Portfolio Risk Management Using the Lorenz Curve

Haim Shalit

Haim Shalit
is a professor of economics at Ben-Gurion University of the Negev in Beer-Sheva, Israel.
shalit@bgu.ac.il

The Lorenz curve was developed in 1905 for the purpose of measuring relative inequality in income distribution. Since then, the curve has been widely used in welfare economics to calculate the share of national income earned by proportions of the population, as ranked by their relative incomes. The application of the Lorenz curve to portfolio risk management is rooted in its ability to measure the distribution and variation of asset returns. In this article, I show how some recent quantitative risk measures can be derived from the Lorenz curve, in order to manage risk and construct efficient portfolios.

In risk studies, the analysis is based on the absolute Lorenz curve (hereafter referred to as the Lorenz), which ranks conditional expected returns with respect to the cumulative probabilities of getting these returns. In risk analysis, the Lorenz originated from Shorrocks [1983], who used absolute curves to derive second-degree stochastic dominance (SSD) conditions. Most portfolio theory and risk management results derived from the Lorenz have appeared in the finance literature with respect to Gini's mean difference (GMD) and the Gini index. Fisher and Lorie [1970] were the first to apply Gini statistics using the standard Lorenz curve to study the variability of single stocks and portfolios. Later, Shalit and Yitzhaki [1984] used the curve to characterize risky assets, apply Gini's mean difference in finance theory, and derive the mean-Gini CAPM.

The Lorenz is also very useful for expressing safety-first risk quantile measures, such as value at risk (VaR) and conditional value at risk (CVaR), that have become very popular in the banking industry. This feature is particularly advantageous when the analyst does not need to specify a particular distribution function. Otherwise, VaR and CVaR measures are quite cumbersome to compute.

In this article, I obtain risk measures from the Lorenz by using discrete probabilities, which are commonly available in financial data. The next article section presents the Lorenz and its relation to SSD. The following section discusses the Lorenz and its relation to GMD. Then, I show the link between VaR, CVaR, and the Lorenz. Finally, I present an investment example that shows how to manage portfolio risk with the Lorenz.

STOCHASTIC DOMINANCE AND THE LORENZ

The main advantage of the Lorenz in financial analysis lies in its simplicity as a tool to rank and evaluate risky assets according to stochastic dominance (SD). Hanoch and Levy [1969], Hadar and Russell [1969], and Rothschild and Stiglitz [1970] independently developed SD rules that provide portfolio efficiency under expected utility maximization, but without resorting to specific utility functions. They obtain SD rules by comparing the cumulative probability distributions of asset returns. First-degree stochastic dominance (FSD) is designed for investors with increasing utilities who are either risk-averters or risk-lovers. Second-degree stochastic dominance (SSD) as mainly for risk-averse investors is the most common model used in portfolio selection. Using cumulative probabilities, SSD rules provide the necessary and sufficient conditions under which all risk-averse expected-utility maximizers prefer risky assets. Nonetheless, the optimal asset choice is more evident with the Lorenz than with the traditional SSD rules.

Consider two risky assets, A and B, whose returns x_i^A for $i = 1,\ldots,N$ and x_i^B for $i = 1,\ldots,N$ are sorted from their lowest yields to their highest. Returns are distributed with probabilities $q^A(x^A)$ and $q^B(x^B)$, respectively, such that $q(x_i^A) \geq 0$ and $\sum_{i=1}^{N} q(x_i^A) = 1$, $q(x_i^B) \geq 0$ and $\sum_{i=1}^{N} q(x_i^B) = 1$. The cumulative probabilities are computed as $\sum_{i=1}^{k} q(x_i^A) = p_k^A$ and $\sum_{i=1}^{k} q(x_i^B) = p_k^B$ for $k = 1,\ldots,N$. Following Levy [2006], we can state the SSD rules by comparing the areas under the cumulative probabilities, namely:

All risk averters prefer asset A to asset B if and only if

$$\sum_{i=1}^{k} p_i^A \leq \sum_{i=1}^{k} p_i^B \quad \text{for all } k = 1,\ldots,N \tag{1}$$

Computing the areas under the cumulative probabilities is not straightforward, as it must be done for all the probabilities and followed by a comparison between the cumulative areas. Following Shorrocks [1983], using the Lorenz to establish SSD involves comparing only between curves. For all risk-averse investors to prefer portfolio A over portfolio B, the Lorenz of A must lie above the Lorenz of B. In other words, asset A dominates asset B if and only if:

$$L_A(p) \geq L_B(p) \quad \text{for all probabilities } 0 \leq p \leq 1, \tag{2}$$

where $L_A(p)$ and $L_B(p)$ are the Lorenz curves for asset A and B, respectively, formulated using Equations (3) and (4) as follows:[1]

In Equation (3), the Lorenz is obtained by summing up the returns x_i times their probabilities $q(x_i)$ up to the counter $k(p)$:

$$L(p) = \sum_{i=1}^{k(p)} x_i q(x_i) \quad \text{for all } p \text{ from 0 to 1} \tag{3}$$

where the counter $k(p)$ is obtained implicitly by:

$$p = \sum_{i=1}^{k(p)} q(x_i) \quad \text{for all } x_i \tag{4}$$

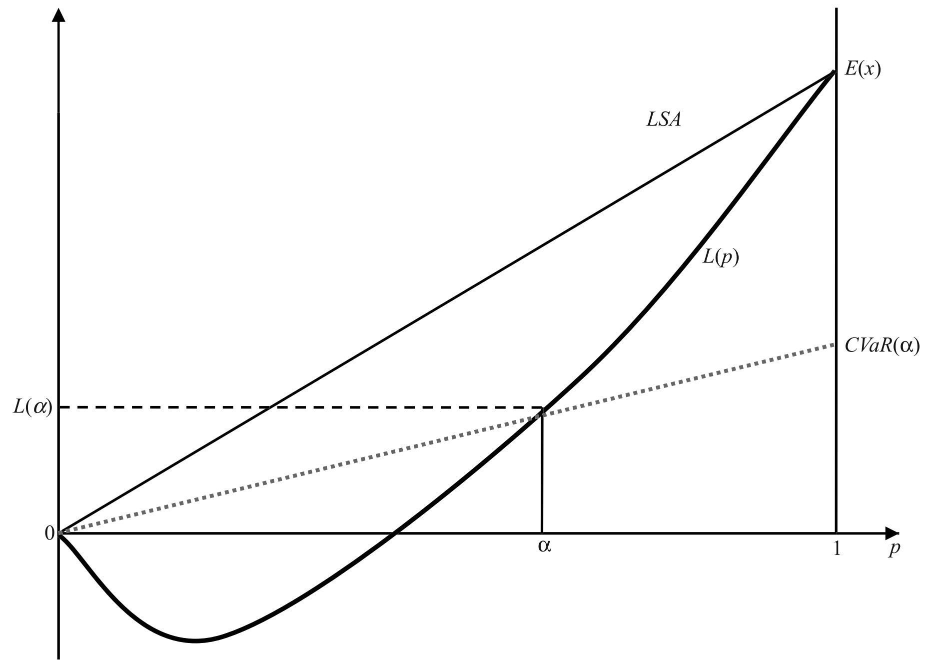
Equation (4) expresses the cumulative probability p that returns are less than a given value specified by $k(p)$.

I shall now explain the Lorenz delineated in Exhibit 1. Cumulative probabilities are exhibited on the horizontal axis, indicating that returns are ranked in increasing value. On the vertical axis, we see cumulative rates of returns weighted by probabilities, as expressed by Equation (3). The Lorenz starts at the origin of axes (0,0) and accumulates the sorted returns multiplied by their probabilities, until all the returns are used up. Because the lowest returns can be losses, the Lorenz may result in negative values. The curve ends at the mean return $E(x)$ on the parallel vertical axis, because at that point, all returns are used up and multiplied by their probabilities.

From its definition in Equation (3), the Lorenz captures the conditional expected return $E(x\,|\,p)$ given the probability p since $L(p) = E(x\,|\,p) \cdot p$ where $E(x|p) = \sum_{i=1}^{k(p)} x_i \frac{q(x_i)}{p}$ is the mean of returns when ranked returns add up to $k(p)$. When all returns are accounted for, i.e., $k(p) = N$, the Lorenz at $p = 1$ is the unconditional mean return of asset $E(x)$.

Now we can look at what can be gained by using the Lorenz in finance. The rationale for using the Lorenz in SSD is rooted in the manner by which the Lorenz characterizes risk and mean return of investments for risk-averse investors. Such investors have concave utility functions that express declining marginal utility. The horizontal axis in Exhibit 1 shows the probabilities of asset returns, ranked from those generating the lowest returns with the highest marginal utility to those generating the highest returns with the lowest marginal utility. The ranking of asset returns is the only information needed to sort an asset according to decreasing marginal utility. This ordering is specified by the cumulative returns, multiplied by the probabilities of getting these returns. This is basically the Lorenz. The principle of distributing resources according to decreasing marginal utility or decreasing marginal product ensures that financial resources are allocated optimally. Using the Lorenz to manage portfolio

Exhibit 1
The Lorenz



risk guarantees that objective. Because the curve expresses asset behavior not as a function of returns over time, but as the occurrence of having lower and higher returns, it provides much more relevant information about risk and return than periodical charts do.

RISK AND GINI'S MEAN DIFFERENCE

Originally, Gini [1912] defined the mean difference as the expected distance between observation pairs as follows:

$$\text{GMD} = E\,|\,x_1 - x_2\,| \tag{5}$$

where x_1 and x_2 are independent replicates of the random variable x. When GMD is divided by the mean, it becomes the well-known Gini coefficient, which is known to evaluate income inequality. GMD is also an attractive measure of risk, because it depends on the spread of the returns among themselves, and not on their deviations from a central value as the mean. The measure has many different representations and formulations, most of which can be found in Yitzhaki [1998]. In finance and portfolio risk management, it is more convenient to use one-half of GMD, which is usually referred to as the Gini Γ. As shown in Equation (6), its formulation uses the covariance between the returns and the cumulative probabilities of getting these returns:

$$\text{GMD}/2 \equiv \Gamma = 2\,\text{cov}[x, p] \tag{6}$$

Being a statistic, the Gini has some advantages, because it lets us express risk with a single number and construct optimal portfolios that are SSD. As I will show, an asset's Gini can be obtained directly from its Lorenz.

Indeed, when investors analyze the features of risky assets, they would like to decompose the returns into two components: one that embraces only the risk of the asset, and the other only the safe return. With the Lorenz, this task is easily accomplished. To show this, I construct a virtual asset that has the same mean return as asset x, but has no risk whatsoever. For each probability the virtual safe asset always yields the same mean return $E(x)$. This riskless asset is depicted in Exhibit 1 by its Lorenz as a straight line that originates at (0, 0) and ends at the mean $(1, E(x))$. This line is called the line of safe asset (LSA), because it expresses the expected return $E(x)$ multiplied by the probability p.

We can now enunciate the risk of asset x as the difference between its LSA, which yields the expected return $E(x)$, and the Lorenz of asset x. It is obvious that, for every probability p, investing in the risky asset earns the cumulative expected return along the Lorenz, while investing in the riskless asset earns a higher cumulative expected return along the LSA. The risk of the asset is quantified by the vertical differences between the LSA and the Lorenz. This area is calculated as the difference between the area below the LSA and the area below the Lorenz. Therefore, the farther the LSA is from the Lorenz, the greater the risk assumed by the asset. Because $E(x)p$ is the LSA equation and $L(p)$ is the Lorenz equation, the area between the two lines is one half the Gini, as shown here:

$$\sum_{i=1}^{N} E(x)p_i - \sum_{i=1}^{N} L(p_i) = \sum_{i=1}^{N} [E(x) - x_i]p_i = \mathrm{cov}[x,p] = \tfrac{1}{2}\Gamma \tag{7}$$

From Equation (7), the Gini can be understood as the pure risk inherent in an asset and therefore can be used together with the mean to characterize investments. Since the mean and the Gini are statistics derived from the Lorenz, they can facilitate the ranking of risky assets. Indeed, using the Lorenz for SSD defines only a partial ordering of investment opportunities. When Lorenz curves intersect, we cannot determine a clear dominance between risky assets, and therefore the relation between all the investments cannot be established. Sometimes a complete ordering is required, although these results can only provide the necessary conditions for SSD. This is the case when the mean and the Gini are used to establish necessary conditions for SSD. To clarify this argument, consider non-intersected Lorenz curves and their relation to SSD. If we choose a linear utility function to determine the optimal portfolio, a necessary condition for the risky portfolio to be preferred by all expected utility maximizers is that it is preferred by the risk-neutral investor whose marginal utility is a constant. As such, only the last data point on the Lorenz, which is the mean, is the relevant gauge for choosing among assets. This explains the first necessary condition for SSD: stating that the mean of the preferred asset is greater than the mean of the dominated asset.

The other necessary condition for SSD is that the area below the Lorenz of the preferred asset be greater than the area below the Lorenz of the dominated asset. This area is one-half the mean return, subtracted by $\tfrac{1}{2}\Gamma = \mathrm{cov}[x,p]$. These two requirements explain the necessary conditions for SSD, using the mean and the Gini. As established by Yitzhaki [1982], these necessary conditions are expressed as

$$\begin{gathered} E(x_A) \geq E(x_B) \\ E(x_A) - \Gamma_A \geq E(x_B) - \Gamma_B \end{gathered} \tag{8}$$

implying that, if portfolio A is SSD preferred to portfolio B, then the mean and the risk-adjusted mean return of A cannot be less than the mean and the risk-adjusted mean return of B, when risk is measured by the portfolio's Gini.[2]

THE LORENZ AND CVAR

As a popular measure of risk, VaR quantifies exposure to risk as the amount of cash to be held in a safe asset to overcome a portfolio's potential total loss. It is a safety-first risk measure defined as the quantile of a given probability p, formulated implicitly as the return $VaR(p)$, such that

$$p = \left\{ \sum_{i=1}^{k(p)} q(x_i) \,\middle|\, x_i \leq VaR(p) \right\} \tag{9}$$

As seen from Equations (3) and (4), $VaR(p)$ is only one single element of the Lorenz that can be obtained directly from the cumulative probabilities p. It is surprising that VaR is so prevalent in finance, as it lacks the following basic properties of a risk measure $\rho(X)$ for it to be coherent (see Artzner et al [1999]):

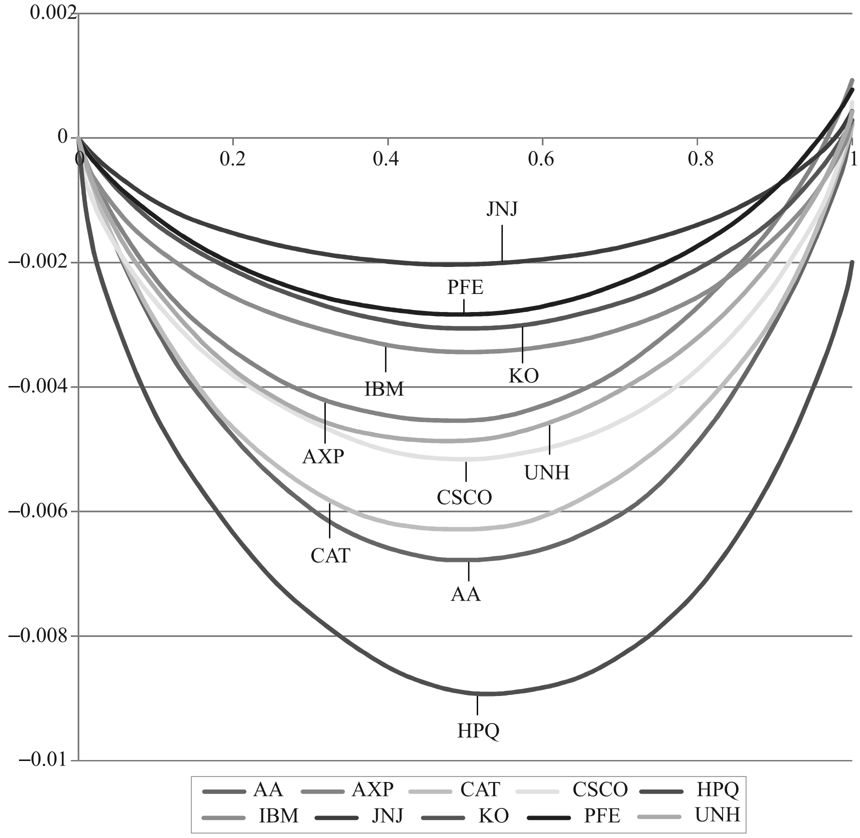
(1) Translation invariance, $\rho(X + R_F) = \rho(X) - R_F$, where R_F is a safe return
(2) Subadditivity, $\rho(A + B) \leq \rho(A) + \rho(B)$
(3) Positively homogeneity, $\rho(\lambda X) = \lambda\rho(X)$
(4) Monotonicity $X \leq Y \Rightarrow \rho(Y) \leq \rho(X)$

Indeed, unless the returns distribution is normal, VaR lacks coherence because it fails to satisfy the subadditivity axiom that would prevent risk reduction in portfolio diversification.

To circumvent VaR's lack of coherence, finance researchers developed conditional value at risk (Rockafellar and Uryasev [2000], Acerbi and Tasche [2002]). The basic idea of this measure is to calculate $CVaR(p)$ as the mean of all the quantiles below the original VaR in the lower

EXHIBIT 2

Lorenz Curves of 10 Select DJIA Stocks Using Daily Returns for 2012



tail of the cumulative probability distribution from 0 to p. In formal terms:

$$CVaR(p) = \frac{1}{p}\sum_{i=0}^{k(p)} VaR(q_i)\,q(x_i) \quad \text{for all } q(x_i) \le p \qquad (10)$$

By comparing Equation (3) and Equation (10), we see that CVaR is easily obtained from the Lorenz:

$$CVaR(p) = -\frac{L(p)}{p} \qquad (11)$$

Consider a specific probability α between 0 and 1. Exhibit 1 shows how CVaR(α) for the probability α is expressed by the slope of the straight line connecting the origin (0,0) to the point (α, $L(\alpha)$). This slope can also be measured on the vertical axis at $p = 1$ by the segment from the horizontal axis, up to the point labeled CVaR(α). As such, it is easier to calculate CVaR for a given asset, because the technique is not restricted to specific probability distributions. Under these provisions, CVaR is obtained from a specific value of the Lorenz, which is estimated by sorting and summing up the returns for a given dataset.

MANAGING RISKY ASSETS: AN INVESTMENT EXAMPLE

To show how the various risk measures are obtained, I calculate the Lorenz for various traded stocks. This is an easy task for a sample of discrete observations, because it involves only ranking returns in ascending order and then summing all the lower returns up to that observation for each given return.

To illustrate the relevance of the Lorenz in ranking securities with respect to risk and return, we use the 250 daily returns of the 30 stocks of the Dow Jones Industrials Average from January 3, 2012, to December 31,

2012. We calculate the Lorenz curves for these 30 stocks as described in Equation (3), where the probability of occurrence is 1/250 for each return. For the sake of clarity, Exhibit 2 plots only 10 Lorenz curves. Now we can isolate the set of SSD-dominated stocks, which are the ones with the Lorenz curves lying on top of the chart. The set of efficient stocks include JNJ, KO, IBM, and PFE, whose Lorenz curves form the non-dominated set. The worst stocks, according to SSD, are the ones lying on the bottom of the chart. These include AA, HPQ, CAT, and CSCO. Not only the position of the Lorenz curves is relevant, but also whether or not the curves intersect.

The goal now is to calculate the stocks' statistics from the Lorenz. The last value on the curve is the stock's mean return. We obtain the Gini as the area under a virtual LSA and the Lorenz for each stock. Alternatively for the Gini, we can use Equation (7) and substitute the value i/N for the cumulative probability p. Exhibit 3 reports the statistics for the 30 DJIA stocks. The risk-adjusted mean return for each stock is shown as the mean minus the Gini. Two CVaRs are also exhibited: one for 5% and the other for 10%. We calculate CVaRs by using Equation (10) or extrapolate them from the Lorenz using Equation (11). Note that the CVaR at 5% is greater than the CVaR at 10%. We can also apply the mean-Gini conditions for SSD, as expressed by the conditions of Equation (8). This is done on Exhibit 4, where the stocks are ranked first according to the mean and then according to the mean, less the Gini. The list shows the most desirable stocks, ranked according to the necessary conditions for SSD. Hence, the top stocks on the list have higher means and higher risk-adjusted means. As such, the list provides a complete ordering of stock choices, by weighing risk and mean return for all risk-averse investors.

Furthermore, we compare these results with the outcomes shown in Exhibit 4, by ranking CVaR from lower (safest stocks) to higher (riskier stocks). As the table shows, there is some correspondence between the Lorenz and the mean-Gini conditions. However, the comparison

EXHIBIT 3
Statistics of Daily Returns for the 30 DJIA Stocks (2012)

Symbol	Mean	Gini	Mean-Gini	CVaR(5%)	CVaR(10%)
AA	0.022%	0.990%	–0.968%	3.682%	3.052%
AXP	0.093%	0.712%	–0.620%	2.799%	2.268%
BA	0.027%	0.626%	–0.599%	2.592%	1.977%
BAC	0.327%	1.372%	–1.045%	4.528%	3.693%
CAT	0.020%	0.922%	–0.902%	3.603%	2.952%
CSCO	0.057%	0.807%	–0.749%	3.620%	2.633%
CVX	0.026%	0.615%	–0.589%	2.720%	2.083%
DD	0.015%	0.655%	–0.640%	2.778%	2.175%
DIS	0.126%	0.617%	–0.491%	2.438%	1.869%
GE	0.084%	0.651%	–0.567%	2.556%	2.001%
HD	0.170%	0.647%	–0.476%	2.595%	2.003%
HPQ	–0.199%	1.157%	–1.357%	6.113%	4.457%
IBM	0.028%	0.531%	–0.502%	2.373%	1.759%
INTC	–0.042%	0.743%	–0.785%	3.194%	2.555%
JNJ	0.043%	0.330%	–0.287%	1.236%	1.025%
JPM	0.140%	0.958%	–0.819%	4.071%	3.104%
KO	0.029%	0.455%	–0.427%	1.682%	1.383%
MCD	–0.035%	0.469%	–0.504%	2.460%	1.811%
MMM	0.066%	0.492%	–0.426%	2.038%	1.528%
MRK	0.054%	0.528%	–0.474%	2.119%	1.678%
MSFT	0.031%	0.710%	–0.679%	2.663%	2.116%
PFE	0.078%	0.456%	–0.378%	1.567%	1.270%
PG	0.024%	0.427%	–0.403%	1.808%	1.373%
T	0.068%	0.473%	–0.405%	1.982%	1.498%
TRV	0.094%	0.550%	–0.456%	2.135%	1.634%
UNH	0.042%	0.732%	–0.690%	3.123%	2.429%
UTX	0.064%	0.683%	–0.619%	2.313%	1.983%
VZ	0.054%	0.494%	–0.441%	1.858%	1.500%
WMT	0.068%	0.532%	–0.464%	2.662%	1.848%
XOM	0.023%	0.514%	–0.491%	2.118%	1.649%

EXHIBIT 4
Stocks Ranked w.r.t. Mean-Gini and w.r.t. CVaR

	Mean	Ranked Mean–Gini	Ranked CVaR(5%)		Ranked CVaR(10%)	
BAC	0.327%	–1.045%	JNJ	1.236%	JNJ	1.025%
HD	0.170%	–0.476%	PFE	1.567%	PFE	1.270%
JPM	0.140%	–0.819%	KO	1.682%	PG	1.373%
DIS	0.126%	–0.491%	PG	1.808%	KO	1.383%
TRV	0.094%	–0.456%	VZ	1.858%	T	1.498%
AXP	0.093%	–0.620%	T	1.982%	VZ	1.500%
GE	0.084%	–0.567%	MMM	2.038%	MMM	1.528%
PFE	0.078%	–0.378%	XOM	2.118%	TRV	1.634%
T	0.068%	–0.405%	MRK	2.119%	XOM	1.649%
WMT	0.068%	–0.464%	TRV	2.135%	MRK	1.678%
MMM	0.066%	–0.426%	UTX	2.313%	IBM	1.759%
UTX	0.064%	–0.619%	IBM	2.373%	MCD	1.811%
CSCO	0.057%	–0.749%	DIS	2.438%	WMT	1.848%
MRK	0.054%	–0.474%	MCD	2.460%	DIS	1.869%
VZ	0.054%	–0.441%	GE	2.556%	BA	1.977%
JNJ	0.043%	–0.287%	BA	2.592%	UTX	1.983%
UNH	0.042%	–0.690%	HD	2.595%	GE	2.001%
MSFT	0.031%	–0.679%	WMT	2.662%	HD	2.003%
KO	0.029%	–0.427%	MSFT	2.663%	CVX	2.083%
IBM	0.028%	–0.502%	CVX	2.720%	MSFT	2.116%
BA	0.027%	–0.599%	DD	2.778%	DD	2.175%
CVX	0.026%	–0.589%	AXP	2.799%	AXP	2.268%
PG	0.024%	–0.403%	UNH	3.123%	UNH	2.429%
XOM	0.023%	–0.491%	INTC	3.194%	INTC	2.555%
AA	0.022%	–0.968%	CAT	3.603%	CSCO	2.633%
CAT	0.020%	–0.902%	CSCO	3.620%	CAT	2.952%
DD	0.015%	–0.640%	AA	3.682%	AA	3.052%
MCD	–0.035%	–0.504%	JPM	4.071%	JPM	3.104%
INTC	–0.042%	–0.785%	BAC	4.528%	BAC	3.693%
HPQ	–0.199%	–1.357%	HPQ	6.113%	HPQ	4.457%

is not complete, because CVaR considers only low-return risks at a given probability, whereas Lorenz statistics consider risk for the entire distribution of returns and therefore provide much more information about risk and mean return.

CONCLUSION

This article has shown how the Lorenz can serve as a basic tool to measure the risk and return of individual assets and portfolios. Not only does the Lorenz comply with SSD, it also facilitates computation of the mean-Gini conditions for SSD when Lorenz curves intersect. Furthermore, the Lorenz allows for calculating the CVaRs for all probabilities of occurrence. Hence, stocks and portfolios can be ranked in terms of risk and return by using only the Lorenz curves, without estimating probability functions.

ENDNOTES

[1]By using a continuous cumulative distribution function $F(x)$ and its inverse, Gastwirth [1971] expressed the Lorenz with a single equation:

$$L(p) = \int_0^p F^{-1}(t)\,dt \quad \text{for } 0 \le p \le 1$$

[2]Yitzhaki [1982] also showed that the mean-Gini conditions for SSD are sufficient whenever cumulative probability distributions functions intersect once at most.

REFERENCES

Acerbi, C., and D. Tasche. "On the Coherence of Expected Shortfall." *Journal of Banking & Finance*, 26 (2002), pp 1487-1503.

Artzner, P., F. Delbaen, J. Eber, and D. Heath. "Coherent Measures of Risk." *Mathematical Finance*, 9 (1999), pp. 203-228.

Fisher, L., and J. Lorie. "Some Studies of Variability of Returns on Investments in Common Stocks." *Journal of Business*, 43 (1970), pp. 99-135.

Gastwirth, J. "A General Definition of the Lorenz Curve." *Econometrica,* 39 (1971), pp. 1037-1039.

Gini, C. "Variabilità e Mutabilità." Studi Economico-Giuridici, Università di Cagliari, tipografia di P. Cuppini, Bologna (1912).

Hadar, J., and W. Russell. "Rules for Ordering Uncertain Prospects." *American Economic Review*, 59 (1969), pp. 25-34.

Hanoch, G., and H. Levy. "The Efficiency Analysis of Choice Involving Risk." *Review of Economic Studies*, 36 (1969), pp. 335-346.

Levy, H. *Stochastic Dominance: Investment Decision Making under Uncertainty*, 2nd edition. New York: Springer Verlag, 2006.

Lorenz, M. "Methods of Measuring the Concentration of Wealth." *Publications of the American Statistical Association*, 9 (1905), pp. 209-219.

Rockafellar, R., and S. Uryasev. "Optimization of Conditional Value-at-Risk." *Journal of Risk*, 2 (2000), pp. 21-41.

Rothschild, M., and J. Stiglitz. "Increasing Risk I: A Definition." *Journal of Economic Theory*, 2 (1970), pp. 66-84.

Shalit, H., and S. Yitzhaki. "Mean-Gini, Portfolio Theory, and the Pricing of Risky Assets." *Journal of Finance*, 39 (1984), pp. 1449-1468.

Shorrocks, A. "Ranking Income Distributions." *Economica*, 50 (1983), pp. 3-17.

Yitzhaki, S. "Stochastic Dominance, Mean-Variance, and Gini's Mean Difference." *American Economic Review*, 72 (1982), pp. 178-185.

——. "More Than a Dozen Alternative Ways of Spelling Gini." *Research on Economic Inequality*, 8 (1998), pp. 13-30.