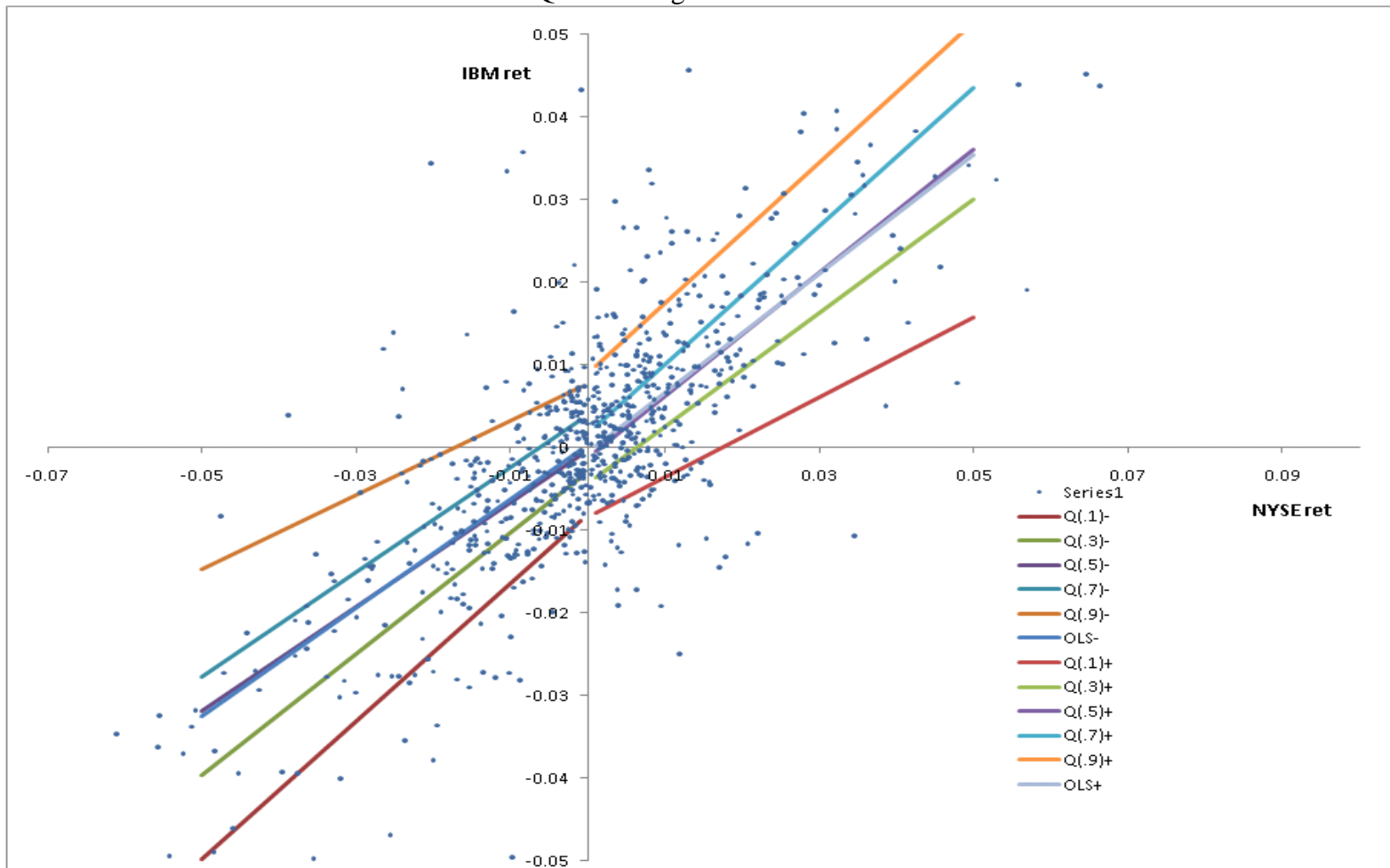


Figure 4f

Similarity function depending on m_t and t : $s(m_t, m_0) = \exp(-\alpha_1 |m_t - m_0| - \alpha_2 t)$

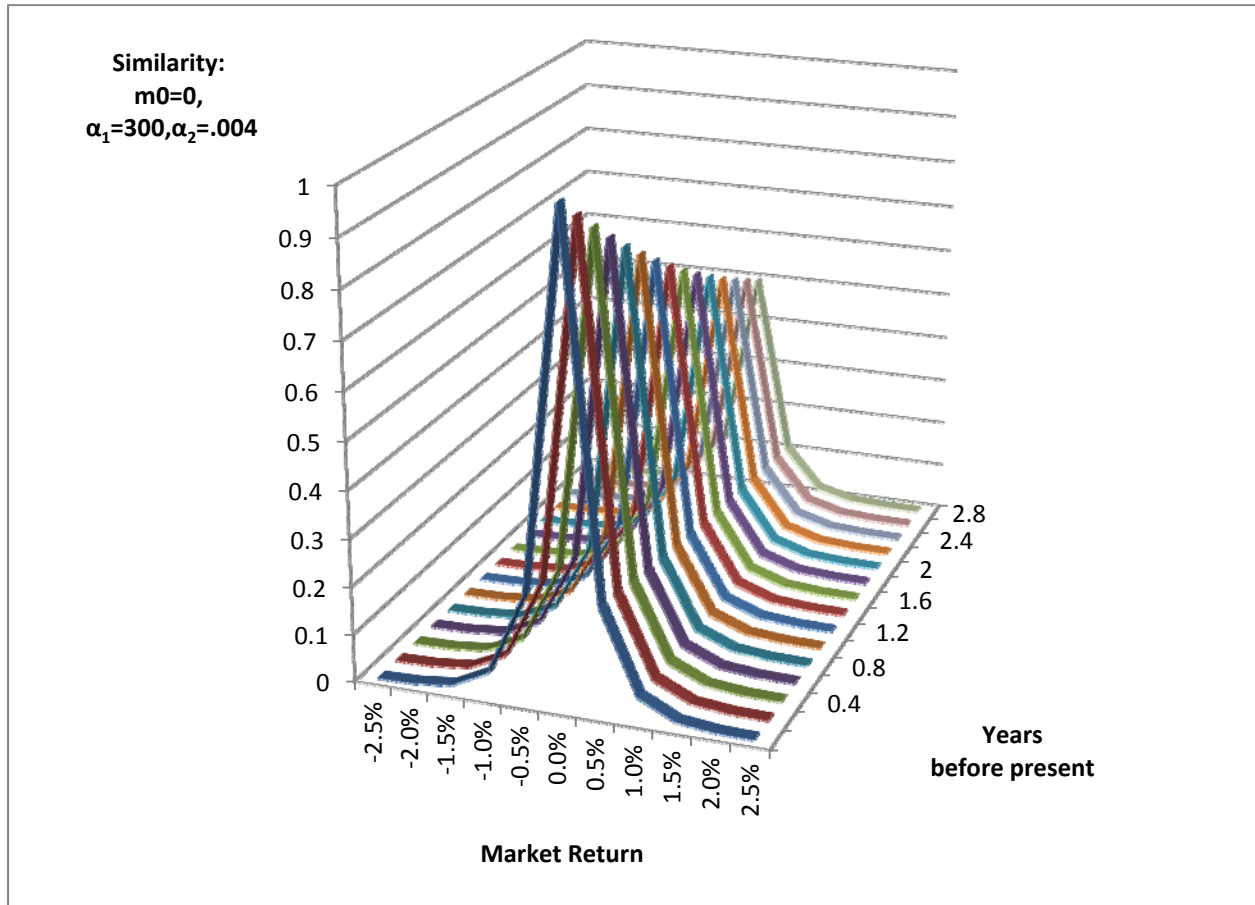


Figure 5a

$\text{VaR}(.1|x), x=+.02, 0, -.02$

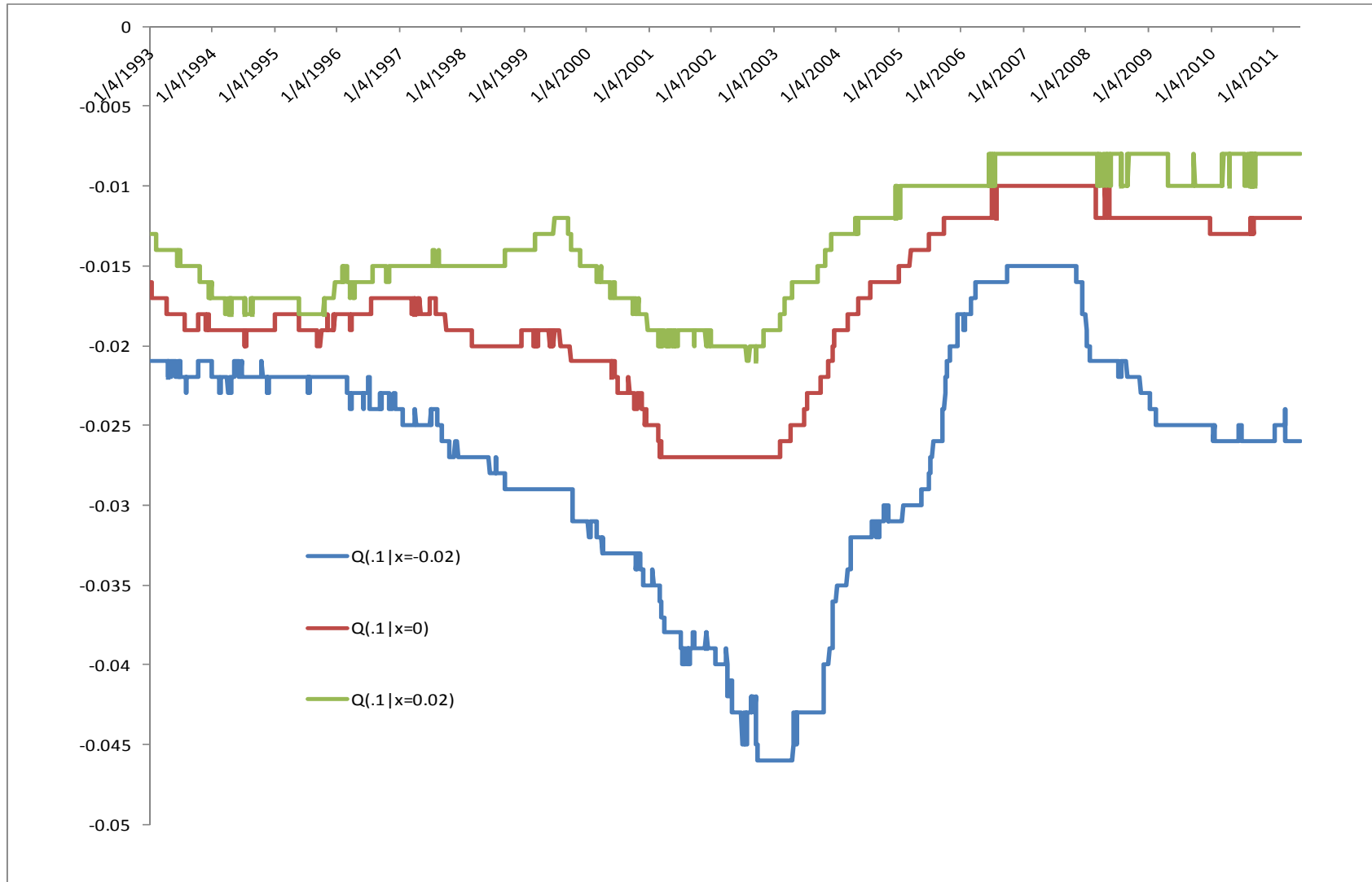


Figure 5b

VaR(.1 | x=-1%); see text

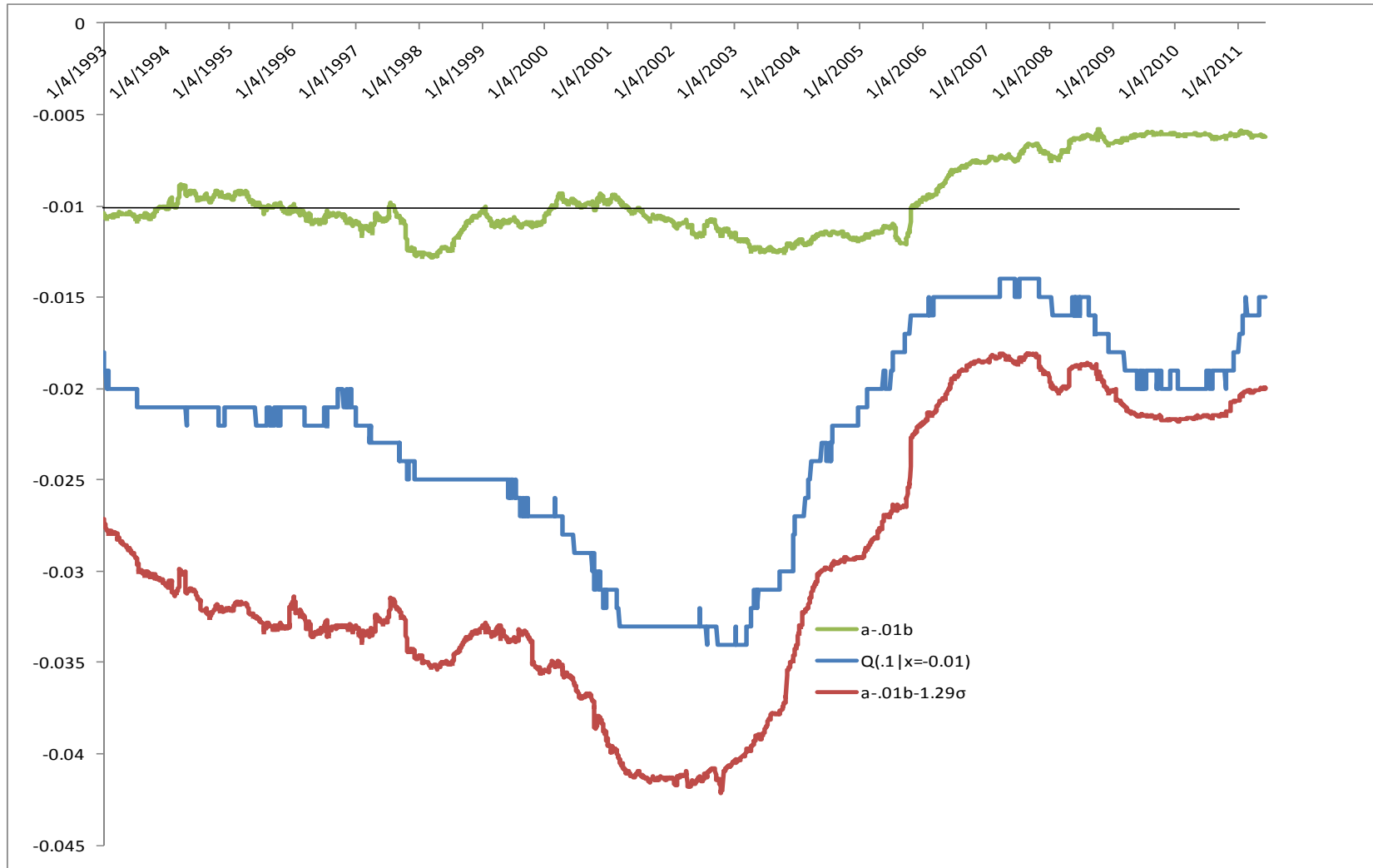


Figure 5c

VaR(.1|x=1%): see text

