Interpreting Expectiles*

Collin S. Philipps[†][‡]

June 2, 2022

Abstract

This article establishes how expectiles should be understood. An expectile is the minimizer of an asymmetric least squares criterion, making it a weighted average. This also means that an expectile is the conditional mean of the distribution under special circumstances. Specifically, an expectile of a distribution is a value that *would be* the mean if values above it were more likely to occur than they are. Expectiles summarize distributions in a manner comparable to quantiles, but quantiles *are* expectiles in location models. The reverse is true in special cases. Expectiles are m-estimators, m-quantiles, and L_p-quantiles, families which connect them to the majority of statistics commonly in use.

Keywords: Expectile Regression, Generalized Quantile Regression

JEL Codes: C0, C21, C46

1 Introduction

Expectile regression is a latent topic that has attracted attention only recently. A continuum of location parameters ranging from the infimum to the supremum of a distribution F and indexed by $\tau \in (0, 1)$, the *expectiles* of F are often compared to quantiles but rarely studied on their own. This article will firmly establish what expectiles *are* and what expectiles *are not.* To achieve that goal, we will present nine interpretations together with a detailed analysis of mathematical properties which set expectiles apart from other statistics.

The τ^{th} expectile of a distribution F would be the mean of F if values above it occurred $\frac{\tau}{1-\tau}$ times as often as they do. In one application of this fact, Philipps [2021a] has applied expectile regression to Boston's Home Mortgage Disclosure Act data and found that racial disparity in mortgage application

*I am grateful to Roger Koenker, without whom I would not have discovered this topic.

†Address correspondence to Collin S. Philipps, Department of Economics and Geospatial Sciences, 2354 Fairchild Drive, Suite 6k-110, United States Air Force Academy, CO 80840; collin.philipps@afacademy.af.edu.

‡The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the United States Air Force Academy, the Air Force, the Department of Defense, or the U.S. Government. PA#: USAFA-DF-2022-20

denial rates is up to twice as large for individuals who are more likely than average to be denied. On the other end of the spectrum, racial differences in denial rates are smaller for individuals who are relatively *un*likely to be denied a mortgage. Expectiles are a valuable addition to the econometrician's toolkit, but they are still a novelty.

First, expectiles can be defined as the minimizers of asymmetrically weighted least squares criteria: in the special case where $\tau = .5$, the weights are symmetric and expectile regression is Ordinary Least Squares. It is obvious in hindsight that an expectile is the mean of a *weighted* distribution, but this fact has been broadly overlooked. It is also worth noting that the asymmetric least squares criterion that elicits expectiles can be directly interpreted in many contexts, such as forecast evaluation or in risk management. These important points establish what expectiles *are*–and do so without leaning on any abstruse analogies.

Next, we consider the relationship between expectiles and quantiles. A great deal of ceremony has been made of the fact that expectiles and quantiles of a distribution are sometimes the same set of values. But in location models, *every* reasonable location parameter is an expectile: expectiles span the full support of a distribution, regardless of whether its quantiles have that property. The behavior of expectile and inverse-expectile functions for F can be compared to quantile and inverse-quantile functions: quantiles are continuous only in cases where the distribution F has density, but continuity of the expectile function is guaranteed in every case. We illustrate these facts with several expectile functions and their inverses. Equivalence of quantile and expectile *regression lines* is also a major topic in the literature, but this equivalence fails except in very special cases. We hope our discussion will serve as a rigorous introduction to expectile and inverse-expectile functions without reinforcing any myth that quantiles and expectiles have the same properties or that quantiles and expectiles of a given distribution coincide in any general case.

Comparisons between expectiles and quantiles are inadequate to establish where expectiles reside on the horizon of statistics. So, finally, we will articulate the relationship between expectiles and broader classes of statistics, such as m-statistics or m-quantiles. Being m-statistics, for example, we see that there is nothing particularly exotic about methods used to estimate expectiles or their asymptotic behavior. Expectile regression is a weighted least squares problem with a simple weighted least squares estimator and standard distribution-free asymptotics which do *not* require knowledge of the underlying distribution or its density. Expectiles are by far the most prevalent example of $m-$quantiles [Breckling and Chambers, 1988]. Expectiles are also among the best examples of the L_p-quantiles, a unified class that contains essentially all elementary summary statistics (mean, median, mode, minimum, maximum, midrange) and is likely to remain near the core of statistical theory far into the future.

1.1 Background

Despite the lack of focus on this topic, expectiles appear with increasing frequency in several branches of the literature. Expectiles' useful properties have made them important in applications such as quantitative finance and risk management. For example, Ziegel [2016] has shown that expectiles are the *only* coherent and elicitable risk measure. It has since become standard to compare every novel risk measure to expectiles. It has been 25 years since Yao and Tong [1996] discovered that expectile and quantile regression sometimes produce the same set of regression lines, but the ramifications of that fact for estimation methods were not well-known until recently. In the special case where these lines co-locate, Waltrup et al. [2015], Daouia et al. [2019b], Philipps [2021a] and others have established that expectile regression is a statistically efficient alternative to quantile regression and that it suffers less from the quantile crossing problem. In cases where quantile and expectile lines do *not* co-locate, it is known that expectile regression remains viable where quantile regression fails, such as in binary response problems and multiple equation models. Let it suffice to say that expectile regression is one of the most intriguing new methods in econometrics.

Because of their simplicity and broad applicability, expectiles are an obvious starting point for practitioners who wish to consider regression lines other than the mean. We hope the menu of topics in this article can encapsulate everything that the econometrician needs in order to begin using and interpreting expectile regression. To aid the reader, we have sequestered several examples, derivations, and proofs in the appendix. Highly-curious readers will find those examples to be valuable.

The body of this text is arranged as follows: the next subsection introduces expectiles formally and lists *nine* stylized interpretations, grouped in three groups as outlined above. Interpretations of an expectile *per se* are discussed in Section 3, relations to quantiles are discussed in Section 4, and relationships to other families including generalized quantiles are discussed in Section 5. Section 6 concludes.

2 Nine Interpretations

Why study expectiles and expectile regression at all? Expectiles are one of only a few mainstream classes of statistics that can characterize tails of a distribution. The virtues of *quantile* regression are universally lauded, but Koenker [2005, ch. 8] has catalogued a number of common data environments where quantile regression is problematic or fails entirely, while least squares regression (read: *expectile* regression) remains widely used. Binary response problems are a well-studied example: see the recent efforts of Chernozhukov et al. [2018] and Philipps [2021a] to extend binary response regression beyond the mean. In any case where quantile regression fails, expectile regression is the most popular alternative. But there is more to the story. Authors who have studied expectile regression often find it to be compelling in unexpected ways: see Kneib [2013], Waltrup et al. [2015], Ziegel [2016]. It is worthwhile to extend regression methods away from the mean in order to explore the full variation of distributions. Aside from true quantiles, expectiles are the only serious contender in that arena.

For a random variable X with distribution F and a finite first moment, $E|X| < \infty$, Newey and Powell [1987] introduced expectiles as the set of minimizers:

$$\mu_\tau \coloneqq \arg\min_{\mu} \int \varsigma_\tau(x-\mu)dF(x) \tag{2.1}$$

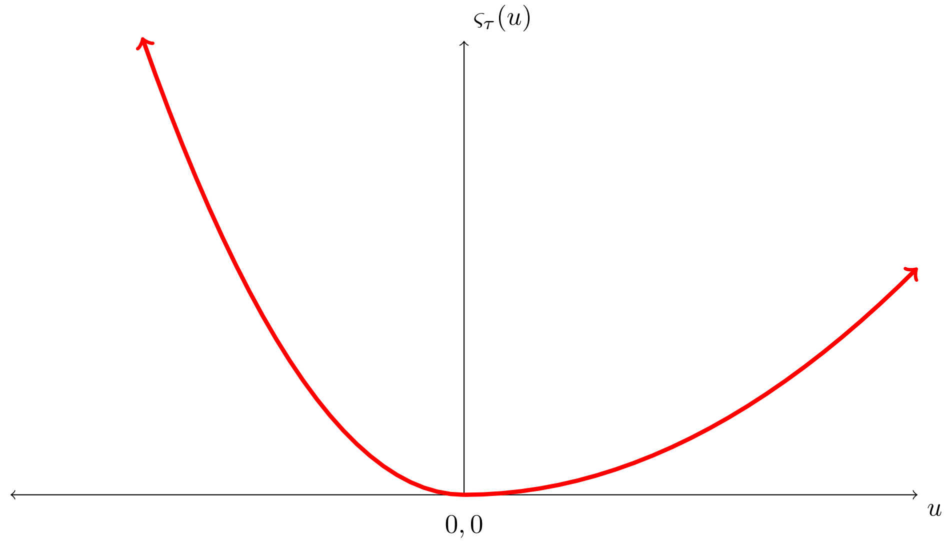

Figure 2.1: The "Swoosh" function $\varsigma_\tau(u)$ for $\tau = .2$.

where the "swoosh" function ς_τ is an asymmetric least squares criterion:

$$\varsigma_\tau(u) = \begin{cases} \tau u^2 & \text{if } u \geq 0 \\ (1-\tau)u^2 & \text{otherwise.} \end{cases} \tag{2.2}$$

This is illustrated in Figure 2.1. The τ^{th} expectile can also be characterized by the first-order-condition from equation 2.1 on the previous page, which leads to a partial moment condition:

$$\tau \int_{\mu_\tau}^{\infty} (x - \mu_\tau) dF(x) = -(1-\tau) \int_{-\infty}^{\mu_\tau} (x - \mu_\tau) dF(x). \tag{2.3}$$

A unique expectile $\mu_\tau(X)$ always exists for every $\tau \in (0,1)$. Following the common convention, we may also treat the $\tau = 0$ and $\tau = 1$ expectile as the infimum and supremum of F, respectively.

2.1 Interpretations

We begin our list of interpretations for the expectile μ_τ with those interpretations that stand *in vacuo*; drawing no ties to the L_1-quantile regression literature or other generalized families. These are interpretations of the expectile as an asymmetric mean, all of which flow directly from equation 2.1.[1]

[1] Expectiles are most often presented and discussed in implicit functional form as in equations 2.1 and 2.3. However, for discrete distributions it is straighforward to represent the τ^{th} expectile using the standard formula for a weighted average:

$$\mu_\tau = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i}$$

where $w_i = \tau$ if $X_i \geq \mu_\tau$ and $w_i = 1 - \tau$ otherwise.

Interpretation 1: An expectile is the optimal predictor under an asymmetric least squares loss function.

Interpretation 2: An expectile is the *conditional mean*, given information about the relative probabilities of positive and negative errors for an atypical observation.

Interpretation 3: An expectile is the *conditional mean* in a particular missing-data problem.

The first of these three is almost certainly obvious—it is a restatement of the definition of an expectile. The asymmetric L_2 loss function can be interpreted in terms of an agent's preferences, such as the preferences a professional forecaster uses to produce point forecasts. There is a substantial body of literature that takes asymmetric least squares loss functions seriously, such as Elliott et al. [2005] or [Ziegel, 2016]. Accordingly, we order it first.

The second and third interpretation are closely related. Both represent potential estimation problems where the un-weighted sample mean is an invalid measure of central tendency. The second interpretation applies when the probability of observing a positive or negative error term differs for *specific* observations or individuals in a data set. This is a natural way to capture unexplored variation (heterogeneity) within a population. The third interpretation applies when observations are drawn from some data-generating process but fail to be observed (go missing) at different rates depending on the sign of the error. In that case, an expectile is a superior location predictor even if the exact details of the missing data process are unknown. From the second and third interpretation, we see that expectiles can characterize data in ways that neither the mean nor quantiles are able.[2]

Next, we have three interpretations that establish a close relationship between quantiles and expectiles:

Interpretation 4: Expectiles characterize the distribution of a variable by mapping its range to the interval $(0, 1)$, with the mean of the variable corresponding to .5.

Interpretation 5: There is a functional mapping from expectiles to quantiles such that expectiles may be used to estimate quantiles by least squares.

Interpretation 6: Expectiles of a distribution F *are* quantiles, but **not** of the distribution F.

The first of these is a basic property. After demonstrating that the expectile function $\mu_\tau := E_X(\tau)$ for any distribution F is monotone increasing and continuous in τ, we need only show that the minimum and maximum of $E_X(\tau)$ are those of F, if they exist. As for Interpretation 5, the existence of a functional mapping between expectiles and quantiles can be stated as a corollary of Interpretation 4: If every *value* in the support of a distribution is an expectile, then every *quantile* must also be

[2]Though it was not stated explicitly, there is a strong hint of Interpretation 2 or Interpretation 3 in the original article by Newey and Powell [1987, footnote 2]. A more concrete version of that interpretation should be inferred immediately from the results of Breckling and Chambers [1988], who did not discuss expectiles in any detail but made an identical claim about quantiles: the τ^{th} quantile is the value that *would be* the median of F if values above it occurred $\frac{\tau}{1-\tau}$ times as often as they actually do. My conversations with others in the quantile regression area, especially my colleagues at the University of Illinois in Urbana-Champaign, lead me to believe that Interpretations 2 and 3 have gone unnoticed for two reasons. First, there have never been a large number of researchers working in this field. Second, zealous comparisons between quantiles and expectiles have led to an apparent conflation of the two topics into one. This article will help to disentangle them again.

an expectile. For the location model, it is trivial to solve for *which* expectile a given value might be: monotonicity of the expectile function for F ensures that there is always an inverse-expectile function $\tau := E_X^{-1}(\mu_\tau)$. Importantly, these properties hold even in regression models as long as the true data generating process is of the location-scale family. In that context, estimators with radically different properties can be used to estimate the same quantile regression lines. See Waltrup et al. [2015], Daouia et al. [2019b] for steps in this direction.[3]

Interpretation 6 should be handled carefully. Expectiles of F are quantiles of a distribution which always has density, which is wonderfully useful to illustrate the properties of the *set* of expectiles indexed by $\tau \in (0,1)$. The inverse of an expectile function $E_X(\tau)$ has all the properties of a CDF. And, because μ_τ is monotone increasing even when the quantile function is not, the inverse expectile function has a "density" even when the original distribution does not. This is another reason why expectiles can been called "generalized" quantiles. But, as we must emphasize, expectiles of a F are *not* its quantiles in general. Koenker [1993] made a famous exercise out of determining whether any distribution had the same quantiles as expectiles. Indeed, there is one special distribution with that property.

Expectiles also belong to three general classes of statistics which are especially well-studied. From this, we can infer that their sample and asymptotic properties are well known and well-familiar. But we may also characterize expectiles in relation to other statistics, which are more familiar to most readers.

Interpretation 7 Expectiles are M-class (and Z-class) statistics

Interpretation 8 Expectiles are m-quantiles.

Interpretation 9 Expectiles are L_p-quantiles.

The latter two, m-quantiles and L_p-quantiles, have been called "generalized" quantiles. As minimizers of loss functions, the $m-$quantile and L_p-quantiles contain the mean, median, mode, midrange, expectiles, quantiles, the minimum, maximum, and other statistics of a given distribution within one family. The L_p-quantiles of Chen [1996] nest those statistics *and their regression counterparts* within a convenient two-dimensional framework.

The following sections develop these nine stylized interpretations in detail.

3 Interpretations as Asymmetric Means

Our first three stylized interpretations follow directly from the definition of an expectile; specifically from equations 2.1 and 2.3. This section may serve as a more detailed introduction.

[3] It is common, even in the regression setting, to consider the possibility that some expectile co-locates with some quantile. Most famously, the mean and median of symmetric distributions are the same, making estimation of either value possible by a wide variety of symmetric loss functions; such as Huber's [1964] famous example.

3.1 Interpretation 1: An expectile is the optimal predictor under an asymmetric least squares loss function.

This interpretation is equivalent to the definition of an expectile given in equation 2.1.

The fundamental goal of statistics is to reduce complicated data into simpler terms so that interpretations and decisions can be made. Often, this involves summarizing a sample using a single location parameter that fits or predicts the data well. Obviously, we cannot achieve that goal without precisely defining what constitutes a good fit or prediction. Depending on the statistician's preferences, an expectile may be the best possible prediction. And this result is not a mere tautology: axiomatic approaches to quantitative finance have led to the result that expectiles may be the only statistics that can possibly satisfy a agent's preferences for risk measures.

Take a random variable X with distribution F. Our hypothetical statistician is tasked to give a point prediction $\hat{x}$ for X. After X is observed, the quality of the predictor $\hat{x}$ is evaluated by comparison with the outcome $X = x$. The most traditional example of analytically representable preferences for this problem (a loss function, in this case) is the least-squares criterion:

$$\begin{aligned} L(x, \hat{x}) &= (x - \hat{x})^2 \\ &= \|x - \hat{x}\|_2^2 \end{aligned}$$

where $\|\cdot\|_p$ denotes the p-norm and a *greater* loss $L(\cdot)$ is considered a *worse* outcome. It is well-known that the arithmetic mean (expected value) of X is the optimal point forecast under the squared error loss function in the sense that it minimizes the expected loss:

$$\text{mean}(X) \in \arg\min_{\hat{x}} \int L(x, \hat{x}) dF(x). \tag{3.1}$$

And ordinary least squares regression achieves the same goal in a regression environment. However, it is not always desirable for the loss function $L(x, \hat{x})$ to be symmetric about the origin as in the OLS case. In financial applications, for example, it is better to earn one dollar *more than expected* than it is to earn one dollar *less*, even when both outcomes are dominated by perfect foresight. Accordingly, there has been a substantial amount of research on asymmetric loss functions and their properties in finance: see Bellini and Di Bernardino [2017], Daouia et al. [2019a, 2020], Taylor [2019]. When the loss function $L(x, \hat{x})$ is similar to the least squares type, but positive and negative errors are weighted differently, we may write the loss function

$$L(x, \hat{x}) \propto (x - \hat{x})^2 |\tau - I(x < \hat{x})| \tag{3.2}$$

for τ such that the ratio of weights for positive and negative errors is $\frac{\tau}{1-\tau}$. In that case, it is clear that the optimal point predictor will be an expectile:

$$\mu_\tau(X) \in \arg\min_{\hat{x}} \int L(x, \hat{x}) dF(x).$$
$$= \arg\min_{\hat{x}} \int \varsigma_\tau(x - \hat{x}) dF(x).$$

In a remarkable result, Ziegel [2016] has found that asymmetric loss functions of this type are the "best" for risk management applications in the sense that expectiles are the *only* coherent and elicitable risk measures.[4] While coherence and elicitability are desirable, it is reasonable to ask whether any alternative loss functions or methods might also elicit parameters with these properties. The answer is a hard no: Gneiting [2011] has shown that the loss function in equation 3.2 on the preceding page is the unique eliciting score function for expectiles for the class of distributions with support on the reals. Thus, if an agent's preferences for a point prediction require coherence and elicitability as advocated by Ziegel [2016] and others, the loss function in equation 3.2 is the *only* loss function that can represent the agent's preferences. Expectiles are the only rational point predictions for an agent with these preferences, and asymmetric least squares is the only consistent method of estimating said parameters.

This is a strong endorsement: under some simple but realistic conditions, the optimal point predictor is an expectile. The same result extends to multivariate applications, where an expectile regression line would be the line of best fit in the colloquial sense.

3.2 Interpretation 2: An expectile is the *conditional mean*, given information about the relative probabilities of positive and negative errors for an atypical observation.

If expectiles are optimal in the sense described above, how should we interpret them? An expectile (for any τ) has a straightforward interpretation as a *conditional* mean of X. The τ^{th} expectile satisfies the moment condition in equation 2.3, which itself can be interpreted directly. From equation 2.3, we have

$$\frac{\tau}{1-\tau} \int_{\mu_\tau}^{\infty} |x - \mu_\tau| dF(x) = \int_{-\infty}^{\mu_\tau} |x - \mu_\tau| dF(x). \tag{3.3}$$

It is well-known that the mean of X is the unique value of μ_τ where these integrals are equal and τ is .5.

But from the equation above, it should be immediately clear that, if values in $[\mu_\tau, \infty)$ were $\frac{\tau}{1-\tau}$

[4] A functional $T(F)$ is *elicitable* relative to a class of probability measures if there exists a scoring function that is strictly consistent for $T(F)$ relative to that class of measures; the scoring function $\varsigma_\tau(x-\hat{x})$ is strictly consistent for the τ^{th} expectile relative to distributions with a finite first moment. For risk measures in financial applications, *coherence* is a collection of desirable properties including monotonicity, subadditivity, positive homogeneity, and translation invariance (see Artzner et al. [1999]).

times as likely to occur as they do according to F, the ratio of partial moments would be one;

$$\int_{\mu_\tau}^{\infty} |x-\mu_\tau| dG(x) = \int_{-\infty}^{\mu_\tau} |x-\mu_\tau| dG(x)$$

$$\text{if } dG(x) \propto \begin{cases} \frac{\tau}{1-\tau} dF(x) & \text{if } x \in [\mu_\tau, \infty) \\ dF(x) & \text{if } x \in (-\infty, \mu_\tau). \end{cases} \tag{3.4}$$

We have just proven the following proposition.

Proposition 1. *For any given $\tau \in (0,1)$ the corresponding expectile μ_τ is the value that* would *be the mean of X if observations in the interval $[\mu_\tau, \infty)$ occurred $\frac{\tau}{1-\tau}$ times as often as they do according to F.*

Despite its obvious applications in economics and finance, this interpretation of the τ^{th} expectile is remarkable. The intuition is similar to but not identical to approaches used in scenario analysis, where external information is provided to explore how a model's outcomes vary contingent on possible events. A wide range of possibilities exist for the *conditional* distribution of X when information about X is available prior to its observation. In the most extreme case, $X|X = x$ is non-stochastic. When we consider the expectiles of $X \sim F$, we assume that heterogeneity exists within a sample and some observations are more likely (or less) to be above (or below) the mean.[5]

When the observed distribution F differs from the true data generating process G in an analytically tractable way as in equation 3.4, it may be straightforward to derive the distribution G and calculate its mean.[6] Three concrete examples are given in the appendix, with examples for known distributions given in Section 4. We remark that the choice of $\tau \in (0,1)$ determines how much more likely a value above μ_τ becomes: $\tau = 2/3$ implies that X_i is twice as likely as usual to exceed its own mean, $\tau = 4/5$ means that X_i is 4 times as likely as usual to exceed its own mean, and so on. When the dependent variable is a measure of wellbeing as is often the case in social sciences, an expectile can be interpreted as the expected value of a variable when things are "better" or "worse" than usual in a particularly simple probabilistic way.

3.3 Interpretation 3: An expectile is the *conditional mean* in a particular missing-data problem.

An expectile can be interpreted as the conditional mean for a particular observation when heterogeneity exists within a sample. There is a trivial extension of this fact that is useful in some cases. Suppose that the observed data X are distributed according to F as before, but observations are missing. If only positive errors ($X_i \geq \mu$) are missing, only negative errors are missing ($X_i < \mu$), or both positive *and* negative errors are missing but with different probabilities, we can revisit the reasoning used in the previous subsection to identify the *un*conditional mean.

[5]Our claim in equation 3.4 should be credited to Breckling and Chambers [1988] who made a more general version of the same point.

[6]We would remark that the re-weighting of different parts of a distribution's support is also a standard approach in survival analysis and censored regression: see Tobin [1958], Amemiya [1973, 1974], famously. Such weighting is used to compensate for bias in a variety of contexts: see Heckman's [2001] extensive work on selection, heterogeneity, and missing data for which he received The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel.

Regardless of which of these scenarios is the case, the true (latent) distribution G has α times the usual ratio of positive errors to negative errors for some $\alpha \in \bar{\mathbb{R}}^{+}$. We can write this as

$$\frac{(1-G(\mu_i)/(1-F(\mu_i))}{G(\mu_i)/F(\mu_i)} = \alpha = \frac{\tau}{1-\tau}.$$

Then the optimal predictor of the latent random variable X is not the mean of the sample. Rather, the same logic as before gives us

$$\begin{aligned}
\mu_i &= \int x dG(x) \\
&= E\left(X_i | \Pr(X_i > \mu_i) = \tau\kappa(1-F(\mu_i))\right) \\
&= (1-\tau)\kappa \int_{-\infty}^{\mu_i} x dF(x) + \tau\kappa \int_{\mu_i}^{\infty} x dF(x) \\
&= \frac{\int x|\tau - I(x < \mu_\tau)| dF}{\int |\tau - I(x < \mu_\tau)| dF}
\end{aligned}$$

–familiar readers will recognize this as the τ^{th} expectile of the observed distribution F. In essence, the fact that data are missing differently for positive and negative residuals means that *every* observation is "atypical" in the same sense as in the previous section: the true data generating process has $\frac{\tau}{1-\tau}$ times the relative odds of positive versus negative errors.

It is not necessary to know the ratio of missing observations. As in the previous section, the set of indices $\tau \in (0,1)$ spans the full continuum of possibilities. This is important for practitioners: if we suspect that asymmetric missing data or atypical observations *might* be occurring, it is no great burden to estimate multiple expectiles to survey the entire spectrum $\tau \in (0,1)$. In a linear regression case, we may similarly estimate a survey of *many* expectiles even if we are only interested in the true mean regression coefficients. If the coefficients have a similar economic interpretation and similar statistical significance across a full range of τ, then we know that the implied results are robust to potential misspecification due to heterogeneity within the sample. In that case, we also know that treatment effects are also relatively constant across the distribution. Following that logic, Newey and Powell [1987] have studied asymptotic test of the hypothesis that expectile regression coefficients are constant across all τ.

The practice of estimating many different expectiles in order to explore the full variation across a distribution is identical in spirit to the way that *quantile* regression is usually used. The next section will establish similarities and differences between these two methodologies.

4 Comparisons to Quantiles

Are expectiles like quantiles? The two are often compared, but they are as different as they are alike. In this section, we will make the comparison between quantiles and expectiles plain. The similarities should be clear and the differences clearer.

Comments regarding the notation used in this section can be found in the appendix. For a detailed introduction to quantiles, see Koenker [2005].

4.1 Interpretation 4: Expectiles characterize the distribution of a variable by mapping its range to the interval $(0, 1)$, with the mean of the variable corresponding to $.5$.

Expectiles *are* like quantiles–albeit only in a very broad sense. Both of these sets of statistics fall within the convex hull of the support of a random distribution. Both of them are indexed by $\tau \in (0, 1)$. In other words, the quantile function $Q_X(\tau)$ and the expectile function $E_X(\tau)$ are both mappings $(0, 1) \mapsto (\inf X, \sup X)$. But these mappings are quite different.

Unlike quantiles, expectiles are a *bijective*, *invertible* mapping from the unit interval to the range of a random variable. As a result, every reasonable location parameter (a value within the range of the distribution) is an expectile.[7] Quantiles sometimes have this property, but only in cases where the distribution F has density almost everywhere in the range of X.

Proposition 2. *For a random variable X with $E|X| < \infty$, the expectile function $E_X(\tau) : (0, 1) \mapsto (\inf X, \sup X)$ is a bijective invertible mapping and $E_X(\tau)$ is continuous, monotone increasing. The inverse expectile function E_X^{-1} is a bijective invertible mapping $(\inf X, \sup X) \mapsto (0, 1)$ and E_X^{-1} is continuous, monotone increasing.*

A straightforward proof is given in the appendix. What can we do with this proposition? Because of continuous monotonicity, every value $\tau \in (0, 1)$ corresponds to a unique value in the convex hull of the support of X. The reverse is also true. As such, this exhaustive set of values within the range of X (expectiles) is the *natural set of location parameters* for the distribution of X–all values that could be used to characterize where X occurs. The mean of X is the most popular location parameter for general distributions and corresponds to the value $\tau = .5$.

In discrete distributions where F is neither continuous nor monotone increasing, not all values within the convex hull of the support will be quantiles. Quite the contrary: the set of quantiles has Lebesgue measure zero within the convex hull of the support: in that case, it is correct to say that almost all expectiles are *not* quantiles. The empirical distribution of a sample $\mathbb{F}_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i \leq x)$ is neither continuous nor monotone increasing, so almost all sample expectiles will also *not* be sample quantiles.

Because both quantiles and expectiles are contained within the convex hull of the support of a distribution, it should be clear that the quantiles of F are a strict subset of the expectiles of F, regardless of whether F has density. As a result, it is possible to estimate quantiles using asymmetric least squares in some special cases.

4.2 Interpretation 5: There is a functional mapping from expectiles to quantiles such that expectiles may be used to estimate quantiles by least squares.

In the previous subsection, we identified that every quantile is an expectile in the location model. Two natural questions arise: can we use expectile estimation methods to estimate quantiles? If so,

[7] The Bernoulli distribution with parameter p is a good example. The parameter $p = E(X)$ is an *expectile* of the distribution, but it is not a *quantile*.

can we apply such a method to regression models? Both of these are possible.

Given some $\alpha-$level quantile $\xi_\alpha = F^{-1}(\alpha)$, which is an expectile, it is mathematically trivial [see Yao and Tong, 1996] to identify *which* expectile ξ_α corresponds to using the inverse expectile function for a generic distribution:

$$\tau = E_X^{-1}(\xi_\alpha) = \frac{\int_{-\infty}^{\xi_\alpha} |z - \xi_\alpha| dF(z)}{\int_{-\infty}^{\infty} |z - \xi_\alpha| dF(z)}. \tag{4.1}$$

Some inverse expectile functions for specific distributions can be found in Table 4.1 on page 14. But regardless of the fact that quantiles are expectiles means that a new pantheon of estimators could become available for each quantile regression line. This is quite an intriguing possibility explored by Kneib [2013], Waltrup et al. [2015], and more generally by Daouia et al. [2019b].

What about regression models? Yao and Tong [1996] were the first to point out that quantile regression and expectile regression coefficients sometimes coincide. Suppose we assume the location-scale model to be true;

$$Y = \mu(X) + \sigma(X)\epsilon \tag{4.2}$$

with μ, σ continuous functions on the support of X, $S_X \in \bar{\mathbb{R}}$, and $E(\epsilon) = 0$, $E(\epsilon^2) < \infty$, and ϵ, X independent. For the model above, the α^{th} conditional quantile and τ^{th} expectile of Y are given by the functions

$$Q_Y(\alpha|X = x) = \mu(x) + \sigma(x)Q_\epsilon(\alpha)$$
$$E_Y(\tau|X = x) = \mu(x) + \sigma(x)E_\epsilon(\tau).$$

Yao and Tong [1996] have given us the following proposition.

Proposition 3. *Let Y, X be a regression model with $E(Y|X) < \infty$ and $Y|X$ distributed according to the location-scale model in equation 4.2. For any $\alpha \in (0, 1)$ and y, x s.t. $p(y, x) > 0$, there exists some function $h(\alpha) : (0, 1) \to (0, 1)$ such that the $\tau = h(\alpha)^{th}$ regression* expectile *of Y given X is equal to the α^{th} regression quantile;*

$$E_{\tau=h(\alpha)}(Y|X = x) = Q_Y(\alpha|X = x).$$

In the expression above, the expectile operator $E_\tau(Y)$ is the linear operator that produces the τ^{th} expectile of Y. It is straightforward to show that the function $h(\alpha)$ exists and is equal to

$$h(\alpha) = \frac{E\left(|Y - E_{h(\alpha)}(Y)|I(Y \le E_{h(\alpha)}(Y))|X\right)}{E\left(|Y - E_{h(\alpha)}(Y)||X\right)}, \tag{4.3}$$

which is a generalization of the inverse expectile function in equation 4.1. The function $h(\alpha)$ can also be expressed as an explicit function of the distribution from which ϵ is drawn, say F, and its partial moments $G(z) = \int_{-\infty}^{z} u dF(u)$. A statement and derivation of that corollary to proposition 3 is given in the appendix. Equation 4.3 is easy to evaluate $h(\alpha)$ in any sample: the values $E_{h(\alpha)}(Y)$

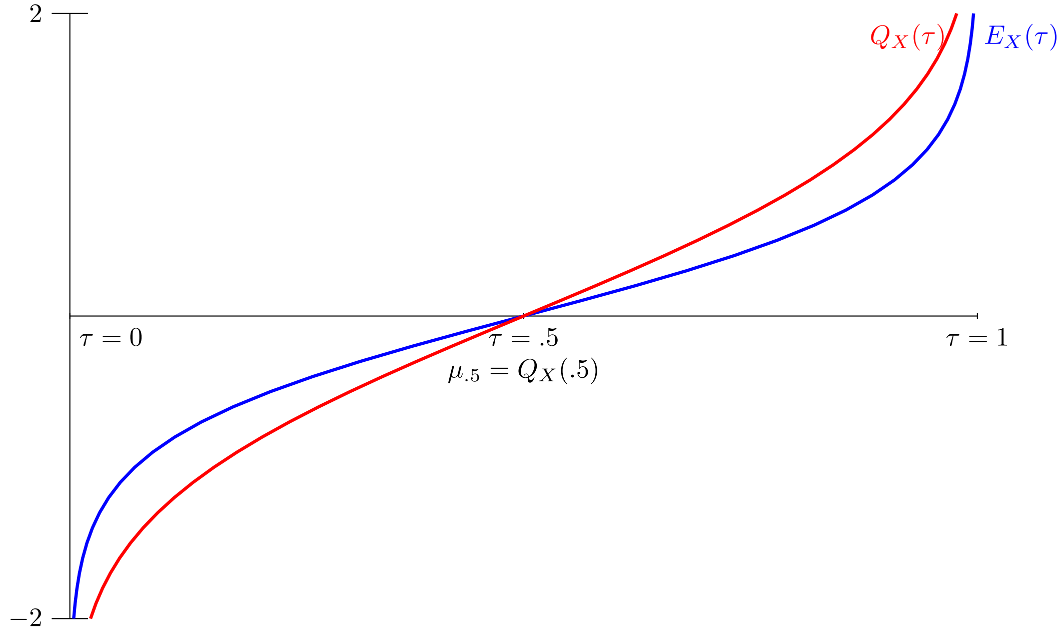

Figure 4.1: The expectile function $\mu_\tau = E_X(\tau)$ of a standard normal distribution in blue and the quantile function $Q_X(\tau)$, which is the inverse of the distribution $F(x)$, of a standard normal distribution in red.

can be replaced with their quantile equivalents and the integrals can be calculated numerically.[8]

What is the possible benefit of re-estimating sample quantiles using expectile regression? Efficiency comparisons between least squares (expectiles) and least absolute deviations (quantiles) are common in the literature [see i.e Bai, 1995, Huber, 2004] and least squares is more efficient when the tails of the distribution are sufficiently thin. Other reasons are more specific to the regression model: regression quantiles in small samples are known to suffer from the quantile crossing problem where regression lines for different α cross each other within a sample. Waltrup et al. [2015] have shown that expectiles suffer far less from this malady. Daouia et al. [2019a] argue further: the limiting distributions of quantile regression estimators involve the density (or sparsity) of the data distribution, which can only be estimated by slowly-converging methods *which fail* when F lacks density. In that worst-case scenario, an alternative density-free estimator based on expectiles is a godsend.

Readers interested in using asymmetric least squares to estimate quantiles are encouraged to see Waltrup et al. [2015] or Daouia et al. [2019b]. But do not be misled: the τ^{th} expectile is not guaranteed to be the τ^{th} quantile of a distribution nor is an expectile guaranteed to be *any* quantile, generally. The next subsection will carefully qualify one claim that appears to contradict these facts.

[8] In figure 4.1, the expectile function $E_X(\tau)$ and the quantile function $Q_X(\tau)$ are shown for a standard normal random variable. Each is a function of $\tau \in (0, 1)$, both are monotonic in this case, and invertibility in this example is obvious. In order to find the τ that produces a given value on the vertical axis (for either function) we need merely reference the horizontal axis. The expressions in equations 6.7 and 4.3 are functional representations of that method.

Distribution	Density or Mass function:	Expectile Function	Inverse Expectile Function
$F(x)$	dF	$\mu_\tau := E_X(\tau)$	$\tau := E_X^{-1}(\mu_\tau)$
Generic $F(x)$	dF	$\mu_\tau = \frac{\int x\lvert\tau - I(x<\mu_\tau)\rvert dF}{\int \lvert\tau - I(x<\mu_\tau)\rvert dF}$	$\tau = \frac{\int_{-\infty}^{\mu_\tau} \lvert x-\mu_\tau\rvert dF}{\int_{-\infty}^{\infty} \lvert x-\mu_\tau\rvert dF}$
Uniform(a,b)	$\frac{dF}{dx} = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a,b) \\ 0 & \text{otherwise} \end{cases}$	$\mu_\tau = \frac{\sqrt{\tau}a + \sqrt{1-\tau}b}{\sqrt{\tau}+\sqrt{1-\tau}}$	$\tau = \frac{(b-\mu_\tau)^2}{(a-\mu_\tau)^2+(b-\mu_\tau)^2}$
Bernoulli(p)	$Y = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{otherwise} \end{cases}$	$\mu_\tau = \frac{\tau p}{\tau p + (1-\tau)(1-p)}$	$\tau = \frac{(1-p)\mu_\tau}{p(1-\mu_\tau)+(1-p)\mu_\tau}$
Two-point(a,b,p)	$Y = \begin{cases} b & \text{w.p. } p \\ a & \text{otherwise} \end{cases}$	$\mu_\tau = \frac{\tau p b + (1-\tau)(1-p)a}{\tau p + (1-\tau)(1-p)}$	$\tau = \frac{(1-p)(\mu_\tau - a)}{(1-p)(\mu_\tau-a)-p(b-\mu_\tau)}$
Normal(0,1), $\Phi(x)$	$\frac{\partial\Phi}{\partial x} = \phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$	$\mu_\tau = \frac{2\tau\phi(\mu_\tau)-1}{(1-2\tau)\Phi(\mu_\tau)+\tau}$	$\tau = \frac{\Phi(\mu_\tau)\mu_\tau + \phi(\mu_\tau)}{2\phi(\mu_\tau)-(1-2\Phi(\mu_\tau))\mu_\tau}$
Koenker's Distribution	$\frac{\partial F}{\partial x} = \frac{2\lvert x\rvert}{(4+x^2)^2\sqrt{1-4/(4+x^2)}}$	$\mu_\tau = \frac{2\tau-1}{\sqrt{\tau(1-\tau)}} = F^{-1}(\tau)$	$\tau = F(\mu_\tau)$

Table 4.1: Expectile functions and inverse expectile functions are shown for a general distribution and a handful of simple parametric distributions. Derivations are given in the appendix.

4.3 Interpretation 6: Expectiles *are* quantiles, but not of the original distribution.

Throughout this article, we have been careful to clarify that quantiles of a distribution F are always expectiles but not vice-versa. Jones [1994] points out that expectiles of a distribution F *are* quantiles in a mathematical sense, but that they are *not* the quantiles of the original distribution F. In this section, we illustrate that result. Because the inverse expectile function of any distribution is nondecreasing and maps $\bar{\mathbb{R}}$ to $[0,1]$, it is true to say that $E_X^{-1}(\mu_\tau) : \bar{\mathbb{R}} \mapsto [0,1]$ will always have the properties of a CDF. Then, the expectile function of a random variable $E_X(\tau)$ would be a quantile function for the distribution $E_X^{-1}(\mu_\tau)$. To illustrate, the CDF and inverse expectile function of a standard normal random variable are shown in Figure 4.2.

But claim made by Jones [1994] should not be stretched to abrade the differences between quantiles and expectiles. The "distribution" which the inverse expectile function E_X^{-1} appears to be is *not* the distribution F, nor will it resemble F in an obvious way. The best way to illustrate this is by counterexample: we already know that the inverse expectile function E_X^{-1} is continuous and monotone increasing. Clearly, not all cumulative distributions (nor quantile functions) have this property.

The result of monotonicity for inverse expectile functions has a dual result in the context of Jones [1994]. Because E_X^{-1} can be interpreted as a CDF, monotonicity implies that E_X^{-1} will have *density* regardless of whether F has the same! In case the dual result is not obvious from the discussion in subsection 4.1, we give an alternate proof in the appendix.

Proposition 4. *For a random variable X with distribution F, the inverse expectile function $E_X^{-1}(\mu_\tau)$: $\bar{\mathbb{R}} \mapsto [0,1]$ is a distribution with strictly positive density on $(\inf X, \sup X)$ given by*

$$f_Z(z) = \frac{\partial E_X^{-1}}{\partial z} = \frac{\tau(1-F(z)) + (1-\tau)F(z)}{\int_{\mu_\tau}^{\infty}(1-F(y))dy + \int_{-\infty}^{\mu_\tau} F(y)dy}. \tag{4.4}$$

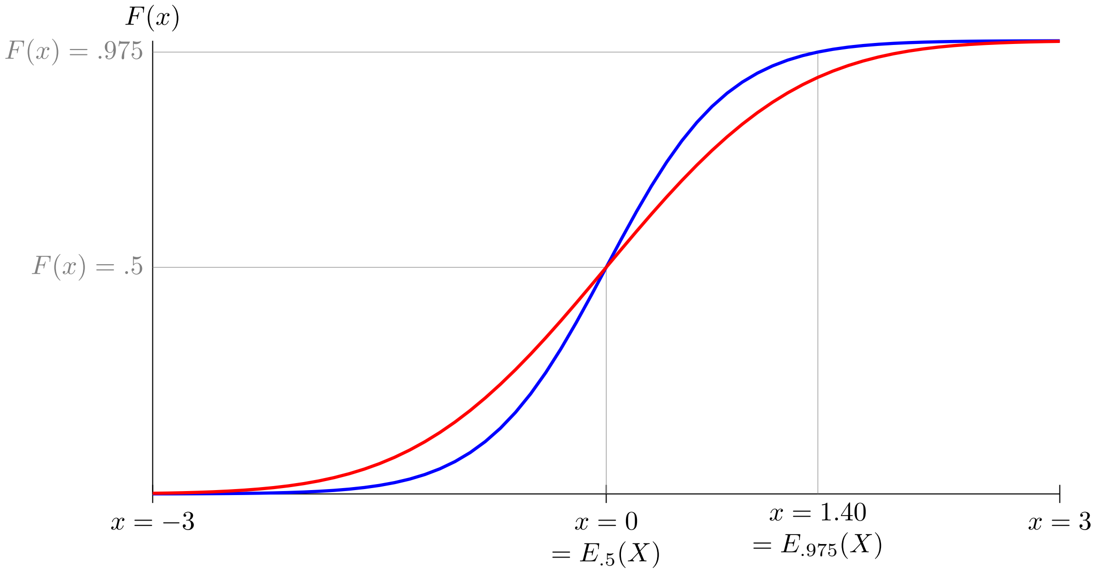

Figure 4.2: The inverse expectile function $E_X^{-1}(\mu_\tau)$ of a standard normal distribution in blue and the inverse quantile function, which is the CDF $\Phi(x)$, in red. Both have all the properties of a cumulative distribution function; the distribution E_X^{-1} has density given in equation 4.4.

These facts led to Koenker's [1993] famous exercise to determine whether any distribution's inverse expectile function—which is a CDF—could be that distribution's *own* CDF. There is one such distribution: any affine transform of a $t-$distribution with two degrees of freedom. A small sample of expectile functions and inverse expectile functions for simple parametric distributions are shown in Table 4.1. Derivations and another illustrated example can be found in the appendix.

As a simple illustration, consider opening a "hole" in the CDF of a random variable X with continuous density. Starting from the example of a standard normal variable, add a probability mass of .5 at $x = 0$;

$$X \sim \frac{1}{2}N(0,1) + \frac{1}{2}\delta$$

where δ is the Dirac function. In that case, we can see that the CDF jumps from .25 to .75 at $x = 0$. The resulting inverse expectile function is shown in figure 4.3 on the following page. Though the CDF of X is no longer continuous, the inverse expectile function remains so. In this example, the mean of X ceases to be a quantile but remains an expectile after the addition of the Dirac contamination.

We have extensively compared expectiles and quantiles. Next, we will explore the relationship between expectiles and other families of statistics.

5 Expectiles Within Larger Families

Expectiles enjoy an important position within the broader taxonomy of statistics. The most common expectile (the mean) resides near the center of that universe. Expectiles belong to the m-class and z-

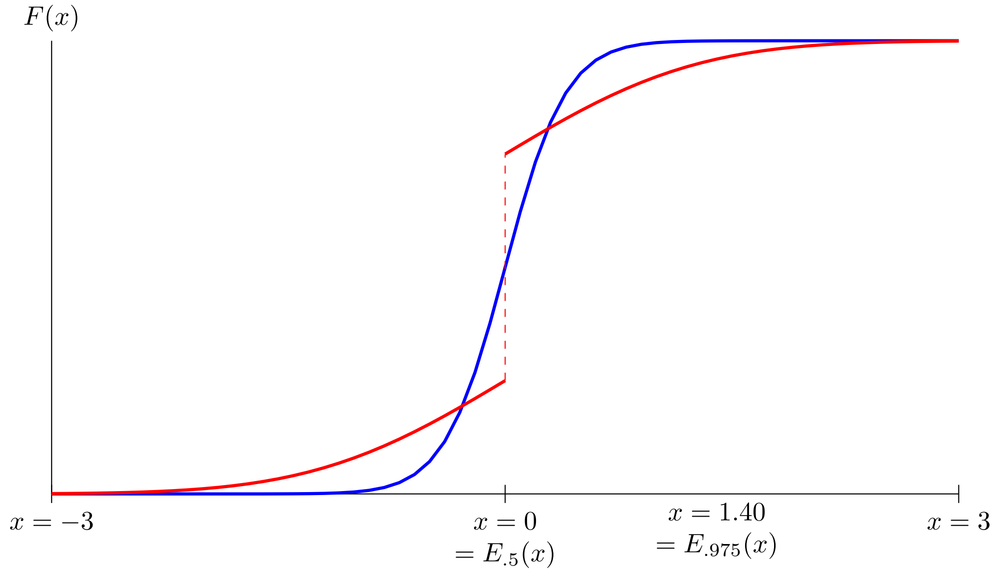

Figure 4.3: The inverse expectile function $\mu^{-1}(x)$ of $\frac{1}{2}N(0,1) + \frac{1}{2}\delta$ in blue and the inverse quantile function, which is the CDF, in red for comparison. It makes no difference whether a distribution is partially discrete: the expectile function is continuous with respect to τ.

class of statistics, but also to the less-ubiquitous m-quantiles and L_p-quantiles. These facts together mean that (1) expectiles can be estimated using very standard methods, (2) their relationship to other simple statistics can be characterized easily and formally.

5.1 Interpretation 7: Expectiles are M-class and Z-class statistics

Expectiles are novel, but their statistical properties are not. Like the mean regression, all regression expectiles are m-class statistics and their asymptotic behavior is easy to characterize using traditional methods. To support pursuits such as these, Huber et al. [1964] introduced a new framework for estimating statistics that are minimae of some objective function. For generality, let m_θ represent a parametric model (objective function) of this type, and we say that we wish to estimate some parameter θ_0 such that

$$\theta_0 = \arg\min_\theta \int m_\theta(x) dF(x). \tag{5.1}$$

Any statistic which can be defined in this way is called an m-statistic, and this class includes maximum likelihood estimators, ordinary least squares, expectiles, and many more. For a detailed discussion, see Huber [2011] or van der Vaart [2000]. The natural plugin m-estimator for each such problem is the m-statistic of the empirical distribution $\mathbb{F}_n(z) = n^{-1}\sum_{i=1}^n I(x_i \leq z)$,

$$\begin{aligned}\hat{\theta}_n &= \arg\min_\theta \int m_\theta(x) d\mathbb{F}_n(x) \\ &= \arg\min_\theta n^{-1} \sum_{i=1}^n m_\theta(x_i).\end{aligned} \tag{5.2}$$

Naturally, expectiles can be produced using $m_\theta = \varsigma_\tau(x-\theta)$, the "swoosh" function in equation 2.1. Similarly, quantiles can be defined as $m-$statistics by choosing $m_\theta(x-\theta)$ equal to Koenker and Bassett's [1978] "check" function. Both the "swoosh" function and the "check" function are convex in θ, so a solution to the minimization problem always exists. And, because the "swoosh" function is *strictly* convex, the minimization problem defines a unique expectile for every τ. This property manifests itself in the monotonicity of the expectile function $E_X(\tau)$, which was discussed in detail in Section 4. Likewise, the swoosh function is differentiable almost everywhere, so a first-order condition using the gradient exists for both location and regression models.

When the first order condition from equation 5.1 or 5.2 exists, that condition can *also* define the parameter θ_0 and its plugin estimator:

$$\begin{aligned}\int \frac{\partial}{\partial\theta} m_\theta(x) dF(x)|_{\theta=\theta_0} &= 0 \\ n^{-1}\sum_{i=1}^{n} \frac{\partial}{\partial\theta} m_\theta(x_i)|_{\theta=\hat{\theta}_n} &= 0. \end{aligned} \tag{5.3}$$

Any estimator which can be defined as the root of such an equation is considered a z-class statistic. The second equation above is called "score" function, which has a root at the estimator $\hat{\theta}_n$ just as the former has a root at the "true" parameter θ_0.

Because expectiles are both m- and z-class statistics, it should be apparent that the m-statistic representation of expectiles corresponding to 5.1 on the previous page

$$\mu_\tau(F) \coloneqq \arg\min_\theta \int |\tau - I(x<\theta)|(x-\theta)^2 dF(x)$$

is precisely the definition of μ_τ given by Newey and Powell [1987]. Similarly, the z-statistic definition of expectiles corresponding to 5.3 is the first order condition

$$\begin{aligned}(1-\tau)\int_{-\infty}^{\mu_\tau} |x-\mu_\tau| dF(x) + \tau\int_{\mu_\tau}^{\infty} |x-\mu_\tau| dF(x) &= 0 \\ \implies \tau\int_{\mu_\tau}^{\infty} |x-\mu_\tau| dF(x) &= -(1-\tau)\int_{-\infty}^{\mu_\tau} |x-\mu_\tau| dF(x)\end{aligned} \tag{5.4}$$

which we gave in 2.3, and which can be interpreted directly. While the *mean* of a distribution is sometimes described as the point of balance where the the integrals $\int_{-\infty}^{\mu_\tau} |x-\mu_\tau| dF(x)$ and $\int_{\mu_\tau}^{\infty} |x-\mu_\tau| dF(x)$ are equal, Philipps [2021b] points out that the τ^{th} expectile is an unbalanced point where the "torque" on the left and right of the fulcrum μ_τ have a ratio $\frac{\tau}{1-\tau}$.

Asymptotic behavior of m- and z-class statistics is well-studied and basic results are applicable here. For example, in cases where the data are i.i.d. with a finite first moment and therefore the empirical CDF converges to F, $\mathbb{F}_n \overset{a.s.}{\to} F$ uniformly almost surely, it is straighforward to show that the integrals $\int_{-\infty}^{\mu} |x-\mu| d\mathbb{F}_n(x)$ and $\int_{\mu}^{\infty} |x-\mu| d\mathbb{F}_n(x)$ also converge to their expected values uniformly almost surely. Thus, the sample counterpart of equation 5.4 converges to its limit, and the root of the equation converges to the true value θ_0. In other words, all sample expectiles $\hat{\mu}_{\tau,n} = \mu_\tau(\mathbb{F}_n)$ for $\tau \in (0,1)$ converge almost surely to the true value $\mu_\tau(F)$.

Standard asymptotic results for m- and z-class estimators can be used to obtain the following three propositions:

Proposition 5. *Let $X_1, ..., X_n$ be i.i.d. according to F with $E|X| < \infty$. Then*

$$\begin{aligned}\hat{\mu}_{\tau,n} &\coloneqq \arg\min_{\theta} \varsigma_\tau(x-\theta)d\mathbb{F}_n(x) \\ &\overset{a.s.}{\to} \mu(F)\end{aligned}$$

Proposition 6. *Let $X_1, ..., X_n$ be i.i.d. according to F with $E|X|^2 < \infty$. Then*

$$\sqrt{n}(\hat{\mu}_{\tau,n} - \mu_\tau) \overset{d}{\to} N\left[0, \frac{E\left(|\tau - I(X < \mu_\tau)|(X-\mu_\tau)\right)^2}{\left(\tau(1-F(\mu_\tau)) + (1-\tau)F(\mu_\tau)\right)^2}\right]$$

for every $\tau : F(\mu_\tau)$ is continuous.

Proposition 7. *Let $Y, X \sim i.i.d.P$ with $E|Y,X|^{4+\delta} < \infty$, $\delta > 0$ and $E(XX')$ nonsingular. Then*

$$\begin{aligned}\hat{\beta}_{\tau,n} = \beta_\tau(\mathbb{P}_n) &\coloneqq \arg\min_{\beta} \varsigma_\tau(y - x'\beta)d\mathbb{P}_n \\ &\overset{a.s.}{\to} \beta_\tau(P) = \beta_\tau \\ \text{and } \sqrt{n}(\hat{\beta}_{\tau,n} - \beta_\tau) &\overset{d}{\to} N(0, H^{-1}VH^{-1})\end{aligned}$$

for every $\tau : EI(Y \leq X'\beta_\tau)$ is continuous,

$$\begin{aligned}H &= E\left(|\tau - I(Y < X'\beta_\tau)|XX'\right) \\ V &= E\left(|\tau - I(Y < X'\beta_\tau)|^2 XX'(Y - X'\beta)^2\right).\end{aligned}$$

Proofs of propositions 5 and 6 can be found in Holzmann and Klar [2016]; proposition 7 follows directly from Theorem 1 in Barry et al. [2021]. We would add that the moment requirements for these results are *identical* to those required for ordinary least squares (or the sample mean) to achieve consistency and asymptotic normality. But of course they are! Ordinary least squares is merely the special case where $\tau = .5$. So, the fact that expectile regression requires no more moments than ordinary least squares should be obvious. The only "interesting" requirement in the propositions above is continuity of the CDF F evaluated at the τ^{th} expectile, which is always satisfied almost everywhere on the convex hull of the support (i.e., for almost every $\tau \in (0,1)$). In practice, this is hardly a limitation.[9]

To summarize: the asymptotic behavior of sample expectile regression coefficients is similar to any other generalized least squares or weighted least squares estimator. This is seen in proposition

[9] In the interesting special case where the distribution is discontinuous at μ_τ, Philipps [2021b] points out that the limiting distribution of $\hat{\mu}_\tau$ is that studied previously by Aigner et al. [1976], which is an asymmetric combination of two half-normal distributions. Asymptotic testing with that distribution is not challenging: one-sided z-tests can still be implemented. However; because $F(\mu_\tau)$ will be continuous for almost every $\tau \in (0,1)$, estimated expectiles with τ chosen randomly from $(0,1)$ will be asymptotically normal *with probability one*. This author is not aware of any empirical application where asymptotic normality fails, but Holzmann and Klar [2016] have constructed a simulation where this can occur.

7. Expectiles are novel, but we hope the reader will agree that they should be no more difficult to use than those GLS and WLS estimators taught in every elementary econometrics course.

5.2 Interpretation 8: Expectiles are m-quantiles

Expectiles are not the only novel m-class statistics that can be used to characterize tails of a distribution. Expectiles belong to a broad class of unusual m-statistics, called *m-quantiles*, which are characterized by simple interpretations similar to those we discussed in Section 3.

In a remarkably prescient paper, Breckling and Chambers [1988] point out that the τ^{th} quantile can be estimated as the *median* of a re-weighted empirical distribution. For a sample distribution $\mathbb{F}_n(x)$, we have the τ^{th} sample quantile $\hat{\xi}_\tau$ obtained as

$$\hat{\xi}_\tau = \arg\min_{\xi} \int |\tau - I(x < \xi)||x - \xi| d\mathbb{F}_n(x),$$

but we could equally define a probability distribution $\mathbb{G}_n(x)$ as:

$$d\mathbb{G}_n(x) \propto \begin{cases} \tau d\mathbb{F}_n(x) & \text{if } x \geq \hat{\xi}_\tau \\ (1-\tau) d\mathbb{F}_n(x) & \text{otherwise.} \end{cases}$$

This transformed version of $\mathbb{F}_n$ is simple enough: clearly $\hat{\xi}_\tau$ is the estimated median of $\mathbb{G}_n$! Breckling and Chambers [1988] made this point clear: the τ^{th} quantile can be interpreted as the value that *would be* the median if values above it were $\frac{\tau}{1-\tau}$ times as likely to occur as they actually are:

$$\hat{\xi}_\tau = \arg\min_{\xi} \int |x - \xi| d\mathbb{G}_n(x).$$

We would emphatically underscore that the same remark should be made regarding expectiles:

$$\hat{\mu}_\tau = \arg\min_{\xi} \int (x - \xi)^2 d\mathbb{G}_n(x) \tag{5.5}$$
$$\text{where } d\mathbb{G}_n(x) \propto \begin{cases} \tau d\mathbb{F}_n(x) & \text{if } x \geq \hat{\mu}_\tau \\ (1-\tau) d\mathbb{F}_n(x) & \text{otherwise} \end{cases}$$

and the τ^{th} expectile of F is the value that would be the mean if values above it were to occur $\tau/(1-\tau)$ times as often as they do according to F.

Indeed, Breckling and Chambers [1988] apply the same logic to all m-class statistics. Those authors suggest adding asymmetric weights τ and $1-\tau$ to any $m-$statistic:

$$\theta_0 = \arg\min_{\theta} \int |\tau - I(x < \theta)| \rho(x - \theta) dF(x)$$
$$\Longrightarrow 0 = \int |\tau - I(x < \theta)| \psi(x - \theta) dF(x). \tag{5.6}$$

This nests the usual statistic when $\tau = .5$ and extends it to non-central cases with any ratio

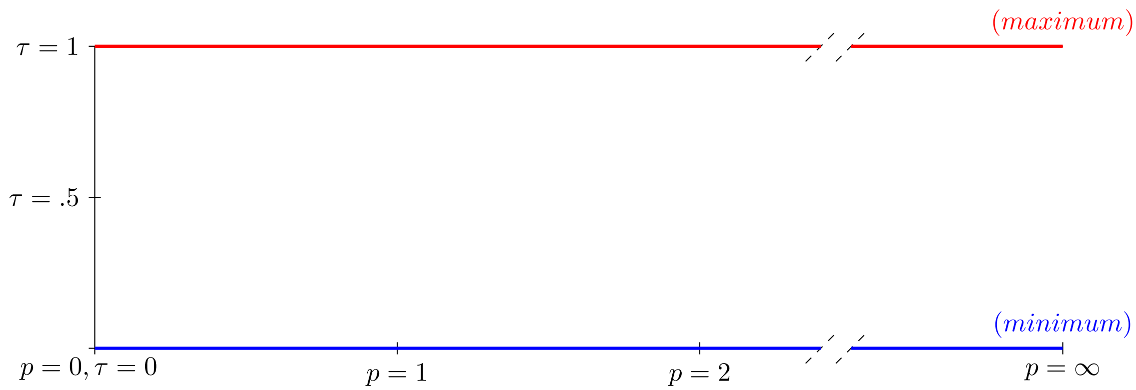

Figure 5.1: The minimum and maximum of a function can be obtained by minimizing the expected asymmetric L_p loss function for $p > 0$ if it exists. The minimum is obtained in the limit as $\tau \to 0$ and the maximum is obtained in the limit as $\tau \to 1$. For distributions where the minimum and maximum do not exist, these will be the infimum and supremum, respectively.

of weights. For $\rho(u) = u^2$, this notation nests expectiles. For $\rho(u) = |u|$, it nests the standard quantiles. Other interesting ρ appear in the next subsection.

Every m-quantile is the minimizer of ρ under an asymmetrically weighted distribution as in $\mathbb{G}_n$, which makes simple interpretations available to *any* set of m-quantiles defined by an objective function ρ. That is, the m-quantile using ρ_τ is the obvious estimator of the minimizer of ρ for an atypical observation if values above it were to occur $\frac{\tau}{1-\tau}$ times as often as they actually do. But only the least-squares criterion $\rho(u) = u^2$ elicits the mean of all distributions [see Gneiting, 2011], so the *mean* of a reweighted distribution $\mathbb{G}_n$ can only be an expectile and expectiles are the only m-quantiles with unambiguous interpretations as the conditional mean of a distribution.

It must be noted that m-quantiles have been studied heavily in environments where traditional quantile regression fails. See Kordas [2006], Dawber and Chambers [2019] for applications to binary response data or Serfling [2002, 2004], Breckling and Chambers [1988], Daouia and Paindaveine [2019] for applications to multiple equation models where quantiles are poorly defined. Invariably, these regression methods have fallen by the wayside because they are *not* true quantiles and, in many cases, minimizers of objective functions ρ can be unintuitive. Expectiles are the glaring exception to that rule: they are well-defined in binary response and multiple equation models, and they are simple and directly interpretable. In the next section, we will compare them to one other family of m-quantiles that can be easy to interpret.

5.3 Interpretation 9: Expectiles are L_p-quantiles

Expectiles are L_p-quantiles–a simple parametric subset of the general m-quantiles. This family should be familiar to all readers, though its *name* may be unfamiliar. After the introduction of m-quantiles, Chen [1996] made a major conceptual refinement. Rather than minimizing an arbitrary

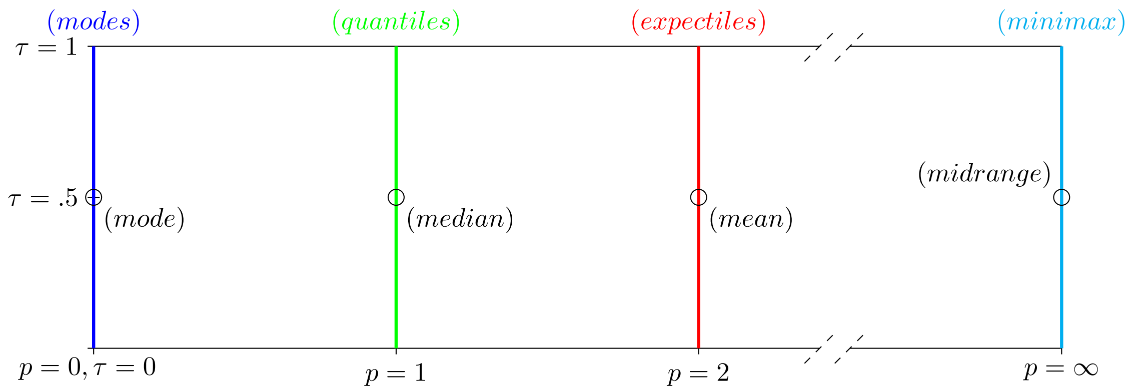

Figure 5.2: Each of four measures of central tendency—the mode, median, mean, and midrange—are nested as L_p-quantiles in the central cases where $\tau = .5$ and $p = 0, 1, 2, \infty$ respectively. Asymmetric versions of these statistics include quantiles and expectiles when $p = 1, 2$ respectively. Asymmetric modes and minimax location parameters are relatively unexplored in the literature.

$\rho(\cdot)$, Chen suggests obtaining the statistics that minimize L_p loss functions, namely

$$L_p(x, \theta) = \|x - \theta\|_p^p = |x - \theta|^p$$

for $p \geq 0$. In this case, the asymmetric minimizers

$$\theta_\tau = \arg\min_\theta \int |\tau - I(x < \theta)| L_p(x, \theta) dF(x) \tag{5.7}$$

are special m-quantiles where $\rho(x - \theta) = L_p(x, \theta)$. The L_p loss function and its weighted counterpart will be strictly convex for $p > 1$. As a result, there is a unique, smooth (in τ) minimizer when $p > 1$.[10]

The wonderful property of these L_p-quantiles is way they nest "standard" statistics as minimizers. This is true for the mean, median, mode, midrange, minimum, and maximum. Essentially, the class of L_p-quantiles parametrizes every elementary statistic taught to young students, together with a variety of others commonly in use. Or, at the very least, these can be attained *in the limit* as either τ or p approach a given value. It is straightforward to see that the minimum (infimum) and maximum (supremum) are nested as τ approaches 0 and 1, regardless of p. This is shown in Figure 5.1 on the previous page. For $p = 1$, the function is the check function of Koenker and Bassett [1978]. For $p = 2$, L_p-quantiles are expectiles. In cases where $\tau = .5$, we also see the standard mode in the limit as $p \to 0$, the median at $p = 1$, the mean at $p = 2$, and the midrange in the limit as $p \to \infty$.

In principle, each of these measures of central tendency—mean, median, mode, midrange—can be extended as suggested by Breckling and Chambers [1988] to create new m-quantile classes of asymmetric estimators. These are shown in Figure 5.2. Extending the median to the non-central case $\tau \neq .5$ produces quantiles. Extending the mean in the same way produces expectiles. Asymmetric versions of the mode and midrange are not well-studied. But, in principle, an asymmetrically

[10]For $p \leq 1$, the L_p loss function is not strictly convex and equation 5.7 may have multiple minima for $\tau \in (0, 1)$.

weighted mode regression could be used to search for different modes in a multimodal distribution. On the other end of the spectrum, the supremum norm is a special p-norm attained as $p \to \infty$. That norm, and the L_p loss function with $p = \infty$, have the minimax property (θ_τ will minimize the maximum possible error), which is easily extensible to an environment where positive and negative errors are given different importance.

Within the L_p-quantile class, only two families are commonly used for regression analysis. These are the L_1-class (quantiles) and the L_2-class (expectiles). As mentioned elsewhere, the L_2-class with $\tau = .5$ (least squares) has historically been the dominant force in econometrics. Extending that class to asymmetric L_p-loss functions helps to root these more novel statistics in a universe of geometric and mathematical projection spaces. The mapping from quantiles to expectiles discussed in Section 4 can also extend to other L_p-quantiles or m-quantiles, making it possible to estimate expectile (or quantile) regression lines using other m-class loss functions; see especially Daouia et al. [2019b]. And much research is yet ongoing. Because so many elementary statistics are nested within the two-parameter family, Chen's L_p-quantiles are arguably the closest thing there is to a Grand Unified Theory of Statistics–with quantiles and expectiles at its center.

6 Conclusions

Expectiles are a fascinating class of statistics. In this article, we have stylized nine distinct interpretations for expectiles. Most of these are simple: expectiles are easy to interpret on their own. It is also interesting and fruitful–so long as we are cautious–to compare expectiles with quantiles and other families. Aside from the true quantiles, expectiles are by far the best-known example from "generalized" quantile families including m-quantiles and L_p-quantiles. And, because mean regression is a special case of expectile regression, it is also accurate to say that expectile regression is the dominant regression methodology in applied sciences—though it is *not* usually identified by that name.

Though the literature has focused heavily on expectiles' relationships with quantiles, this is probably the *least* obvious or useful way to look at them. Instead, we would encourage practitioners to note expectiles' relationships with other m-class statistics and L_p-quantiles. Even more so, applied researchers should enjoy the benefits of expectiles' natural interpretation as the mean of a conditional distribution. Expectiles are easy to interpret and span the entire distribution of a random variable, so they have explanatory power similar to quantiles. And in environments where true quantiles are poorly defined, such as binary response problems, simultaneous equation models, and others outlined by Koenker [2005, ch. 8], expectile regression may be the simplest and most prudent choice–if not the *only* choice. Expectiles are m-class statistics with very standard analytical properties, so there is no substantial cost to using them.

While every quantile of a distribution F is an expectile, this result is vacuously true: *every location parameter* is an expectile. Expectile functions of a distribution F are invertible bijective mappings from the unit interval to the convex hull of the support of F. Accordingly, any value between the infimum and supremum of F is an expectile and it is trivial to solve *which* expectile corresponds to that value for any F. Importantly, these facts are globally true for distributions with a first moment and require no special assumptions about whether F has density, for example.

The τ^{th} expectile of a distribution F is the value that *would be* F's mean if observations above it were $\frac{\tau}{1-\tau}$ times as likely to occur as they are according to F. This fact is immediately useful to analyze individual observations that are different from average. It is also useful to analyze scenarios where missing data occurs at different frequencies depending on the sign of the error term. The same interpretation applies in regression models, and extends the power of the mean regression to exploring the tails of a distribution. But their power remains largely unadvertised: expectile regression may be the best-kept secret in econometrics.

References

Dennis J Aigner, Takeshi Amemiya, and Dale J Poirier. On the estimation of production frontiers: maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review*, pages 377–396, 1976.

Takeshi Amemiya. Regression analysis when the dependent variable is truncated normal. *Econometrica: Journal of the Econometric Society*, pages 997–1016, 1973.

Takeshi Amemiya. Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica: Journal of the Econometric Society*, pages 999–1012, 1974.

Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

Jushan Bai. Least absolute deviation estimation of a shift. *Econometric Theory*, 11(3):403–436, 1995.

Donald R Barr and E Todd Sherrill. Mean and variance of truncated normal distributions. *The American Statistician*, 53(4):357–361, 1999.

Amadou Barry, Karim Oualkacha, and Arthur Charpentier. A new gee method to account for heteroscedasticity using asymmetric least-square regressions. *Journal of Applied Statistics*, pages 1–27, 2021.

Fabio Bellini and Elena Di Bernardino. Risk management with expectiles. *The European Journal of Finance*, 23(6):487–506, 2017.

Jens Breckling and Ray Chambers. M-quantiles. *Biometrika*, 75(4):761–771, 1988.

Zehua Chen. Conditional lp-quantiles and their application to the testing of symmetry in non-parametric regression. *Statistics & probability letters*, 29(2):107–115, 1996.

Victor Chernozhukov, Iván Fernández-Val, and Ye Luo. The sorted effects method: discovering heterogeneous effects beyond their averages. *Econometrica*, 86(6):1911–1938, 2018.

Abdelaati Daouia and Davy Paindaveine. From halfspace m-depth to multiple-output expectile regression. *arXiv preprint arXiv:1905.12718*, 2019.

Abdelaati Daouia, Irène Gijbels, and Gilles Stupfler. Extremiles: A new perspective on asymmetric least squares. *Journal of the American Statistical Association*, 114(527):1366–1381, 2019a.

Abdelaati Daouia, Stéphane Girard, Gilles Stupfler, et al. Extreme m-quantiles as risk measures: From l^1 to l^p optimization. *Bernoulli*, 25(1):264–309, 2019b.

Abdelaati Daouia, Stephane Girard, and Gilles Stupfler. Tail expectile process and risk assessment. *Bernoulli*, 26(1):531–556, 2020.

James Dawber and Raymond Chambers. Modelling group heterogeneity for small area estimation using m-quantiles. *International Statistical Review*, 87:S50–S63, 2019.

Graham Elliott, Allan Timmermann, and Ivana Komunjer. Estimation and testing of forecast rationality under flexible loss. *The Review of Economic Studies*, 72(4):1107–1125, 2005.

Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

James J Heckman. Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of political Economy*, 109(4):673–748, 2001.

Hajo Holzmann and Bernhard Klar. Expectile asymptotics. *Electronic Journal of Statistics*, 10(2): 2355–2371, 2016. doi: 10.1214/16-EJS1173.

Peter J Huber. *Robust statistics*, volume 523. John Wiley & Sons, 2004.

Peter J Huber et al. Robust estimation of a location parameter. *The annals of mathematical statistics*, 35(1):73–101, 1964.

PJ Huber. Robust statistics (pp. 1248-1251). *Springer Berlin Heidelberg. Rehabilitation Psychology*, 46:382–399, 2011.

M Chris Jones. Expectiles and m-quantiles are quantiles. *Statistics & Probability Letters*, 20(2): 149–153, 1994.

David G Kleinbaum and Mitchel Klein. *Survival analysis*. Springer, 2010.

Thomas Kneib. Beyond mean regression. *Statistical Modelling*, 13(4):275–303, 2013.

Roger Koenker. When are expectiles percentiles? *Econometric Theory*, 9(3):526–527, 1993.

Roger Koenker. *Quantile Regression (Econometric Society monographs; no. 38)*. Cambridge University Press, 2005.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

Gregory Kordas. Smoothed binary regression quantiles. *Journal of Applied Econometrics*, 21(3): 387–407, 2006.

Whitney K Newey and James L Powell. Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, pages 819–847, 1987.

Collin S Philipps. When is an expectile the best linear unbiased estimator? Available at SSRN: http://ssrn.com/abstract=3881181, 2021a.

Collin S Philipps. The mle of aigner, amemiya, and poirier is not the expectile mle. *Econometric Reviews*, 2021b. doi: 10.2139/ssrn.3660526.

Robert Serfling. Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica*, 56(2):214–232, 2002.

Robert Serfling. Nonparametric multivariate descriptive measures based on spatial quantiles. *Journal of statistical Planning and Inference*, 123(2):259–278, 2004.

James W Taylor. Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. *Journal of Business & Economic Statistics*, 37(1): 121–133, 2019.

James Tobin. Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society*, pages 24–36, 1958.

A.W. van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Linda Schulze Waltrup, Fabian Sobotka, Thomas Kneib, and Göran Kauermann. Expectile and quantile regression–david and goliath? *Statistical Modelling*, 15(5):433–456, 2015.

Qiwli Yao and Howell Tong. Asymmetric least squares regression estimation: a nonparametric approach. *Journal of Nonparametric Statistics*, 6(2-3):273–292, 1996.

Johanna F Ziegel. Coherence and elicitability. *Mathematical Finance*, 26(4):901–918, 2016.

Hui Zou. Generalizing koenker's distribution. *Journal of Statistical Planning and Inference*, 148: 123–127, 2014.

Appendix to "Interpreting Expectiles"

Remarks regarding notation

Parts of this article make extensive use of the notation for expectile functions of a random variable X, $E_X(\tau) : (0,1) \mapsto \bar{\mathbb{R}}$, which express μ_τ as a function of τ, and the inverse expectile function $E_X^{-1}(\mu_\tau) : \bar{\mathbb{R}} \mapsto (0,1)$ which does the reverse. Expectile and inverse-expectile functions can be compared extensively to quantile functions $Q_X(\tau) : (0,1) \mapsto \bar{\mathbb{R}}$ and inverse-quantile functions $Q_X^{-1}(\xi_\tau) : \bar{\mathbb{R}} \mapsto (0,1)$, the latter of which is equivalent to the CDF of X in cases where a unique inverse exists. In our notation, $\bar{\mathbb{R}}$ is the set of extended real numbers $\mathbb{R} \cup \{-\infty, \infty\}$.

For cases where a unique inverse of F does not exist, the quantile function is defined using a generalized inverse. By convention, we define the τ^{th} quantile to be the *infimum*

$$Q_X(\tau) \coloneqq \inf\{x : \ F(x) \geq \tau\}$$

of the set of values where $F(x)$ is at least τ. For a detailed introduction to quantiles, see Koenker [2005].

The 0^{th} and 1^{th} quantile and expectile, defined as minimizers, are not unique. However, we will rely on the fact that the "ends" of the quantile and expectile functions for a random variable are uniquely that variable's infimum and supremum;

$$\begin{aligned}
\lim_{\tau \to 1^-} Q_X(\tau) &= \lim_{\tau \to 1^-} E_X(\tau) = \sup X \\
\lim_{\tau \to 0^+} Q_X(\tau) &= \lim_{\tau \to 0^+} E_X(\tau) = \inf X.
\end{aligned}$$

And, in cases where the minimum or maximum of X exist, these values are often considered to be the 0^{th} and 1^{th} expectile, respectively. Our list of interpretations begins below.

Supplement to Section 3.2

The following examples obtain the mean of a latent distribution by weighting an observed distribution.

Example 8. Suppose that a particular observation X_i will be distributed according to $F(x|\Omega_i)$ and we know only that X_i is greater than 5, it is clear that $F(x|X > 5)$ can be written as a function of the unconditional CDF $F(x)$:

$$\begin{aligned}
F(x|X > 5) &= \int_{-\infty}^{x} dF(x|X > 5) \\
&= \int_{-\infty}^{x} w(x) dF(x) \qquad (6.1)
\end{aligned}$$

where these weights $w(x)$ are zero for $x \leq 5$ and $[\int_5^\infty dF(x)]^{-1}$ otherwise. In that case, standard statistics such as the conditional mean can be estimated using the same transformation of the original distribution F:

$$
\begin{aligned}
E(X_i|X_i > 5) &= \int_{-\infty}^{\infty} x dF(x|X > 5) \\
&= [\int_{5}^{\infty} dF(x)]^{-1} \int_{5}^{\infty} x dF(x). \qquad (6.2)
\end{aligned}
$$

This sort of transformation is common censored data applications and survival analysis; see Kleinbaum and Klein [2010] and the similar example below.[11]

Example 9 (Truncated Normal)**.** A common example in censored regression models: Let $X^* \sim N(\mu, \sigma^2)$, but assume $X = X^*$ is only observed if $X^* \geq 0$. In that case, the probability density of X^* is given by

$$
f_{X^*}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} = \phi(x; \mu, \sigma^2)
$$

But the observed data X are distributed according to the conditional distribution of $X^*|X^* \geq 0$; we have

$$
\begin{aligned}
f_{X^*|X^* \geq 0}(x) &= \frac{f_{X^*}(x)}{\Pr(X^* \geq 0)} \\
&= \frac{\phi(x, \mu, \sigma^2)}{1 - \Phi(0; \mu, \sigma^2)}
\end{aligned}
$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the Gaussian probability density and cumulative distribution functions, respectively. Accordingly, the mean of X^* is μ but the mean of the observed X is given by

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x \frac{\phi(x, \mu, \sigma^2)}{1 - \Phi(0; \mu, \sigma^2)} dx \\
&= \frac{1}{1 - \Phi(0; \mu, \sigma^2)} \int_{-\infty}^{\infty} x\phi(x, \mu, \sigma^2) dx \\
&= \mu + \frac{\phi(0; \mu, \sigma^2)}{1 - \Phi(0; \mu, \sigma^2)}.
\end{aligned}
$$

For a thorough analysis of truncated normal distributions, see Barr and Sherrill [1999]. In this example, every value $X^* < 0$ goes missing from the sample. If, instead, the probability that the observation $X^* < 0$ goes missing is $\alpha \in [0, 1]$, it is straightforward to find that

$$
E(X) = \mu + \alpha \frac{\phi(0; \mu, \sigma^2)}{1 - \Phi(0; \mu, \sigma^2)}.
$$

Example 10 (The τ^{th} Expectile)**.** Suppose we know an unconditional distribution F, and we know

[11] The representation of $F(x|X > 5)$ in equations 6.1 and 6.2 follow directly from the definition of conditional probability

$$
\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}
$$

and extend to joint distributions of variables defined on the same probability space. For clarity, we consider only the univariate case.

that the observation X_i has some conditional mean μ_i not known *a priori*, but we know that the probability of $X_i > \mu_i$ is α. We allow α to be more or less than the probability of any other $X_j > \mu_j$ without loss of generality.

Is the conditional mean μ_i well defined, given only this information? Yes. Incorporating this information in the same way as the previous example, specifically $\Pr(X_i \leq \mu_i) \neq F(\mu_i)$, we have

$$\begin{aligned}
\mu_i &= E(X_i | \Pr(X_i > \mu_i) = \alpha) \\
&= (1-\alpha)E(X_i|X_i \leq \mu_i) + \alpha E(X_i|X_i > \mu_i) \\
&= (1-\alpha)[\int_{-\infty}^{\mu_i} dF(x)]^{-1} \int_{-\infty}^{\mu_i} xdF(x) + \alpha[\int_{\mu_i}^{\infty} dF(x)]^{-1} \int_{\mu_i}^{\infty} xdF(x)
\end{aligned}$$

For any α, it is straightforward to find $\tau \in (0,1)$ such that the probability $\alpha = \Pr(X_i > \mu_i)$ is proportional to τ times the unconditional probability of the same. The conditional probability of $X_i \leq \mu_i$ will be proportional to $1-\tau$ times the *un*conditional probability of the same. So:

$$\begin{aligned}
\mu_i &= E\left(X_i | \Pr(X_i > \mu_i) = \tau\kappa(1-F(\mu_i))\right) \\
&= (1-\tau)\kappa F(\mu_i)E(X_i|X_i \leq \mu_i) + \tau\kappa(1-F(\mu_i))E(X_i|X_i > \mu_i) \\
&= (1-\tau)\kappa \int_{-\infty}^{\mu_i} xdF(x) + \tau\kappa \int_{\mu_i}^{\infty} xdF(x) \\
&= \mu_\tau
\end{aligned}$$

the mean is the τ^{th} expectile! In summary, the τ^{th} expectile is the *conditional* mean given the information that an observation's relative odds of exceeding the conditional mean are $\frac{\tau}{1-\tau}$.

$$\frac{\Pr(X_i > \mu_i)/(1-F(\mu_i))}{\Pr(X_i \leq \mu_i)/F(\mu_i)} = \frac{\tau}{1-\tau} \tag{6.3}$$

Example 11. Suppose we know only a conditional τ^{th} quantile, $Q_\tau(X_i) = \xi_{\tau,i}$, of the distribution of X_i. In that case, we may also determine the conditional distribution of X_i. Similarly, we may calculate the conditional expected value of X_i as:

$$\begin{aligned}
E(X_i|\xi_{\tau,i}) &= E\left[X_i | \Pr(X_i \leq \xi_{\tau,i}) = \tau\right] \\
&= \tau E(X_i|X_i \leq \xi_{\tau,i}) + (1-\tau)E(X_i|X_i > \xi_{\tau,i}) \\
&= \tau[\int_{-\infty}^{\xi_{\tau,i}} dF(x)]^{-1} \int_{-\infty}^{\xi_{\tau,i}} xdF(x) + (1-\tau)[\int_{\xi_{\tau,i}}^{\infty} dF(x)]^{-1} \int_{\xi_{\tau,i}}^{\infty} xdF(x).
\end{aligned}$$

The conditional means in the two examples above are easy to calculate and even easier to interpret. *Expectiles* of a distribution F are identical, in principle, to these two examples—albeit perhaps even simpler. The τ^{th} expectile, μ_τ, is the conditional mean of X_i if values above μ_τ occur $\frac{\tau}{1-\tau}$ times as often as they usually do.

Proofs of Propositions

Proof of proposition 2: The τ^{th} expectile is characterized by

$$\mu_\tau := \arg\min_\mu \int \varsigma_\tau(x-\mu)dF(x)$$

and its first order condition, which leads to the ratio of partial moments

$$\frac{\tau}{1-\tau}\int_{\mu_\tau}^{\infty}|x-\mu_\tau|dF(x) = \int_{-\infty}^{\mu_\tau}|x-\mu_\tau|dF(x). \tag{6.4}$$

The representation above is sufficient to prove that the expectile function is *surjective.* That is, every element μ_τ in the codomain $(\inf X, \sup X)$ has at least one $\tau \in (0,1)$ such that $E_X(\tau) = \tau$. To prove this, simply point out that every element in the codomain $(\inf X, \sup X)$ has at least one element above and one element below it. Plainly, there exists some $\delta > 0$ such that

$$\begin{aligned} \mu_\tau + \delta &\in (\inf X, \sup X) \\ \mu_\tau - \delta &\in (\inf X, \sup X). \end{aligned}$$

Because $\mu_\tau \pm \delta \in (\inf X, \sup X)$, we know that

$$\begin{aligned} \int_{\mu_\tau+\delta}^{\sup X} dF(x) > 0 \\ \int_{\inf X}^{\mu_\tau-\delta} dF(x) > 0 \end{aligned}$$

which implies

$$\begin{aligned} \int_{\mu_\tau}^{\infty}|x-\mu_\tau|dF(x) &\geq \int_{\mu_\tau+\delta}^{\sup X}|x-\mu_\tau|dF(x) > 0 \\ \int_{-\infty}^{\mu_\tau}|x-\mu_\tau|dF(x) &\geq \int_{\inf X}^{\mu_\tau-\delta}|x-\mu_\tau|dF(x) > 0. \end{aligned}$$

In other words, *both* integrals in equation 6.4 must be strictly positive. However,

$$\begin{aligned} \int_{\mu_\tau}^{\infty}|x-\mu_\tau|dF(x) &\leq \int_{-\infty}^{\infty}|x-\mu_\tau|dF(x) < \infty \\ \int_{-\infty}^{\mu_\tau}|x-\mu_\tau|dF(x) &\leq \int_{-\infty}^{\infty}|x-\mu_\tau|dF(x) < \infty \end{aligned}$$

with finiteness assured by the existence of a finite moment. But then

$$\underbrace{\frac{\tau}{1-\tau}}_{a(\tau)} = \underbrace{\frac{\int_{-\infty}^{\mu_\tau}|x-\mu_\tau|dF(x)}{\int_{\mu_\tau}^{\infty}|x-\mu_\tau|dF(x)}}_{b(\mu_\tau)} > 0$$

implies that τ is strictly positive, while inverting both ratios above will imply that $1 - \tau$ is strictly positive. So $\tau \in (0, 1)$. Finally, observe that the ratio of integrals above is continuous and monotonic in μ_τ. The numerator is continuous and increasing in μ_τ, while the denominator is continuous and decreasing in μ_τ (while strictly positive). And, because the function $\frac{\tau}{1-\tau} : (0, 1) \mapsto (0, \infty)$ is continuous and strictly increasing some value $\tau \in (0, 1)$ *must* solve the equation above. Both the function $a(\tau)$ and the function $b(\mu_\tau)$ are continuous and monotone, and compositions of continous monotone functions on open sets are also continuous and monotone. Accordingly, we have

$$
\begin{aligned}
E_X(\tau) &= b^{-1}(a(\tau)) \\
E_X^{-1}(\mu_\tau) &= a^{-1}(b(\tau))
\end{aligned}
$$

As compositions of continuous monotone functions, both of these must be continuous and monotone. It is obvious that $E_X^{-1}(\mu_\tau)$ is the inverse of $E_X(\tau)$ and vice-versa, so trivially both of these are invertible i.e. their inverses exist. $\square$

Alternate Proof of Continuity of the Expectile Function $E_X(\tau)$ and its inverse: As an alternative, to illustrate continuous monotonicity of the expectile function we may rely on an un-obvious representation of the τ^{th} expectile;

$$
\mu_\tau - E(X) = \left(\frac{(2\tau - 1)}{(1 - \tau)} \int_{\mu_\tau}^{\infty} (x - \mu_\tau) dF(x) \right) \tag{6.5}
$$

which was given by Newey and Powell [1987, eq. 2.7].

Informally, it is possible to see this result from equation 6.5. The integral on the right hand side is always positive, and the fraction $\frac{(2\tau-1)}{(1-\tau)}$ is negative only when $\tau < .5$. As you see, μ_τ will be greater than the mean when $\tau > .5$ and less than the mean when $\tau < .5$. Although the integral on the right *decreases* with μ_τ, it does so slowly as μ_τ becomes large. On the other hand, the fraction $\frac{(2\tau-1)}{(1-\tau)}$ diverges rapidly as $\tau \to 1$. Behavior in the other direction can be determined by considering $-X$ rather than X or by noting that the integral on the right diverges as μ_τ approaches $-\infty$. $\square$

Corollary 12 (Corollary to Proposition 3.)**.**

Yao and Tong [1996] were the first to point out that the quantiles of a distribution F can be expressed as expectiles (and sometimes vice-versa). Equivalence of a particular quantile and expectile *regression line* can occur under special circumstances, such as in the famous example where the mean and median of a symmetric distribution are the same.

We begin with the location-scale regression model,

$$
Y = \mu(X) + \sigma(X)\epsilon \tag{6.6}
$$

and its conditional quantile function:

$$
Q_Y(\alpha | X = x) = \mu(x) + \sigma(x) Q_\epsilon(\alpha).
$$

The result given by Yao and Tong [1996] indicates that the τ^{th} expectile regression line (curve) is equal to the α^{th} quantile regression line (curve), and that we can solve for the function $\tau = h(\alpha)$ as an explicit function of F, α:

$$h(\alpha) = \frac{\alpha F^{-1}(\alpha) - G(F^{-1}(\alpha))}{\mu_{.5}(F) - 2G(F^{-1}(\alpha)) - (1-2\alpha)F^{-1}(\alpha)}. \tag{6.7}$$

Proof. To simplify notation, $G(z) = \int_{-\infty}^{z} u dF(u)$. We can solve for the functional representation of $h(\alpha)$ by pointing out or assuming that $E_{h(\alpha)}(Y) = Q_Y(\alpha)$. Then we use equation 6.13 to write

$$\begin{aligned}
h(\alpha) &= \frac{E\left(|Y - Q_Y(\alpha)| I(Y \leq Q_Y(\alpha)) | X\right)}{E\left(|Y - Q_Y(\alpha)| | X\right)} \\
&= \frac{E\left(|\epsilon - Q_\epsilon(\alpha)| I(\epsilon \leq Q_\epsilon(\alpha)) | X\right)}{E\left(|\epsilon - Q_\epsilon(\alpha)| | X\right)} \\
&= \frac{E\left(\left(Q_\epsilon(\alpha) - \epsilon\right) I(\epsilon \leq Q_\epsilon(\alpha))\right)}{E\left(\left(Q_\epsilon(\alpha) - \epsilon\right) I(\epsilon \leq Q_\epsilon(\alpha)) + \left(\epsilon - Q_\epsilon(\alpha)\right) I(\epsilon > Q_\epsilon(\alpha))\right)} \\
&= \frac{\tau Q_\epsilon(\alpha) - G(Q_\epsilon(\alpha))}{\tau Q_\epsilon(\alpha) - G(Q_\epsilon(\alpha)) + E(\epsilon) - G(Q_\epsilon(\alpha)) - (1-\alpha)Q_\epsilon(\alpha)} \\
&= \frac{\tau Q_\epsilon(\alpha) - G(Q_\epsilon(\alpha))}{E(\epsilon) - 2G(Q_\epsilon(\alpha)) - (1-2\alpha)Q_\epsilon(\alpha)} \\
&\frac{\alpha F^{-1}(\alpha) - G(F^{-1}(\alpha))}{\mu_{.5}(F) - 2G(F^{-1}(\alpha)) - (1-2\alpha)F^{-1}(\alpha)}
\end{aligned}$$

which is the same as equation 6.7. Then the α^{th} quantile is the $\tau = h(\alpha)^{th}$ expectile with the function h defined as above. □

Proof of proposition 4. The proof proceeds by proving that both E_X and E_X^{-1} are differentiable at continuity points of F and their derivatives are unambiguously positive. Differentiability of both the function and its inverse rules out "jump" discontinuities; then E_X^{-1} is continuous. Then E_X^{-1} is a continuous monotone CDF, the derivative $\partial E_X^{-1}(z)/\partial z$ can be interpreted directly as its density.

From the definition in equation 2.1, we can solve for the τ^{th} expectile of X in implicit functional form.

$$\mu_\tau = \frac{\int x|\tau - I(x < \mu_\tau)|dF}{\int |\tau - I(x < \mu_\tau)|dF} \tag{6.8}$$

$$= \frac{(1-\tau)G(\mu_\tau) + \tau(\mu_{.5} - G(\mu_\tau))}{(1-\tau)F(\mu_\tau) + \tau(1 - F(\mu_\tau))} \tag{6.9}$$

with the left partial moment $G(z) = \int_{-\infty}^{z} u dF(u)$. From the representation above, it is tedious but straightforward to verify that μ_τ is left and right differentiable everywhere;

$$\frac{\partial^\pm \mu_\tau}{\partial \tau} = \frac{\int_{\mu_\tau}^{\infty}(1 - F(x))dx + \int_{-\infty}^{\mu_\tau} F(x)dx}{(1-\tau)F^\pm(\mu_\tau) + \tau(1 - F^\pm(\mu_\tau))} \tag{6.10}$$

where F^+ denotes the right limit of F and F^- its left limit. To obtain this result, begin with the proof of proposition 1 in Holzmann and Klar [2016], who assume F is continuous at μ_τ, then add that F has left and right limits at any point where it is *not* continuous. The result follows.

Next, obtain the derivative $\partial E_X^{-1}(z)/\partial z$. From the generic inverse expectile function for a distribution F,

$$E_X^{-1}(z) = \frac{\int_{-\infty}^{z} |x-z| dF(x)}{\int_{-\infty}^{\infty} |x-z| dF(x)}$$

and apply the quotient rule together with Leibniz' rule for integrals[12];

$$E_X^{-1}(z) = \frac{\int_{-\infty}^{z} (z-x) dF(x)}{\int_{-\infty}^{z} (z-x) dF(x) + \int_{z}^{\infty} (x-z) dF(x)}$$

so:

$$\begin{aligned}
\frac{\partial E_X^{-1}(z)}{\partial z} &= \frac{\left(\int_{-\infty}^{\infty} |x-z| dF(x)\right) F(z) - \int_{-\infty}^{z} (z-x) dF(x) \,(2F(z)-1)}{\int_{-\infty}^{\infty} |x-z| dF(x) \left(\int_{-\infty}^{z} (z-x) dF(x) + \int_{z}^{\infty} (x-z) dF(x)\right)} \\
&= \frac{F(z) - \overbrace{\frac{\int_{-\infty}^{z} (z-x) dF(x)}{\int_{-\infty}^{\infty} |x-z| dF(x)}}^{\tau,\text{ notice}} (2F(z)-1)}{\left(\int_{-\infty}^{z} (z-x) dF(x) + \int_{z}^{\infty} (x-z) dF(x)\right)} \\
&= \frac{(1-\tau)F(z) + \tau(1-F(z))}{+\int_{-\infty}^{z} F(x)dx + \int_{z}^{\infty} (1-F(x))dx}.
\end{aligned}$$

This is recognizably the inverse of $\frac{\partial \mu_\tau}{\partial \tau}$ as in equation 6.10. It is also clearly positive.

The argument to complete the proof is as follows: So long as F has a first moment, it is clear that $E_X(\tau)$ produces a finite value for every $\tau \in (0,1)$. Similarly, it is clear that $E_X^{-1}(z)$ produces a value in $(0,1)$ for every z in the convex hull of the support of F. Existence and uniqueness of both the function and its inverse, together with the fact that both strictly positive derivatives (or left and right derivatives, at discontinuities in F), imply that both are continuous and monotone increasing. We already know that E_X^{-1} can be interpreted as a CDF, so continuity everywhere implies density everywhere, though the density may be discontinuous at discontinuities of F. $\square$

[12] It may be helpful to point out that

$$\begin{aligned}
\int_{-\infty}^{z} (z-x) dF(x) &= \int_{-\infty}^{z} z dF(x) - \int_{-\infty}^{z} x dF(x) \\
&= zF(z) - G(z)
\end{aligned}$$

and

$$\begin{aligned}
\frac{d}{dz} \int_{-\infty}^{z} (z-x) dF &= F(z) \\
\frac{d}{dz} \int_{z}^{\infty} (x-z) dF &= F(z) - 1
\end{aligned}$$

at continuity points of F. This also leads to the identity:

$$\int_{-\infty}^{z} (z-x) dF = \int_{-\infty}^{z} \frac{d}{dz} \int_{-\infty}^{z} (z-x) dF dz = \int_{-\infty}^{z} F(x) dx.$$

Expectile and Inverse-Expectile Functions Shown in Table 4.1

Here we derive expectile and inverse-expectile functions for a variety of distributions.

General Distribution F

As before, the τ^{th} expectile of a distribution F can be defined as either the minimizer (m-statistic)

$$\mu_\tau := \arg\min_{\mu} \int \varsigma_\tau(x-\mu)dF(x)$$

or using the first order condition from the same minimization problem (z-statistic)

$$\tau \int_{\mu_\tau}^{\infty}(x-\mu_\tau)dF(x) + (1-\tau)\int_{-\infty}^{\mu_\tau}(x-\mu_\tau)dF(x) = 0.$$

Using the latter definition, it should be fairly clear that μ_τ exists and is finite for every $\tau \in (0,1)$ so long as F has a finite first moment. Accordingly, the value μ_τ can be obtained in implicit functional form as follows:

$$\begin{aligned}
&\tau \int_{\mu_\tau}^{\infty}(x-\mu_\tau)dF(x) + (1-\tau)\int_{-\infty}^{\mu_\tau}(x-\mu_\tau)dF(x) = 0 \\
\implies &\tau \int_{\mu_\tau}^{\infty} xdF(x) - \tau\mu_\tau \int_{\mu_\tau}^{\infty} dF(x) + (1-\tau)\int_{-\infty}^{\mu_\tau} xdF(x) - (1-\tau)\mu_\tau \int_{-\infty}^{\mu_\tau} dF(x) = 0
\end{aligned}$$

and

$$\mu_\tau = \frac{\tau \int_{\mu_\tau}^{\infty} xdF(x) + (1-\tau)\int_{-\infty}^{\mu_\tau} xdF(x)}{\tau \int_{\mu_\tau}^{\infty} dF(x) + (1-\tau)\int_{-\infty}^{\mu_\tau} dF(x)}$$

which can be expressed in other ways including the simplified form:

$$\mu_\tau = \frac{\int x|\tau - I(x<\mu_\tau)|dF}{\int |\tau - I(x<\mu_\tau)|dF}. \tag{6.11}$$

Alternative representations exist, but it is this author's opinion that the best representation, from an interpretability standpoint, is either the "weighted average" formula in equation 6.11, or the "point of balance" formula below

$$\tau \int_{\mu_\tau}^{\infty}(x-\mu_\tau)dF(x) = -(1-\tau)\int_{-\infty}^{\mu_\tau}(x-\mu_\tau)dF(x) \tag{6.12}$$

which makes clear that the ratio of left to right partial moments is $\frac{\tau}{1-\tau}$.

The inverse-expectile function for a distribution F is straightforward. Given a value $\mu_\tau \in [\inf F, \sup F]$, we take equation 6.12 and solve for τ.

$$\tau \int_{\mu_\tau}^{\infty}(x-\mu_\tau)dF(x) - \tau\int_{-\infty}^{\mu_\tau}(x-\mu_\tau)dF(x) = \int_{-\infty}^{\mu_\tau}|x-\mu_\tau|dF(x)$$

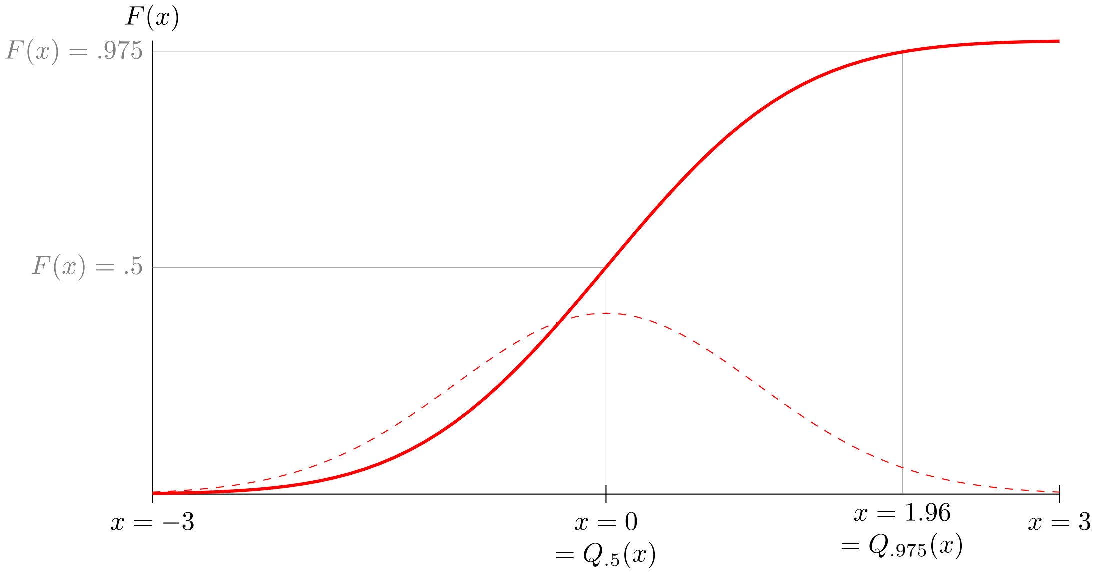

Figure 6.1: The CDF of a standard normal distribution in solid red with the corresponding density function in dashed red.

and

$$\tau = \frac{\int_{-\infty}^{\mu_\tau} |x - \mu_\tau| dF(x)}{\int_{-\infty}^{\infty} |x - \mu_\tau| dF(x)}. \tag{6.13}$$

This is synonymous with the representations given in equations 6.7 and 4.3.

Standard Normal Distribution $N(0,1)$

As one example, assume that the variable X is distributed according to the standard normal distribution

$$X \sim N(0,1).$$

The CDF and density of this distribution are shown in Figure 6.1. In that case, $F = \Phi(x)$, we have the left-moment $G(z)$ given by

$$\begin{aligned} G(z) &= \int_{-\infty}^{z} u d\Phi(u) \\ &= \int_{-\infty}^{z} u\phi(u) du \\ &= -\phi(z) \end{aligned}$$

so the expectile function can be written as an implicit function

$$\mu_\tau = E_X(\tau) = \frac{-(1-\tau)\phi(\mu_\tau) + \tau\phi(\mu_\tau)}{(1-\tau)\Phi(\mu_\tau) + \tau(1-\Phi(\mu_\tau))}$$
$$= \frac{2\tau\phi(\mu_\tau) - 1}{(1-2\tau)\Phi(\mu_\tau) + \tau}$$

which is relatively simple.[13] Of course, this mapping holds for any value in the convex support of the distribution, so the inverse of $E_X(\tau)$ can be obtained by rewriting equation 6.7 as

$$E_X^{-1}(\mu_\tau) = \frac{\Phi(\mu_\tau)\mu_\tau + \phi(\mu_\tau)}{2\phi(\mu_\tau) - (1-2\Phi(\mu_\tau))\mu_\tau} \tag{6.14}$$

with $E_X^{-1}(\mu_\tau)$ always defined in $(0,1)$. The inverse expectile function in equation 6.14 is shown together with the normal CDF in Figure 4.2.[14]

Uniform

Suppose next that Y is distributed uniformly on the interval (a,b).

$$Y \sim \mathcal{U}(a,b)$$

The expectile function $E_Y(\tau)$ can be attained as follows. From the definition,

[13]These follow by way of the identity

$$\int_a^b x\phi(x)dx = \frac{1}{\sqrt{2\pi}}\int_a^b xe^{-\frac{1}{2}x^2}dx = \phi(a) - \phi(b).$$

[14]In the case of a standard normal distribution, the inverse expectile function has density given by

$$\frac{\partial E_X^{-1}(z)}{\partial z} = \frac{-\Phi(z) + \Phi(z)z + \phi(z)}{[2\phi(z) - (1-2\Phi(z))z]^2}.$$

$$
\begin{aligned}
\mu_\tau(Y;\tau) &= \arg\min_\mu \int_a^b \varsigma_\tau(y-\mu)\frac{1}{b-a}dy \\
\implies \tau\int_a^{\mu_\tau} \frac{y-\mu_\tau}{b-a}dy &= -(1-\tau)\int_{\mu_\tau}^b \frac{y-\mu_\tau}{b-a}dy \\
\frac{\tau}{b-a}\int_a^{\mu_\tau}(y-\mu_\tau)dy &= -\frac{(1-\tau)}{b-a}\int_{\mu_\tau}^b (y-\mu_\tau)dy \\
=\frac{\tau}{b-a}\frac{(y-\mu_\tau)^2}{2}|_a^{\mu_\tau} &= -(1-\tau)\frac{(y-\mu_\tau)^2}{2}|_{\mu_\tau}^b \\
-\tau\frac{(a-\mu_\tau)^2}{2} &= -(1-\tau)\frac{(b-\mu_\tau)^2}{2} \\
\sqrt{\tau}(\mu_\tau-a) &= \sqrt{1-\tau}(b-\mu_\tau) \\
\mu_\tau(\sqrt{\tau}+\sqrt{1-\tau}) &= (\sqrt{\tau}a+\sqrt{1-\tau}b) \\
\mu_\tau &= \frac{\sqrt{\tau}a+\sqrt{1-\tau}b}{\sqrt{\tau}+\sqrt{1-\tau}}.
\end{aligned}
$$

And the inverse-expectile function can be found in the same way. Starting from the fifth line above, we have

$$
\begin{aligned}
\tau(a-\mu_\tau)^2 &= (1-\tau)(b-\mu_\tau)^2 \\
\implies \tau &= \frac{(b-\mu_\tau)^2}{(a-\mu_\tau)^2+(b-\mu_\tau)^2}.
\end{aligned}
$$

Bernoulli

A Bernoulli random variable takes a value of 1 with probability p and 0 with probability $1-p$, so

$$
Y = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{otherwise.} \end{cases}
$$

The expectile function $\mu_\tau = E_Y(\tau)$ of a Bernoulli random variable can be found as follows:

$$
\begin{aligned}
\mu_\tau(Y;\tau) &= \arg\min_\mu \int \varsigma_\tau(y-\mu)dF(y) \\
&= \arg\min_\mu \tau p(y-1)^2 + (1-\tau)(1-p)y^2 \\
\implies \tau p(\mu_\tau-1) &= -(1-\tau)(1-p)\mu_\tau \\
\tau p\mu_\tau + (1-\tau)(1-p)\mu_\tau &= \tau p \\
\implies \mu_\tau &= \frac{\tau p}{\tau p+(1-\tau)(1-p)}.
\end{aligned}
$$

This variation is quite intuitive: it is clear that μ_τ approaches one as τ or p approach one and that μ_τ approaches zero as τ or p approach zero.

The inverse-expectile function $\tau = E_Y^{-1}(\mu_\tau)$can be found by a similar method. From the third line above, we have:

$$
\begin{aligned}
\tau p(\mu_\tau - 1) &= -(1-\tau)(1-p)\mu_\tau \\
\implies \tau\left(p(1-\mu_\tau) + (1-p)\mu_\tau\right) &= (1-p)\mu_\tau \\
\text{so } \tau &= \frac{(1-p)\mu_\tau}{p(1-\mu_\tau) + (1-p)\mu_\tau}.
\end{aligned}
$$

Two-Point Distribution

Next we consider a generalized version of the Bernoulli distribution, where the random variable takes a value of b with probability p or a value of a with probability $1-p$. Without loss of generality, we assume that $b > a$.

$$
Y = \begin{cases} b & \text{w.p. } p \\ a & \text{otherwise.} \end{cases}
$$

The expectile function $E_Y(\tau)$ can be obtained as follows:

$$
\begin{aligned}
\mu_\tau(Y;\tau) &= \arg\min_\mu \int \varsigma_\tau(y-\mu)dF(y) \\
&= \arg\min_\mu \tau p(y-b)^2 + (1-\tau)(1-p)(y-a)^2 \\
\implies \tau p(\mu_\tau - b) &= -(1-\tau)(1-p)(\mu_\tau - a) \\
\tau p\mu_\tau + (1-\tau)(1-p)\mu_\tau &= \tau p + (1-\tau)(1-p)a \\
\implies \mu_\tau &= \frac{\tau pb + (1-\tau)(1-p)a}{\tau p + (1-\tau)(1-p)}.
\end{aligned}
$$

In this example, μ_τ approaches b as either τ or p approach unity and μ_τ approaches a as either τ or p approach zero. The inverse expectile function can be found starting from the third line above;

$$
\begin{aligned}
-\tau p(\mu_\tau - b) &= (1-\tau)(1-p)(\mu_\tau - a) \\
\implies \tau\left((1-p)(\mu_\tau - a) - p(\underbrace{\mu_\tau - b}_{<0})\right) &= (1-p)(\mu_\tau - a) \\
\implies \tau &= \frac{(1-p)(\mu_\tau - a)}{(1-p)(\mu_\tau - a) - p(b-\mu_\tau)}.
\end{aligned}
$$

Newey and Powell [1987] found that expectiles are equivariant to affine transforms, so it is also possible to derive these two results using transformations of the Bernoulli expectile and inverse expectile functions, or vice-versa. We leave that solution as an exercise.

Standard Normal Distribution

Supposing now that Y is a standard normal random variable:

$$
Y \sim N(0,1).
$$

We can obtain the expectile function $E_Y(\tau)$:

$$\begin{aligned}
\mu_\tau(Y;\tau) &= \arg\min_\mu \int \varsigma_\tau(y-\mu)dF(y) \\
&= \arg\min_\mu \int_{-\infty}^{\mu} (1-\tau)(y-\mu)^2\phi(y)dy \\
&\quad + \int_{\mu}^{\infty} \tau(y-\mu)^2\phi(y)dy \\
\implies \tau \int_{\mu}^{\infty} (y-\mu)\phi(y)dy &= -(1-\tau)\int_{-\infty}^{\mu}(y-\mu)\phi(y)dy \\
\text{so } \mu_\tau &= \frac{-(1-\tau)\phi(\mu_\tau)+\tau\phi(\mu_\tau)}{(1-\tau)\Phi(\mu_\tau)+\tau(1-\Phi(\mu_\tau))} \\
&= \frac{2\tau\phi(\mu_\tau)-1}{(1-2\tau)\Phi(\mu_\tau)+\tau}.
\end{aligned}$$

It is then straightforward to produce the inverse expectile function $E_Y^{-1}(\mu_\tau)$ by manipulating the function for μ_τ;

$$\tau = \frac{\Phi(\mu_\tau)\mu_\tau+\phi(\mu_\tau)}{2\phi(\mu_\tau)-(1-2\Phi(\mu_\tau))\mu_\tau}.$$

Koenker's Distribution

The distribution given by Koenker [1993] is designed specifically such that its inverse expectile function is equal to its CDF.

$$E_Y^{-1}(\mu_\tau) = F(\mu_\tau).$$

Koenker's distribution is a Student t-distribution with two degrees of freedom. This distribution has density on the reals. Accordingly, F is invertible and the quantile function $F^{-1}(\tau)$ is equal to the expectile function $E_Y(\tau)$. See also Zou [2014].