Measuring the Informativeness of Market Statistics

Kyungmin Kim

Measuring the Informativeness of Market Statistics

Kyungmin Kim*

Federal Reserve Board

September 14, 2016

Abstract

Market statistics can be viewed as noisy signals for true variables of interest. These signals are used by individual recipients of the statistics to imperfectly infer different variables of interest. This paper presents a framework under which the 'informativeness' of statistics is defined as their efficacy as the basis of such inference, and is quantified as expected distortion, a concept from information theory. The framework can be used to compare the informativeness of a set of statistics with that of another set or its theoretical limits. Also, the proposed informativeness measure can be computed as solutions to familiar problems under a range of assumptions. As an application, the measure is used to explain the difference in usage levels of temperature derivatives across different base weather stations. The informativeness measure is found to be at least as effective as city size measures in explaining the difference in usage levels.

1 Introduction

Market statistics convey information on a large number of variables. Individual recipients use market statistics to infer their own variables of interest, but the inference is usually far from perfect. If the set of statistics included all the variables that the recipients were interested in learning about, the statistics would perfectly satisty the recipients' informational need. However, the number of published statistics is typically small and not sufficient to cover all the variables of interest.

Given this limitation, it is natural to ask for a measure to quantify the informativeness of market statistics as an input for such inference. This paper proposes a framework under which a measure of informativeness can be defined. It is based on the idea that if there are M market statistics that convey information on K variables of interest, it is generally not possible to recover the K variables precisely from the M statistics, especially if $M < K$. Then, the expected distance between the K variables recovered from the M statistics and the actual values of the K variables can serve as a measure of informativeness. A smaller distance means that the statistics are more informative. This formulation of informativeness as the expected distance between actual variables and recovered variables is known as expected distortion in information theory.[1]

This definition of informativeness is distinct from two popular alternative ways to measure the quantity of information in the economics literature: Price impact and entropy. Many studies have measured the change in the price of an asset in response to new information, such as changes in bond ratings (Goh and Ederington (1993) and Pinches and Singleton (1978)) and divident announcement (Aharony and Swary (1980)). This method can be applied to the problem of this paper, by measuring the impact of changes in the market statistics on each variable of interest. This paper's contribution is to provide a framework under which the impact on multiple variables can be interpreted.

Entropy (Shannon entropy, as in Shannon (1948), or derived measures such as transfer entropy, as in Schreiber (2000)) is sometimes used as a general measure of the quantity of information conveyed by a set of random variables, for example, by Dimpfl and Peter (2012). Rate distortion theory relates entropy-related measures to informativeness, or expected distortion as it is known in information theory (see Cover and Thomas (2006), chapter 10). However, the complex form of the relationship discovered by rate distortion theory is not directly applicable within the scope of this paper.

The framework prposed by this paper consists of four elements: (i) The market statistics, (ii) the variables of interest, (iii) the recovery rule, and (iv) the distance measure. The

---

[1]For example, Cover and Thomas (2006), chapter 10.

framework can be used to evaluate a given set of statistics and to establish limits on the expected distance. Specifying the variables of interest is essential, as there is no sense in discussing informativeness without reference to which variables are being represented by the statistics. The recovery rule is restricted to be either linear or discrete, and either L^1 or L^2 norm is used as the distance measure. With these assumptions, the informativeness measure can be computed by routine methods.

The framework is useful both as a practical tool and as a component of empirical research. The measures can be used to compare the informativeness of different statistics and can tell how much potential improvement in informativeness is possible. Such measures can inform a decision between competing choices of statistics to publish or communicate. Also, they can be used to study how the informativeness of availabe statistics affects economic decisions, as a source of information or as a reference value for financial contracts. The use of informativeness measures can be readily extended to any economic problem in which the distance between actual and recovered variables matters, for example, in assessing the efficiency of insurance contracts.

As an application, this paper computes the informativeness of the temperature recordings at certain weather stations to explain the variation in the trade volume of temperature derivatives across weather stations. The informativeness measures are found to be as effective as city size measures in explaining the variation.

The rest of paper is organized as follows: Section 2 defines the informativeness measure and discusses their theoretical limits. Section 3 shows that how the measure can be computed as solutions to familiar problems, such as group means/medians, linear regression, k means/medians clustering and principal component analysis (PCA) under L_1 and L_2 norms. Section 4 applies the informativeness measure to explaining the cross-sectional variation in the trade volume of temperature derivatives across weather stations. Section 5 concludes.

2   Framework

Market statistics imperfectly communicate variables of interest. The informativeness measure represents the degree of imperfection.

Let $I_1, I_2, ..., I_M \in \mathbb{R}$ be M market statistics and let $z_1, z_2, ..., z_K \in \mathbb{R}$ be K variables of interest. The statistics and the variables of interest follow a static joint probability distribution, whose density is denoted by $p(I_1, ..., I_M, z_1, ..., z_K)$. Let $g : \mathbb{R}^M \to \mathbb{R}^K$ denote a recovery rule, which is the value of $(z_1, ..., z_K)$ inferred from $(I_1, ..., I_M)$. For $i = 1, 2, ..., K$, let g_i denote the i-th component of g, which is the recovery rule for z_i.

$d_i : \mathbb{R}^2 \to \mathbb{R}, i = 1, ..., K$, is the distance or penalty associated with the difference between

3

recovered and actual z_i. At this point, d_i can be any function, but for simplicity, $d_i = \alpha_i d$ for every $i = 1, ..., K$ and some metric $d : \mathbb{R}^2 \to \mathbb{R}$ that is common across the K variables of interest. α_i is a positive real number representing the weight given to the i-th dimension.

With these d_i's, the expected distance measure, D, is defined as follows:

$$D = \mathbb{E} \sum_{i=1}^{K} \alpha_i d_i(z_i, g_i(I_1, ..., I_M)). \tag{1}$$

As usual, $\mathbb{E}$ denotes expected value. The expected distance measure depends on three sets parameters: p, the joint distribution of the M market statistics and the K variables of interest; $(\alpha_1, ..., \alpha_K, d)$, representing the penalty for an imperfect recovery; and the recovery rule g. Therefore, D can be viewed as a function of the parameters, and written as $D(p, (\alpha_1, ..., \alpha_K, d), g)$.

This formulation is known as expected distortion in information theory, as mentioned in the introduction. However, this term will not be used in this paper because distortion has an unrelated established meaning in economics. Instead, the measure D will be simply referred to as informativeness measure or, sometimes, as expected distance.

The first two parameters, p and $(\alpha_1, ..., \alpha_K, d)$, define the objective of communicating market statistics. g can be interpreted both as the recovery rule that the individual recipients use or the recovery rule that the publisher of statistics expect the recipients to use. Given this ambiguity, there are alternative ways to choose a reasonable g. In this paper, g will be defined as the minimizer of D among a class of simple functions. The following two types of functions, or rules, are considered:

(i) Linear rule: g is a linear function of market statistics: For $i = 1, ..., K$, $g_i(I_1, ..., I_M) = \beta_{i,0} + \sum_{j=1}^{M} \beta_{i,j} I_j$.

(ii) Discrete rule: M partitions of $\mathbb{R}$ into n sets, $S(i, j)$ for $i = 1, ..., M$ and $j = 1, ..., n$, are given: $S(i, 1), S(i, 2), ..., S(i, n)$ form a partition, and i is an index that distinguishes the M partitions. g is constant on each set $\prod_{i=1}^{M} S(i, j_i)$ for any $j_1, j_2, ..., j_M \in \{1, ..., n\}$

In reality, the recipients may not achieve the minimum distance under either rule because they do not know the joint distribution p. The measures are useful only if the recipients are 'smart enough' to derive a reasonably good recovery rule. The simple forms that the linear rule and the discrete rule force on g partially reflect this limitation. In addition, these restrictions make D easily computable.

Let D_L be the minimum of D under linear g: $D_L = \min_g D(p, (\alpha_1, ..., \alpha_K, d), g)$, where g is a linear rule. Similarly, let D_D be the minimum of D under discrete g, with given partitions. D_L and D_D can be used to compare two sets of statistics in terms of their informativeness:

4

The set of statistics with smaller D_L or D_D is a better basis of inference for the variables of interest.

In absolute terms, how does a set of market statistics compare with the *best* possible set and the *worst* possible set? These questions are useful in thinking about how much improvement can be made by choosing more informative statistics and in interpreting the magnitude of the expected distance D_L and D_D. With a fixed number of statistic M, the expected distance of the best possible statistics, $\overline{D_L}$, is simply defined as the minimum of D under linear g and under an arbitrary joint density p: $\overline{D_L} = \min_{p,g} D(p, (\alpha_1, ..., \alpha_K, d), g)$. Similarly, $\overline{D_D}$ is defined as the minimum of D under discrete g for a given partition and an arbitrary p.

The worst set of statistics is the set of statistics that does not help infer the variables of interest at all. This happens if the statistics are statistically independent of the variables of interest, or more simply, if the statistics are constant. This condition is equivalent to g being a vector of constants.

Let D_0 be the expected distance from using the worst set of statistics. Then,

$$D_0 = \min_{g \in \mathbb{R}^M} \sum_{i=1}^{K} \alpha_i d_i(z_i, g_i). \tag{2}$$

Proposition 1. $D_0 \geq D_L \geq \overline{D_L}$ *and* $D_0 \geq D_D \geq \overline{D_D}$.

By definition, $D_L \geq \overline{D_L}$ and $D_D \geq \overline{D_D}$. Also, $D_0 \geq D_L$ because a constant recovery rule is a linear recovery rule with zero slopes. Similarly, $D_0 \geq D_D$ because a contant recovery rule is a discrete recovery rule with the same value on all the n^M products of partitioning sets.

Unfortunately, there is no closed-form solution for D_L, D_D, $\overline{D_L}$, $\overline{D_D}$ or D_0 in general, and no general computational technique to approximate the solutions exists either. However, if d is linear distance (Euclidean), $d(x, y) = |x - y|$, or if d is square distance, $d(x, y) = (x - y)^2$, there are known closed-form solutions or computational techniques. The next section describes them in more detail.

3   Solutions to the Minimization Problems

In this section, $(I_1, ..., I_M, z_1, ..., z_K)$ is assumed to follow a sample distribution of size T: Given T vectors in $\mathbb{R}^{M+K}$ indexed by t, denoted $(I_{t,1}, ..., I_{t,M}, z_{t,1}, ..., z_{t,K})$, the probability distribution of $(I_1, ..., I_M, z_1, ..., z_K)$ is a discrete random variable over $(I_{t,1}, ..., I_{t,M}, z_{t,1}, ..., z_{t,K})$ with uniform probability $1/T$.

The following four subsections explain how to compute D_L, $\overline{D_L}$, D_D, $\overline{D_D}$ and D_0.

3.1 Computation of D_L

The minimum expected distance under a linear recovery rule, D_L, can be computed by linear regression. Let $\hat{\beta}_{i,j}, 0 \leq j \leq M$ denote the coefficients of the linear regression of z_i on $I_1, ..., I_M$. The coefficients solve the minimization problem for each i: $\min\limits_{\beta_0, \beta_1, ..., \beta_M} \sum\limits_{t=1}^{T} d_i(z_{t,i}, \beta_0 + \beta_1 I_{t,1} + ... + \beta_M I_{t,M})$. SR_i is the sum of residuals transformed by d: $SR_i = \sum\limits_{t=1}^{T} d_i(z_{t,i}, \hat{\beta}_0 + \hat{\beta}_1 I_{t,1} + ... + \hat{\beta}_M I_{t,M})$.

Proposition 2. $D_L = (\frac{1}{T}) \sum\limits_{i=1}^{K} \alpha_i SR_i$.

Proof. By definition,

$$D_L = \min_{\beta_{j,k}} (\frac{1}{T}) \sum_{t=1}^{T} (\frac{1}{T}) \sum_{i=1}^{K} \alpha_i d(z_{t,i}, \beta_{i,0} + \beta_{i,1} I_{t,1} + ... + \beta_{i,M} I_{t,M}). \tag{3}$$

Since the term $d(z_{t,i}, \beta_{i,0} + \beta_{i,1} I_{t,1} + ... + \beta_{i,M} I_{t,M})$ does not include any $\beta_{j,k}$ for $j \neq i$, the minimization can be done before the summation, and the order of the summation can be switched:

$$D_L = (\frac{1}{T}) \sum_{i=1}^{K} \alpha_i SR_i, \tag{4}$$

where $SR_i = \min\limits_{\beta_{j,k}} \sum\limits_{t=1}^{T} d(z_{t,i}, \beta_{i,0} + \beta_{i,1} I_{t,1} + ... + \beta_{i,M} I_{t,M})$. This completes the proof. $\square$

If $d(x, y) = (x - y)^2$, SR_i can be simply computed from the least squares regression of z_i on $I_1, ..., I_M$. If $d(x, y) = |x - y|$, SR_i can be computed from median regression (see Portnoy and Koenker (1997), for example). More generally, for $d(x, y) = |x - y|^p, p \geq 1$, the linear regression is a convex optimization problem, also known as ℓ_p regression. For an example of computational methods to solve the optimization problem, see Dasgupta et al. (2009).

3.2 Computation of $\overline{D_L}$

Computing $\overline{D_L}$ requires finding $I_{t,i}$ that minimize D_L given $z_{t,i}$. If $M \geq K$, $\overline{D_L} = 0$ because the first K statistics, $I_{t,1}, ..., I_{t,K}$, can simply be set to equal $z_{t,1}, ..., z_{t,K}$ and D_L will be zero as a consequence. Therefore, it is assumed that $M < K$ in the following discussion of $\overline{D_L}$.

The problem of $\overline{D_L}$ can be transformed to a problem similar to principal component analysis (PCA):

Proposition 3. *For $z, w \in \mathbb{R}^K$, let $d_v(z, w) = \sum_{i=1}^{K} \alpha_i d(z_i, w_i)$, where the subscript i on a vector denotes its i-th component. Then,*

$$\overline{D_L} = (\frac{1}{T}) \min_{B_0, B_1} [\sum_{t=1}^{T} \min_u d_v(\mathbf{z}_t, B_0 + B_1 u)], \qquad (5)$$

where $\mathbf{z}_t = (z_{t,1}, z_{t,2}, ..., z_{t,K}), u \in \mathbb{R}^M, B_0 \in \mathbb{R}^K$ and B_1 is a $K \times M$ matrix.

Proof.

$$\overline{D_L} = (\frac{1}{T}) \min_{B_0, B_1, I_{t,i}} [\sum_{t=1}^{T} d_v(\mathbf{z}_t, B_0 + B_1 \mathbf{I}_t)], \qquad (6)$$

where $\mathbf{I}_t = (I_{t,1}, ..., I_{t,M})$. Equation (5) is obtained by moving the minimum over $I_{t,i}$ inside the summation and replacing $\mathbf{I}_t$ by u. □

In particular, with $d(x, y) = |x - y|^p$ for $p \geq 1$, the minimum exists: B_0 and u can be bounded given the $\mathbf{z}_t$'s and the columns of B_1 can be restricted to be orthonormal.

In addition, $d_v(\mathbf{z}_t, \mathbf{w}_t) = \sum_{i=1}^{K} \alpha_i |z_{t,i} - w_{t,i}|^p = \sum_{i=1}^{K} |\alpha_i^{1/p} z_{t,i} - \alpha_i^{1/p} w_{t,i}|^p$ for any $\mathbf{z}_t, \mathbf{w}_t \in \mathbb{R}^K$. Therefore, the weights α_i can be moved inside the absolute value expression, which yields

$$\overline{D_L} = (\frac{1}{T}) \min_{B_0, B_1} \sum_{t=1}^{T} [\min_u \sum_{i=1}^{K} |\alpha_i^{1/p} z_{t,i} - A(B_0 + B_1 u)|^p], \qquad (7)$$

where A is a $K \times K$ diagonal matrix with $A_{ii} = \alpha_i^{1/p}$. By redefining B_0 and B_1 as AB_0 and AB_1, respectively, A can be simply droppped from the expression for $\overline{D_L}$.

With $p = 2$, $\overline{D_L}$ can be computed by PCA using the following procedure:[2]

(1) Set $B_{0,i}$ to equal the mean of $z_{t,i}$ over t for each $i = 1, ..., K$.

(2) Define $z'_{t,i} = \sqrt{\alpha_i}(z_{t,i} - B_{0,i})$.

(3) Set B_1 to be the matrix of the first M principal components of $\mathbf{z}'_t, t = 1, ..., T$.

(4) Let $\mathbf{w}'_t$ be the projection of $\mathbf{z}'_t$ onto the column space of B_1.

(5) Sum the square of the Euclidean distance between $\mathbf{z}'_t$ and $\mathbf{w}'_t$ over t and divide by T to obtain $\overline{D_L}$.

With $p = 1$, $\overline{D_L}$ can also be computed by PCA generalized to L^1 norm by finding an M dimensional subspace that minimizes the sum of distance between $\mathbf{z}_t$ and the subspace. Some care must be taken with the use of the term PCA with L^1 norm, as it has been used

---

[2]The equivalence between the problem of finding the first M principal components and that of finding the minimum distance M-dimensional subspace follows from basic properties of PCA. See section 8 of Theil (1983), for example.

to describe at least three distinct optimization problems in recent studies.[3] Step 2 must be changed so that $z_{t,i} - B_{0.i}$ is multiplied by α_i, not by $\sqrt{\alpha_i}$. In addition, the objective function to be minimized is not convex and there is no known process to obtain its exact solution, as opposed to the ordinary PCA with square distance. Therefore, only computational solutions of B_0 and B_1 are available. For example, both Ke and Kanade (2005) and Brooks et al. (2013) develop methods to find B_1 with a given B_0. B_0 is sometimes chosen to be the vector consisting of medians of $z_{t,i}$ over t, but such a choice does not guarantee minimum in general.[4]

3.3 Computation of D_D

D_D minimizes $(\frac{1}{T}) \sum_{t=1}^{T} \sum_{i=1}^{K} \alpha_i d(z_{t,i}, g_i(I_{t,1}, ..., I_{t,M}))$ over all possible functions g_i that is constant on n^M subsets of $\mathbb{R}^M$ that are products M partitions of $\mathbb{R}$ into n sets. Let $L = n^M$ and let $P_1, P_2, ..., P_L$ be the L sets that partition $\mathbb{R}^M$, where $g_i(P_j)$ is a singleton for any $i = 1, ..., K$ and $j = 1, ..., L$. With $c_{i,j}$ defined as $g_i(P_j)$, D_D can be computed by minimizing the objective function with respect to $c_{i,j}$. This can be achieved by choosing $c_{i,j}$ to minimize the sum of the distance between $c_{i,j}$ and $z_{t,i}$ for t such that $\mathbf{I}_t \in P_j$.

Proposition 4.

$$D_D = (\frac{1}{T}) \sum_{j=1}^{L} \sum_{i=1}^{K} \alpha_i \min_c \sum_{\mathbf{I}_t \in P_j} d(z_{t,i}, c). \tag{8}$$

Proof.

$$D_D = (\frac{1}{T}) \min_{c_{i,j}} \sum_{t=1}^{T} \sum_{i=1}^{K} \alpha_i d(z_{t,i}, c_{i,j(t)}), \tag{9}$$

where $j(t)$ is the value of j such that $\mathbf{I}_t \in P_j$. Dividing the summation over T into L groups according to the value of $j(t)$ and moving min inside the summations produce the equation of the proposition. $\square$

If $d(x, y) = |x - y|^p$ and $p \geq 1$, a minimizer c of $\sum_{\mathbf{I}_t \in P_j} |z_{t,i} - c|^p$ exists because the minimand is a convex function of c. Especially, if $p = 1$ or $p = 2$, the minimizer c is the median or the

---

[3]Ke and Kanade (2005) and Brooks et al. (2013) present computational techniques to find an approximate minimal distance subspace, and their objective function is the same as that in equation (7). Kwak (2008) analyzes the problem of maximizing L_1 dispersion of Euclidean projection of given points. Euclidean projection does not find the point on a subspace that is closest to the original point under L_1 norm. In addition, unlike with L_2 norm, maximizing the dispersion of projections is not equivalent to minimizing the distance between original points and their projections with L_1 norm. Park and Klabjan (2014) studies the same problem as Kwak (2008) and another problem of minimizing the L_1 distance between given points and their Euclidean projection into a subspace.

[4]For example, Brooks and Jot (2012) describes R codes that 'center' datapoints with their median.

mean, respectively, of $z_{t,i}$ over t such that $\mathbf{I}_t \in P_j$. In other words, c can be computed as the group median or the group mean for each of the K dimensions of $\mathbf{z}_t$, and the group is defined by the index j such that $\mathbf{I}_t \in P_j$.

3.4   Computation of $\overline{D_D}$

$\overline{D_D}$ can be computed by finding $\mathbf{I}_t$ that minimizes D_D, given $\mathbf{z}_t$ and a partition of $\mathbb{R}^K$ into $L = n^M$ sets. As with D_D, The number L can be any number, not just n^M, but $L = n^M$ is used to be consistent with earlier sections. Also, the shape of the partition, other than the number of sets in it, is irrelevant because $\mathbf{I}_t$ can be chosen in an arbitrary way. Since $\mathbf{I}_t$ is arbitrary, it is sufficient to consider a partition of $\{1, 2, ..., T\}$ into L sets $P_1', .., P_L'$. Then, following proposition 4, computing $\overline{D_D}$ reduces to a problem of finding a partition of $\{1, ..., T\}$ into L sets minimizing D_D:

Proposition 5.

$$\overline{D_D} = \left(\frac{1}{T}\right) \inf_{P_1', ..., P_L'} \sum_{j=1}^{L} \sum_{i=1}^{K} \alpha_i \min_c \sum_{t \in P_j'} d(z_{t,i}, c) \tag{10}$$

With $d(x, y) = |x - y|^p$, the expression for $\overline{D_D}$ can be rewritten as follows:

$$\overline{D_D} = \left(\frac{1}{T}\right) \inf_{P_1', ..., P_L'} \sum_{j=1}^{L} \sum_{i=1}^{K} \inf_c \sum_{t \in P_j'} |\alpha_i^{1/p} z_{t,i} - c|^p \tag{11}$$

This problem is an example of partitional clustering (see Chapter 3.3 of Jain and Dubes (1988)). With $p = 2$, the minimizer c is simply the mean of $\sqrt{\alpha_i} z_{t,i}$ over $t \in P_j'$ for each $i = 1, ..., K$ and $j = 1, ..., L$. Finding the partition $P_1', ..., P_L'$ that minimizes the square distance between the group means and $\mathbf{z}_t$ has no closed-form solution, but the solution can be approximated by a k-means algorithm. Similarly, with $p = 1$, the minimizer c is the median of $\alpha_i z_{t,i}$, and a computational solution can be found by a k-medians algorithm.

3.5   Computation of D_0

$D_0 = D_D = \overline{D_D}$ if $n = 1$ and hence, $L = n^M = 1$. Therefore, D_0 can be computed by proposition 4.

4 Use of Temperature Derivatives

4.1 Market Structure and Variation in Trade Volume

The purpose of this section is to demonstrate the usefulness of informativeness measures in explaining the variation in trade volume in monthly HDD (heat degree days) options across 24 cities in the US. Payoff from temperature derivatives depends on the temperature recorded at certain weather stations. For example, Chicago Mercantile Exchange (CME) has various options and futures listed, which are based on HDD and CDD (cold degree days). HDD is the average of daily maximum and minimum temperature minus 65 degreees Fahrenheit, if that number is positive, and zero otherwise. Similarly, CDD is 65 degrees Fahrenheit minus the average of daily maximum and minimum temparature, if that number is positive, and zero otherwise.[5] HDD and CDD broadly represent the demand for energy for heating and cooling, respectively.

CME first introduced these temperature derivatives between 1999 and 2000 on 10 weather stations. As trade volume grew, they listed contracts on 14 additional weather stations by 2008. However, trade volume started declining rapidly after 2010, and derivatives on 16 stations have been completely or partially delisted. As of June 2016, only 8 stations have a full range of derivatives, and 8 additional stations have only a limited range of derivatives.[6] Table 1 lists the metropolitan areas, or cities, represented by these stations, and figure 1 shows the annual volume of trade on HDD monthly options, which account for a large part of temperature derivatives trading on CME.

Firms whose profit is sensitive to temperature, such as gas and electricity untilities, may trade these derivatives to insure themselves against the risk of too high or too low temperature. Pérez-González and Yun (2013) provides evidence that some firms, especially gas and electricity utilities, use temperature derivatives to hedge risk and argues that such hedging increases the value of those firms.

The volume of trades on derivatives based on different weather stations varies wildly. Across the 24 weather stations, the mean and the standard variation of the mean annual trade volume of monthly HDD options between 2009 and 2014 are 994 and 1710, respectively. The relatively large standard deviation is driven by a few weather stations with disproportionately large volumes, such as New York.

This variation in trade volume across the weather stations cannot simply be explained by the variation in the size of the cities represented. For example, the average trade volume

---

[5]This definition of HDD and CDD is from Chapter 403 of CME Rulebook.

[6]This history of CME temperature derivatives is based on Purnanandam and Weagley (2016) and public announcements by CME.

Full listing	Partial listing	Delisted
Atlanta, GA	Des Moines, IA	Baltimore, MD
Chicago, IL	Philadelphia, PA	Detroit, MI
Cincinnati, OH	Portland, OR	Salt Lake City, UT
New York, NY	Tucson, AZ	Colorado Springs, CO
Dallas-Fort Worth, TX	Boston, MA	Jacksonville, FL
Las Vegas, NV	Houston, TX	Little Rock, AR
Minneapolis-St. Paul, MN	Kansas City, MO	Los Angeles, CA
Sacramento, CA	Washington, DC	Raleigh-Durham, NC

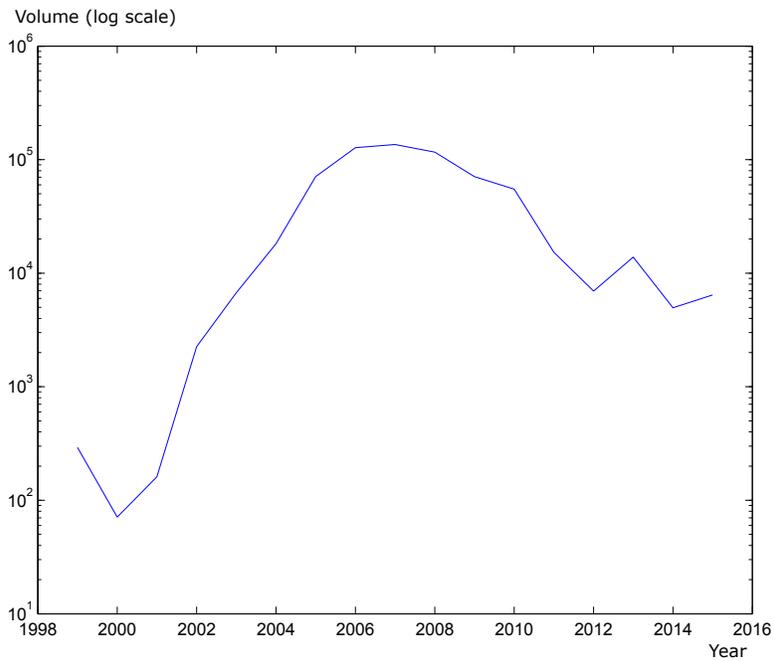Table 1: List of Represented Cities

Figure 1: Annual Trade Volume

for Boston and that for Washington were both lower than that for Portland, Sacramento, or Colorado Springs, even though Boston and Washington are both considerably larger than any of the three cities with larger trade volume.

The lack of close correlation between trade volume and city size may be explained by the fact that there is a close substitute for Boston- or Washington-based contracts, which is contracts based on New York. New York-based contracts can serve as a substitute because trade volume on New York-based contracts is large and the deviation of daily HDD from its monthly average shows a highly positive correlation between New York, Boston, and Washington, reflecting their geographic proximity.

Given this highly positive correaltion, considering eastern US as a whole, rather than a collection of individual cities, makes sense in understanding variation in trade volume. However, doing so involves an arbitrary choice of regional boundaries, and does not reflect varying temperature correlation across different pairs of eastern cities. The informativeness measure can address this problem by quantifying the amount of temperature variation represented by any group of cities. Indeed, informativeness measure does as well as city size measures in explaining the different levels of trade volume, as shown in the rest of this section.

4.2 Data and Method

Three types of data are used: (i) temperature, (ii) city size measures, and (iii) trade volume, on each of the 24 cities. The historical daily HDD values are publicly available on the CME website.[7] Six different measures of city size are used, which are population, GDP, annual and winter gas consumption, and annual and winter expenditure on gas.[8] City-level population comes from the Census Bureau,[9] and GDP data from the the Bureau of Economic Analysis.[10] The data are reported for each Metopolitan Statistical Area (MSA) and publicly available on the websites. State-level gas consumption and price are from the Energy Information Administration, again available on its public website.[11] State-level gas consumption is converted to city-level consumption by multiplying state consumption by the ratio of city population to state population. The list of cities in table 1 shows in which state

---

[7]CME Group Inc. 'Heating Degree Day (HDD), Historical Daily Data.' `http://www.cmegroup.com/market-data/reports/historical-weather-data.html` (accessed August, 2016).

[8]Gas consumption is measured by the volume of gas consumed, while expenditure is volume times price. These two measures are different because different prices apply to different cities. Winter is defined to be seven months from October to April.

[9]US Census Bureau. 'City-Level Population.' `http://www.census.gov/popest/` (accessed August, 2016).

[10]US Department of Commerce. 'Gross Domestic Product.' Bureau of Economic Analysis. `http://www.bea.gov/national/Index.htm` (accessed August, 2016).

[11]US Energy Information Administration. 'State-Level Consumption and Prices.' `http://www.eia.gov/petroleum/data.cfm` (accessed August, 2016).

each city is located. However, assigning one state to each city ignores the fact that certain city areas intersect with multiple states. As a consequence, the city-to-state population ratio is close to one for New York and much greather than one for Washington. Trade volume has been collected from a Bloomberg terminal.[12]

The general idea tested in this section is that the set of cities with large combined trade volume should explain a large amount of variation in daily HDD across cities. More specifically, let the integers $1, 2, ..., 24$ denote the 24 cities, with $U = \{1, 2, ..., 24\}$ denoting the universe of cities. For a subset C of U, $v(C)$ is the sum of trade volumes for the cities in C, and $D(C)$ is the informativeness of their HDDs in representing the HDDs of all the cities in U for winter months. Using the language of previous chapters, the HDDs of the cities in C are market statistics and those of the cities in U are variables of interest. This section tests whether a high $v(C)$ implies a small $D(C)$ (small expected distance means more informative), while a low $v(C)$ does not necessarily imply a large $D(C)$. This hypothesis implies both smaller mean and smaller standard deviation of $D(C)$ for larger $v(C)$, considering the distribution of $D(C)$ as functionally dependent on $v(C)$.

Following is the logic behind this hypothesis: A large $v(C)$ means that it is enough to trade in C to hedge against a large part of temperature variation. Therefore, temperature indices for C must capture a large part of temperature variation across U, leading to a small $D(C)$. At the same time, there can be alternative choices of C that have small $D(C)$ but do not have large $v(C)$, so a small $v(C)$ is compatible with both a small $D(C)$ and a large $D(C)$.

In comparing different C's, all possible choices of C as a six-element set has been considered. Sets that have too few or too many elements will automatically have both small $v(C)$ and $D(C)$ or both large $v(C)$ and $D(C)$, respectively. Sets with six elements were chosen because they generated a good mix of high $v(C)$ and low $v(C)$, and the number of possibilities was small enough to require only modest computational power.

In computing $v(C)$, the averge annual trade volume from 2009 to 2014 was used. In computing $D(C)$, the variables of interest were the deviations of daily HDD from its monthly average for each city. One interpretation of this choice is that mean monthly temperature can be perfectly predicted in advance, and the risk that needs to be insured against is the deviation of daily HDD from the predicted monthly average. Daily HDD data from May 2008 to the end of 2015 are used, because May 2008 is the first month with available historical data from the source. Finally, for normalization, each of the six city size measures for each year betwen 2009 and 2014 is normalized by dividing an individual city's measure by the sum of the measure over the 24 cities. Then, the average of each city's size measure over

---

[12]Bloomberg Finance LP. 'Trade Volume.' (accessed August, 2016).

	Population	GDP	Gas con.[2]	Gas exp.[3]	Gas con. (winter)	Gas exp. (winter)
Mean[4]	4.2	4.2	4.2	4.2	4.2	4.2
Median	2.4	2.4	2.4	2.5	2.5	2.4
St. dev.	4.3	4.8	5.2	5.4	5.2	5.4
Maximum (New York)	18.8	21.5	22.2	24.4	21.8	24.4

[1] Numbers are in percent.

[2] Shorthand for consumption.

[3] Shorthand for expenditure.

[4] Mean is identical by construction.

Table 2: Statistics on Size Measures

from 2009 to 2014 is used as the city's weight in computing $D(C)$.

$D(C)$ is computed as D_L under linear distance, and divided by D_0 for normalization. The other choices of informativeness measure, D_L with square distance or D_D with linear or square distance, had been tried as well for robustness and produced similar results.

For comparison, aggregate city size measures $S(C)$ were also computed, simply as the sum of city size measures over the cities in C.

4.3 Descriptive Statistics

The pairwise correlation between the six city size measures is close to one. Table 2 shows that the mean size measure is larger than the median, which is consistent with the fact that there are a few very large cities. New York is larger than any other city by far, and certainly much larger than the mean. This can be seen from both tables 2 and 3, the latter of which ranks the cities by their population.

Daily HDD minus its monthly average tends to show higher correaltion between cities that are geographically closer. Table 4 shows the correlation matrix for the following six cities: New York, Boston, Washington, Portland, Sacramento and Colorado Springs. The highly positive correlation between New York, Boston and Washington is consistent with the idea that temperature risks for Boston and Washington can be effectively insured by New York-based contracts, while temperature risks for the other cities cannot be. Figure 2 plots daily HDD minus monthly average in Boston and Colorado Springs against that in

Under 2	Between 2 and 5	Over 5
Des Moine (0.6)	Cincinnati (2.0)	Atlanta (5.2)
Colorado Springs (0.6)	Sacramento (2.1)	Washington (5.5)
Little Rock (0.7)	Portland (2.1)	Philadelphia (5.7)
Tucson (1.0)	Baltimore (2.6)	Houston (5.8)
Salt Lake City (1.1)	Minneapolis-St. Paul (3.2)	Dallas-Fort Worth (6.4)
Raleigh-Durham (1.1)	Detroit (4.1)	Chicago (9.1)
Jacksonville (1.3)	Boston (4.4)	Los Angeles (12.4)
Las Vegas (1.9)		New York (18.8)
Kansas City (2.0)		

[1] Numbers in parentheses are normalized population.
[2] The cities are listed in the order of increasing population in each column.

Table 3: Population Ranking of Cities

New York.

The city size measures and informativeness measures for individual cities are weakly correlated with trade volumes. The rank correlation between trade volume, $v(\{i\})$, and population, $S(\{i\})$, is 0.36, while the rank correlation between trade volume and informativeness, $D(\{i\})$, using population as weights, is -0.15. Figure 3 plots the rank of population and that of informativeness measure against the rank of trade volume, which shows no strong relationship.

	New York	Boston	Washington	Portland	Sacramento	Colorado Springs
New York	1.00	0.91	0.89	0.10	0.05	-0.09
Boston		1.00	0.76	0.11	0.07	-0.08
Washington			1.00	0.07	0.01	-0.08
Portland				1.00	0.49	0.38
Sacramento					1.00	0.32

Table 4: Correlation Matrix of HDD
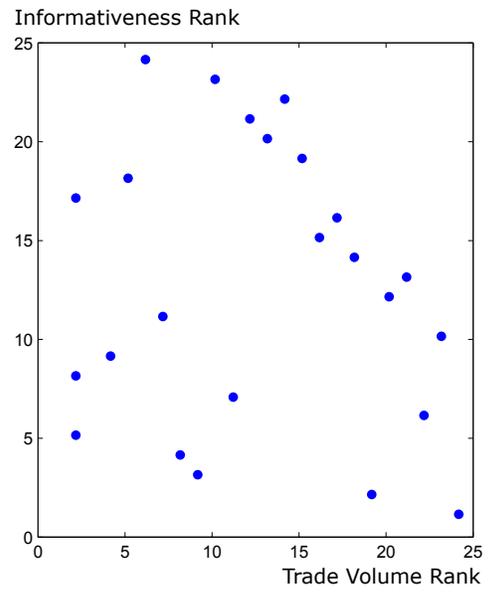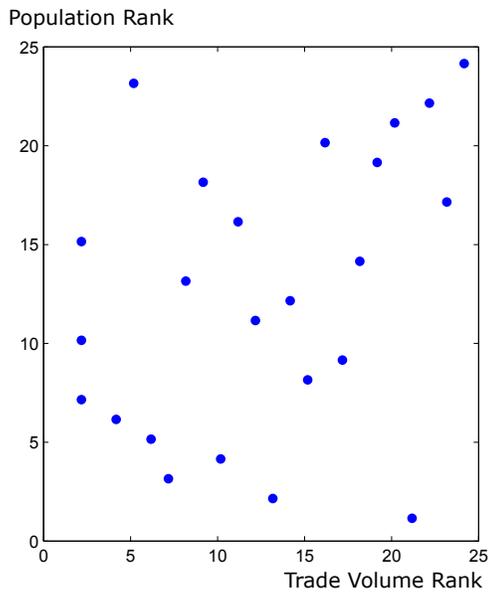

Figure 2: Correlation in HDD between Cities

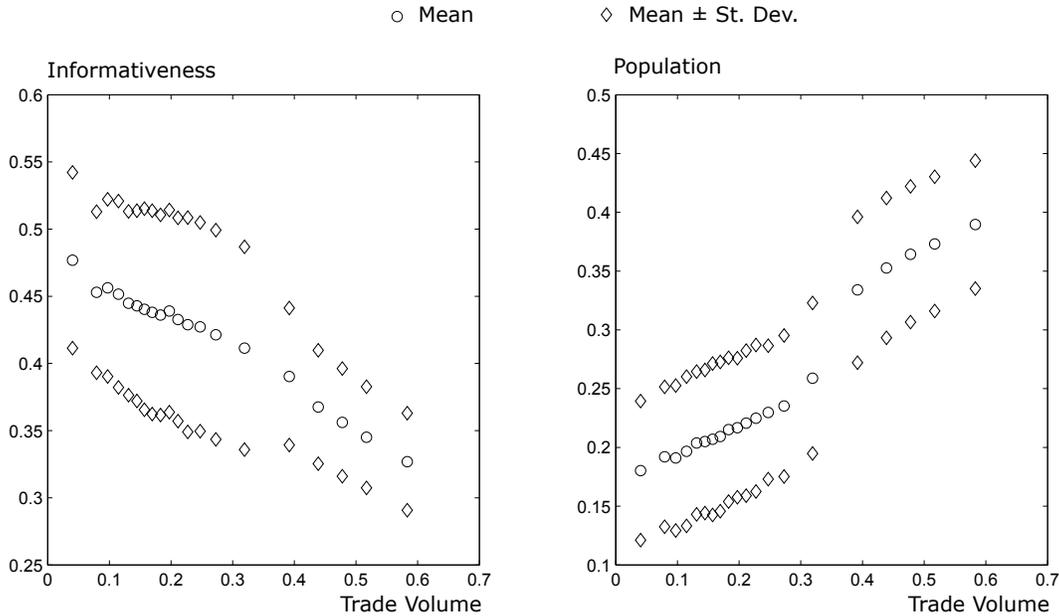Figure 3: Relationship between Trade Volume and City Size or Informativeness

Figure 4: Informativeness and Population as Functions of Trade Volume

4.4 Results

Figure 4 shows the distribution of informativeness and population as a function of volume. All the possible sets of 6 cities are divided into 20 equally-sized bins in the order of increasing total trade volume, $v(C)$. For each bin, the means of informativeness and population are computed, along with standard deviation, which represents how dispersed informativeness and popluation are for different levels of trade volume. As a reminder, the measure of informativeness referred to in this section is D_L (linear rule) with linear distance. As mentioned earlier, using other informativeness measures, either with square distance or with D_D (discrete rule), produces similar results.

Consistent with the hypothesis proposed in section 4.2, informativeness measure $D(C)$ decreases with trade volume, while population increases with trade volume. Consistent with the figure, the rank correlation between information distance and trade volume is -0.54, and that between population and trade volume is 0.65. Also, the standard deviation tends to be smaller for larger trade volume.

A potential problem with looking at every possible six-element set of cities is that New York dominates all other cities both in trade volume and population. Therefore, the negative relationship between informativenss measure and trade volume and the positive relationship between population and trade volume may just reflect the difference between the sets of cities which contain New York and those which do not. To address this issue, the exercise is repeated only with all six-element sets that contain New York.
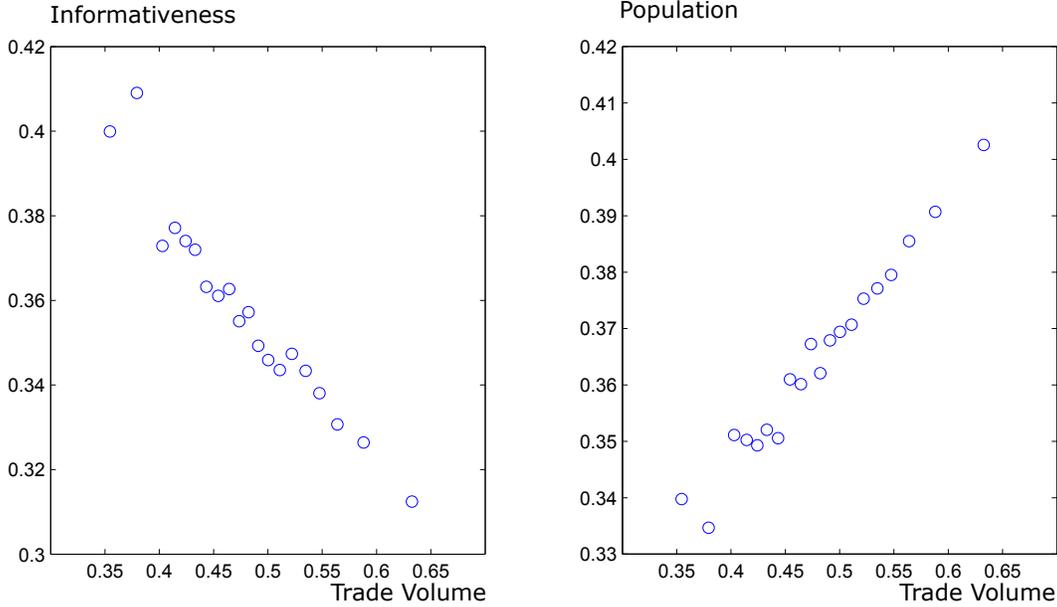
18

Figure 5: Mean of Informativeness and Population with New York

The result of this second exercise is consistent with that from the first exercise. Figure 5 shows the mean of informativeness and population as a function of trade volume, and figure 6 shows their standard deviation. The decrease in standard deviation is more evident with informativeness measure. In addition, the rank correlation between informativeness and trade volume is stronger at -0.48, compared with 0.28 between population and trade volume. This result suggests that the dominating effect of New York is stronger with population than with informativeness measure.

Among all the six-element sets, the mean value of informativeness for the top quintile of trade volume is 0.35, compared with 0.44 on average for the other four quintiles. Similarly, the mean population for the top quintile is 0.37, compared with 0.22 for the others. The standard deviation of informativeness measure for the top quintile is 0.42, compared with 0.72 for the others. This decrease in the standard deviation is more pronounced than that for population, with 0.59 for the top quintile and 0.65 for the rest. Table 5 shows the mean and the standard deviation for each quintile.

Redoing the exercise only with sets containing New York produces similar results, and the standard deviation of the informativeness measure decreases more consistently with trade volume than that of population. Table 6 shows the mean and the standard deviation for each quintile, using only the sets containing New York.

Overall, both the informativeness measure and population show behavior consistent with the hypothesis that (i) smaller informativeness measure (more informative) and higher size are associated with higher trade volume, and (ii) the variability in informativeness and size
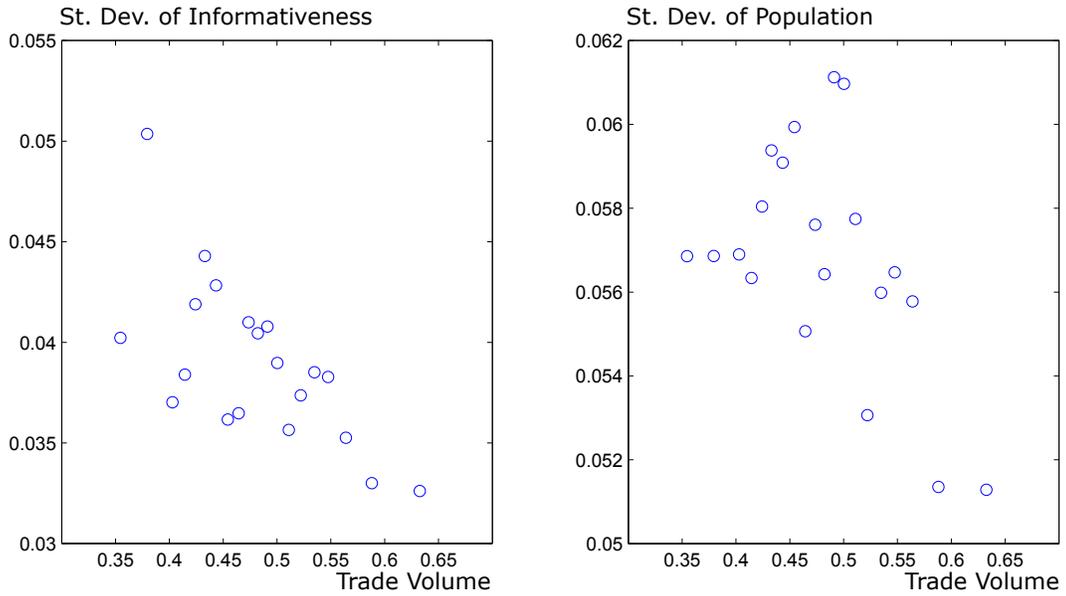
19

Figure 6: St. Dev. of Informativeness and Population with New York

All Sets

Quintiles	Informativeness	Population	Trade Volume
Top	0.35 (0.042)[1]	0.37 (0.059)	0.50 (0.057)
Second	0.41 (0.073)	0.26 (0.074)	0.31 (0.057)
Third	0.43 (0.076)	0.22 (0.061)	0.20 (0.017)
Fourth	0.44 (0.073)	0.21 (0.062)	0.15 (0.015)
Bottom	0.46 (0.066)	0.19 (0.061)	0.08 (0.029)

[1] Numbers in ( ) are standard deviations.

Table 5: Informativeness and Population for Each Volume Quintile

Sets Containing New York

Quintiles	Informativeness	Population	Trade Volume
Top	0.33 (0.036)[1]	0.39 (0.054)	0.58 (0.035)
Second	0.35 (0.038)	0.37 (0.057)	0.52 (0.013)
Third	0.36 (0.040)	0.36 (0.058)	0.48 (0.010)
Fourth	0.37 (0.042)	0.35 (0.059)	0.44 (0.012)
Bottom	0.39 (0.045)	0.34 (0.057)	0.39 (0.024)

[1] Numbers in ( ) are standard deviations.

Table 6: Informativeness and Population for Each Volume Quintile for Sets Containing New York

is smaller for higher trade volume.

5 Conclusion

This paper introduced a framework to quantify the informativeness of market statistics under different assumptions on how the recipients of the statistics infer their variables of interest from the statistics and how the deviation from true value is penalized. In particular, linear and discrete inference rules and linear and square penalties for deviation were shown to lead to easily computable measures for both informativeness and its theoretical bounds.

This paper also used the informativeness measures to explain the different levels of trade volume in temperature derivatives across different weather stations within the US. Informativeness explains the variation in trade volume at least as well as city cize measures do. This example shows that the proposed informativeness measures are useful in understanding why some market statistics, or 'numbers/variables' more generally, are more frequently referred to or adapted as bases of financial contracts.

References

Aharony, Joseph and Itzhak Swary (1980) "Quarterly Dividend and Earnings Announcements and Stockholders' Returns: An Empirical Analysis," *Journal of Finance*, Vol. 35, pp. 1–12.

Brooks, J. P., J. H. Dulá, and E. L. Boone (2013) "A Pure L_1-Norm Principal Component Analysis," *Computational Statistics & Data Analysis*, Vol. 61, pp. 83–98.

Brooks, J. Paul and Sapan Jot (2012) "pcaL1: An Implementation in R of Three Methods for L1-Norm Principal Component Analysis," *Optimization Online*.

Cover, Thomas M. and Joy A. Thomas (2006) *Elements of Information Theory*, Hoboken, NJ: John Wiley and Sons.

Dasgupta, Anirban, Petros Drineas, Boulos Harb, Ravi Kumar, and Michael W. Mahoney (2009) "Sampling Algorithms and Coresets for ℓ_p Regression," *SIAM Journal on Computing*, Vol. 38, pp. 2060–2078.

Dimpfl, Thomas and Franziska J. Peter (2012) "Using Transfer Entropy to Measure Information Flows between Financial Markets," *SFB 649 Discussion Paper*.

Goh, Jeremy C. and Louis H. Ederington (1993) "Is a Bond Rating Downgrade Bad News, Good News, or No News for Stockholders?" *Journal of Finance*, Vol. 48, pp. 2001–2008.

Jain, Anil K. and Richard C. Dubes (1988) *Algorithms for Clustering Data*, Englewood Cliffs, NJ: Prentice Hall.

Ke, Qifa and Takeo Kanade (2005) "Robust L_1 Norm Factorization in the Presence of Outliers and Missing Data by Alternative Convex Programming," in *IEEE Conference on Computer Vision and Pattern ¿ Recognition (CVPR 2005)*, July.

Kwak, Nojun (2008) "Principal Component Analysis Based on L1-Norm Maximization," *IEEE Transactions on Pattern Aanlysis and Machine Intelligence*, Vol. 30, pp. 1672–1680.

Park, Young Wook and Diego Klabjan (2014) "Algorithms for L1-Norm Principal Component Analysis."

Pérez-González, Francisco and Hayong Yun (2013) "Risk Management and Firm Value: Evidence from Weather Derivatives," *Journal of Finance*, Vol. 68.

Pinches, George E. and J. Clay Singleton (1978) "The Adjustment of Stock Prices to Bond Rating Changes," *Journal of Finance*, Vol. 33, pp. 29–44.

Portnoy, Stephen and Roger Koenker (1997) "The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators," *Statistical Science*, Vol. 12, pp. 279–300.

Purnanandam, Amiyatosh and Daniel Weagley (2016) "Can Markets Discipline Government Agencies? Evidence from the Weather Derivatives Market," *Journal of Finance*, Vol. 71.

Schreiber, Thomas (2000) "Measuring Information Transfer," *Physical Review Letters*, Vol. 85, pp. 461–464.

Shannon, C. E. (1948) "A Mathematical Theory of Communication," *The Bell System Technical Journal*, Vol. 27, pp. 379–423.

Theil, Henri (1983) "Linear Algebra and Matrix Methods in Econometrics," in Zvi Griliches and Michael D. Intriligator eds. *Handbook of Econometrics*, Vol. 1: Elsevier B. V. Chap. 1, pp. 3–65.