3-2009

Stochastic Volatility Models and Financial Risk Measures: Towards New Justifications

Hung T. Nguyen

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Songsak Sriboonchitta

Comments:
Technical Report: UTEP-CS-09-11
Published in the *Proceedings of the 2009 Singapore Economic Review Conference*, Singapore, August 6-8, 2009.

Stochastic Volatility Models and Financial Risk Measures: Towards New Justifications

Hung T. Nguyen[1,2], Vladik Kreinovich[3,2]
and Songsak Sriboonchitta[2],

[1]New Mexico State University
Las Cruces, NM 88003, USA, hunguyen@nmsu.edu
[2]Faculty of Economics, Chiang Mai University
Chiang Mai 50200 Thailand, songsak@econ.cmu.ac.th
[3]University of Texas at El Paso
El Paso, TX 79968, USA, vladik@utep.edu

Abstract

We provide theoretical justifications for the empirical successes of (1) the asymmetric heteroskedasticity models of stochastic volatility in mathematical finance and (2) Wang's distorted probability risk measures in actuarial and investment sciences, using a unified framework of symmetry groups.

Keywords: stochastic volatility, ARCH, GARCH, risk measures, distorted probability, symmetry groups.

1 Introduction

Stochastic volatility and risk management: important economic problems. Time-varying stochastic volatility is an important feature of financial markets, a feature that drastically affects the outcomes of different investments. It is therefore important to take stochastic volatility into account in financial economics and mathematical finance.

There are two related aspects of financial economics and mathematical finance in which stochastic volatility must be taken into account.

First, we must be able to *predict* stochastic volatility based on the past behavior of the corresponding financial instruments (and of the markets in general). Several *stochastic volatility models* have been developed for such prediction; see, e.g., [6, 11, 14, 28]. These models take into account the time-varying character of the corresponding stochastic processes and are therefore called heteroskedasticity models.

Second, we must be able to gauge the corresponding *risk* associated with different financial instruments. Several *risk measures* have been proposed in actuarial and investment sciences to describe and manage this risk. Among the most widely used measures are "distorted probability" measures developed originally by S. S. Wang.

Many existing stochastic volatility models and risk measures are semi-heuristic: practical need for justification. Many existing models and measures are semi-empirical, semi-heuristic. The semi-empirical nature of these models and measures means that they have been selected largely based on the past behavior of the financial markets. However, the financial markets change. It is therefore necessary to decide which features of these models and measures are still applicable in the changed markets and which need to be modified. To be able to make this decision, it is necessary to analyze:

- which features of the existing models and measures follow from the general economic principles; these features can be clearly applied to the changing markets as well, and
- which features only follow from the past data; these features may need to be modified when the financial markets undergo a drastic change.

In other words, to decide which features of the existing models and measures are applicable in the changing markets, we must try to find theoretical *justifications* for these models and measures.

What we do in this paper. In this paper, we provide theoretical justifications of heteroskedasticity models of stochastic volatility and of Wang's risk measures. In our justification, we use the idea of symmetry, an idea that appears very naturally in the context of economics.

To exploit this idea, we use the technique of *symmetry groups*, a mathematical techniques which was developed to study symmetries. To be able to apply symmetry groups to economics, we also use notions from decision theory, and then use known mathematical techniques to solve the resulting functional equations.

The paper consists of two main parts. In the first part, we develop symmetry-based justification of stochastic volatility models. In the second part, we describe a symmetry-based justification of the related risk measures.

Most of our results are new; other results further develop ideas outlined in our previous publications [19, 27].

This paper may also be of interest to physicists. While the main purpose of this paper is to provide rationale for some main successful proposed stochastic volatility models and risk measure models in economics, it could be interesting for physicists as well, as the technique of symmetry group that we used is quite familiar to them; see, e.g., [8].

Starting from the quark theory which was originally formulated in terms of an appropriate symmetry groups, symmetry groups have been one of the main research tools of modern physics. The corresponding ideas of invariance are behind Einstein's special and general relativity theory, symmetries distinguish liquids from gases from solids, symmetries and supersymmetries abound in quantum physics.

The successes of this idea in physics has led to many successful applications of the invariance idea in other disciplines, including engineering and computer science; see, e.g., [22]. In this paper, we show that similar techniques can be also very useful in econometrics: namely, they help to justify (and thus, to better understand) many successful heuristical methods.

Because of our desire to attract the attention of physicists, in each part, we first describe the corresponding models before explaining how we can justify these models.

2 Part 1: Asymmetric Heteroskedasticity Models of Stochastic Volatility

Let us start with a brief introduction to the existing models for describing stochastic volatility.

One of the main goals of econometrics. One of the main objectives of econometrics is to use the known values x_t, x_{t-1}, x_{t-2}, ..., of different economic characteristics x at different moments of time t, $t-1$, $t-2$, ..., to predict the future values x_{t+1}, x_{t+2}, ..., of these characteristics.

First approximation: engineering models. A similar problem of analyzing time series x_t exists in engineering applications. So, historically the first econometric models simply used the formulas developed in engineering applications.

In engineering, most processes are stationary. It is known that stationary processes x_t can be well-described by auto-regression (AR) models:

$$x_t = a_0 + \sum_{i=1}^{q} a_i \cdot x_{t-i} + \varepsilon_t, \tag{1}$$

where ε_t are independent normally distributed random variables with 0 means and standard deviation σ – i.e., $\varepsilon_t = \sigma \cdot z_t$, where z_t is normally distributed with 0 means and standard deviation 1. The more terms we take in the AR model, i.e., the larger the value q, the better the corresponding AR(q) model describes the stationary process.

Heteroskedisticity (non-stationarity): a specific feature of econometric time series. In contrast to engineering time series, economic time series are usually non-stationary (*heteroskedastic*).

Specifically, in the economic time series, the empirical standard deviation σ of the remainder term ε_t depends on time. In other words, instead of a single value σ, at different moments of time t, we have different values σ_t. Thus, to appropriately describe the corresponding time series, we also need to know how this value σ_t changes with time.

Second approximation: models that take heteroskedisticity into account. The heteroskedisticity phenomenon was first taken into account by Engle [7] who proposed a linear regression model of the dependence of σ_t on the previous deviations:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \cdot \varepsilon_{t-i}^2. \tag{2}$$

This model is known as the Autoregression Conditional Heteroskedacity model, or ARCH(q), for short.

An even more accurate Generalized Autoregression Conditional Heteroskedacity model GARCH(p,q) was proposed in [3]. In this model, the new value σ_t^2 of the variance is determined not only by the previous values of the squared differences, but also by the previous values of the variance:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \cdot \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \cdot \sigma_{t-i}^2. \tag{3}$$

Several modifications of these models have been proposed. For example, Zakoian [33] proposed to use regression to predict the standard deviation instead of the variance:

$$\sigma_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \cdot |\varepsilon_{t-i}| + \sum_{i=1}^{p} \beta_i \cdot \sigma_{t-i}. \tag{4}$$

Nelson [21] proposed to take into account that the values of the variance must always be non-negative – while in most existing autoregression models, it is potentially possible to get negative predictions for σ_t^2. To avoid negative predictions, Nelson considers the regression for $\log \sigma_t^2$ instead of for σ_t:

$$\log \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \cdot |\varepsilon_{t-i}| + \sum_{i=1}^{p} \beta_i \cdot \log \sigma_{t-i}^2. \tag{5}$$

Asymmetry: an additional feature of economic time series that needs to be taken into account. The above models such as ARCH(q) and GARCH(p,q) models are still not always fully adequate in describing the actual econometric time series. One of the main reasons for this fact is that these models do not take into account a clear *asymmetry* between the effects of positive shocks $\varepsilon_t > 0$ and negative shocks $\varepsilon_t < 0$.

It is therefore desirable to modify the ARCH(q) and GARCH(p,q) models by taking asymmetry into account.

Models that take asymmetry into account. Several modifications of the ARCH(q) and GARCH(p,q) models have been proposed to take asymmetry into account.

For example, Glosten et al. [13] proposed the following modification of the GARCH(p,q) model:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} (\alpha_i + \gamma_i \cdot I(\varepsilon_{t-i})) \cdot \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \cdot \sigma_{t-i}^2, \tag{6}$$

where $I(\varepsilon) = 0$ fir $\varepsilon \geq 0$ and $I(\varepsilon) = 1$ for $\varepsilon < 0$.

Similar modifications were proposed by Zakoian [33] and Nelson [21] for their models. The asymmetric version of Zakoian's model has the form

$$\sigma_t = \alpha_0 + \sum_{i=1}^{q} (\alpha_i^+ \cdot \varepsilon_{t-i}^+ + \alpha_i^- \cdot \varepsilon_{t-i}^-) + \sum_{i=1}^{p} \beta_i \cdot \sigma_{t-i}, \tag{7}$$

where:

- $\varepsilon^+ = \varepsilon$ for $\varepsilon > 0$ and $\varepsilon^+ = 0$ for $\varepsilon \leq 0$;
- $\varepsilon^- = \varepsilon$ for $\varepsilon < 0$ and $\varepsilon^- = 0$ for $\varepsilon \geq 0$.

The asymmetric version of Nelson's model has the form

$$\log \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} (\alpha_i \cdot |\varepsilon_{t-i}| + \gamma_i \cdot \varepsilon_{t-i}) + \sum_{i=1}^{p} \beta_i \cdot \log \sigma_{t-i}^2. \tag{8}$$

Asymmetric heteroskedactic models: a general description. All the above formulas can be viewed as particular cases of the following general scheme:

$$\sigma_t = f\left(\alpha_0 + \sum_{i=1}^{q} \alpha_i \cdot f_i(\varepsilon_{t-i}) + \sum_{i=1}^{p} \beta_i \cdot g_i(\sigma_{t-i})\right). \tag{9}$$

For example:

- ARCH(q) and GARCH(p,q) correspond to using $f(x) = \sqrt{x}$ and $f_i(x) = g_i(x) = x^2$.
- Glosten's formula corresponds to using $f_i(x) = (1 + (\gamma_i/\alpha_i) \cdot I(x)) \cdot x^2$.
- A symmetric version of the Nelson's formula corresponds to $f(x) = \sqrt{\exp(x)}$, $f_i(x) = |x|$, and $g_i(x) = \log x^2$; the general (asymmetric) version of this formula corresponds to $f_i(x) = |x| + (\gamma_i/\alpha_i) \cdot x$.

Problem: the existing asymmetric modifications are very heuristic. The main problem with the existing asymmetric models is that they are very *ad hoc*. These model are obtained by simply replacing a symmetric expression ε^2 or $|\varepsilon|$ by an asymmetric one, without explaining why these specific asymmetric expressions have been selected.

From the purely mathematical viewpoint, one can envision many other different functions $f(x)$, $f_i(x)$, and $g_i(x)$. The asymmetric models – corresponding to very specific selections of these functions – have worked well in describing the past econometric time series. To find out to what extent these models can be applicable to other situations, when the financial markets may have changed, it is desirable to analyze which features of these models follows from the general principles.

Such an analysis is proposed in this section. As a result, we justify the specific form of the dependence on ε_{t-i} by the scale-invariance requirement.

Scale invariance: a natural feature of econometric descriptions. The numerical value of each economic variable depends on the choice of a measuring unit. If we choose another measuring unit, the situation remains the same, but the numerical values change.

For example, for a US investor, it is natural to describe all the prices in dollars. For a European investor, it is equally natural to translate all the prices into Euros. If we replace the original unit with a new unit which is λ times smaller, then all numerical values need to be multiplied by λ.

The models should not change if we simply change the units. When we replace the original values ε_{t-i} by new numerical values $\varepsilon_{t-i} = \lambda \cdot \varepsilon_{t-i}$ of the same quantity, then each corresponding term $f_i(\varepsilon_i)$ is replaced with a new term $f_i(\varepsilon'_i) = f_i(\lambda \cdot \varepsilon_i)$. Thus, the overall contribution of all these terms changes from the original value $I = \sum\limits_{i=1}^{q} \alpha_i \cdot f_i(\varepsilon_{t-i})$ to the new value $I' = \sum\limits_{i=1}^{q} \alpha_i \cdot f_i(\lambda \cdot \varepsilon_{t-i})$.

It is reasonable to require that the relative quantity of different contributions does not change, i.e., that if two different sets $x_i^{(1)} \stackrel{\text{def}}{=} \varepsilon_{t-i}^{(1)}$ and $x_i^{(2)} \stackrel{\text{def}}{=} \varepsilon_{t-i}^{(2)}$ lead to the same contributions $I^{(1)} = I^{(2)}$, then after re-scaling, they should also lead to the same contributions I'. Thus, we arrive at the following condition:

Scale invariance: precise formulation of the requirement. Let the values $\alpha_1, \ldots, \alpha_q$ be fixed. Then, the functions $f_1(x), \ldots, f_p(x)$ should satisfy the following condition: if for two sets $x_1^{(1)}, \ldots, x_q^{(1)}$ and $x_1^{(2)}, \ldots, x_q^{(2)}$, we have

$$\sum_{i=1}^{q} \alpha_i \cdot f_i(x_i^{(1)}) = \sum_{i=1}^{q} \alpha_i \cdot f_i(x_i^{(2)}), \tag{10}$$

then for every $\lambda > 0$, we must have

$$\sum_{i=1}^{q} \alpha_i \cdot f_i(\lambda \cdot x_i^{(1)}) = \sum_{i=1}^{q} \alpha_i \cdot f_i(\lambda \cdot x_i^{(2)}). \tag{11}$$

Analysis of the problem: from scale invariance to the functional equation. For simplicity, let us start with the case when the values $x_i^{(2)}$ are very close to $x_i^{(1)}$, i.e., when $x_i^{(2)} = x_i^{(1)} + k_i \cdot h$ for some constants k_i and for a very small real number h. For small h, we have

$$f_i(x_i^{(1)} + k_i \cdot h) = f_i(x_i^{(1)}) + f_i'(x_i^{(1)}) \cdot k_i \cdot h + O(h^2). \tag{12}$$

Substituting the expression (12) into the formula (10), we conclude that

$$\sum_{i=1}^{q} \alpha_i \cdot f_i'(x_i^{(1)}) \cdot k_i \cdot h + O(h^2) = 0. \tag{13}$$

Dividing both sides by h, we get

$$\sum_{i=1}^{q} \alpha_i \cdot f_i'(x_i^{(1)}) \cdot k_i + O(h) = 0. \tag{14}$$

Similarly, the condition (11) leads to

$$\sum_{i=1}^{q} \alpha_i \cdot f_i'(\lambda \cdot x_i^{(1)}) \cdot k_i + O(h) = 0. \tag{15}$$

In general, the condition (14) lead to (15). In the limit $h \to 0$, we therefore conclude that for every vector $k = (k_1, \ldots, k_q)$, if

$$\sum_{i=1}^{q} k_i \cdot (\alpha_i \cdot f_i'(x_i^{(1)})) = 0, \tag{16}$$

then

$$\sum_{i=1}^{q} k_i \cdot (\alpha_i \cdot f_i'(\lambda \cdot x_i^{(1)})) = 0. \tag{17}$$

Functional equation: geometric analysis. The sum (16) is a scalar (dot) product between the vector k and the vector a with components $\alpha_i \cdot f_i'(x_i^{(1)})$. Similarly, the sum (17) is a scalar (dot) product between the vector k and the vector b with components $\alpha_i \cdot f_i'(\lambda \cdot x_i^{(1)})$. Thus, the above implication means that the vector b is orthogonal to every vector k which is orthogonal to a, i.e., to all vectors k from the hyperplane consisting of all the vectors orthogonal to a.

It is easy to see geometrically that the only vectors which are orthogonal to the hyperplane are vectors collinear with a. Thus, we conclude that $b = \delta \cdot a$ for some constant δ, i.e., that

$$\alpha_i \cdot f_i'(\lambda \cdot x_i^{(1)}) = \delta \cdot \alpha_i \cdot f_i'(x_i^{(1)}). \tag{18}$$

Dividing both sides by α_i, we conclude that

$$f_i'(\lambda \cdot x_i^{(1)}) = \delta \cdot f_i'(x_i^{(1)}). \tag{19}$$

Analysis of dependence and the resulting new differential equation. In principle, δ depends on λ and on values $x_i^{(1)}$. From the equation (19) corresponding to $i = 1$, we see that

$$\delta = \frac{f_1'(\lambda \cdot x_1^{(1)})}{f_i'(x_1^{(1)})}. \tag{20}$$

Thus, δ only depends on $x_1^{(1)}$ and does not depend on any other value $x_i^{(1)}$. Similarly, by considering the case $i = 2$, we conclude that δ can depend only on $x_2^{(1)}$ and thus, does not depend on $x_1^{(1)}$ either. Thus, δ only depends on λ, i.e., the condition (19) takes the form

$$f_i'(\lambda \cdot x_i^{(1)}) = \delta(\lambda) \cdot f_i'(x_i^{(1)}). \tag{21}$$

From the solution to the new differential equation to the solution of our original problem. It is known that every continuous function $f_i'(x)$ satisfying the equation (19) has the following form:

- $f_i'(x) = C_i^+ \cdot x^{a_i}$ for $x > 0$, and
- $f_i'(x) = C_i^- \cdot |x|^{a_i}$ for $x < 0$,

for some values $C_i^\pm$ and a_i; see, e.g., [1], Section 3.1.1, or [22]. (This result was first proven in [24].) For differentiable functions, the easiest way to prove this result is to differentiate both sides of (19) by λ, set $\lambda = 1$, and solve the resulting differential equation.

For the corresponding functions, the condition (21) is satisfied with $\delta(\lambda) = \lambda^{a_i}$. Since the function $\delta(\lambda)$ is the same for all i, the value a_i is therefore also the same for all i: $a_1 = \ldots = a_q$. Let us denote the joint value of all these a_i by a.

Thus, all the derivatives have $f_i'(x)$ are proportional to x^a. Hence, the original functions are proportional

- either to x^{a+1} (for $a \neq -1$)
- or to $\log(x)$ (when $a = -1$).

The additive integration constant can be absorbed into the additive constant α_0, and the multiplicative constants can be absorbed into a factor α_i.

Thus, without losing generality, we can conclude that in the scale invariant case, either $f_i(x) = x_i^a \cdot (1 + b \cdot I(x))$, or $f_i(x) = \log(|x|)$.

Conclusion. We have proven that the natural scale-invariance condition implies that each function $f_i(x)$ has either the form $\log(x)$, or the form $f_i(x) = x_i^a \cdot (1 + b \cdot I(x))$. This conclusion covers all the functions which are efficiently used to describe asymmetric heteroskedasticity:

- the function $f_i(x) = 1 + (\gamma_i/\alpha_i) \cdot I(x)) \cdot x^2$ used in Glosten's model;

- the function $f_i(x) = x^+ + (\alpha^-/\alpha^+) \cdot x^-$ used in Zakoian's model; and
- the function $f_i(x) = (1 + (\gamma_i/\alpha_i) \cdot \text{sign}(x)) \cdot |x|$ used in Nelson's model.

It is worth mentioning that this result also covers the functions $g_i(x) = x^2$ and $g_i(x) = \log(x^2) = 2\log(x)$ used to describe the dependence on σ_{t-i}.

Thus, the exact form of the dependence on ε_{t-i} has indeed been justified by the natural scale invariance requirement – as well as the dependence on σ_{t-i}.

Comment. It is worth mentioning that scale-invariance of the econometric formulas describing heteroskedasticity was noticed and actively used in [5]. However, our approaches are somewhat different:

- In [5], the econometric *formulas* were taken as *given*, and scale invariance was used to analyze heteroskedasticity tests.
- In contrast, we use scale invariance to *derive* the econometric *formulas*.

3 Part 2: Wang Transform Operators in Financial Risk Analysis

Pricing under risk: a problem. The price of most financial instruments unpredictably (randomly) fluctuates. As a result of these unpredictable fluctuations, investing in a financial instrument is risky: there is a always a possibility that the price of this particular instrument will go down, and as a result, the investors will suffer losses.

In many practical situations, based on the prior performance of a financial instrument (and on the additional information that we may have), we can get a pretty good understanding of the probabilities of different future prices. In other words, we usually know the probability distribution of the future price X. This distribution can be described, e.g., by describing the probabilities that the future price does not go below a certain threshold, i.e., by the *decumulative distribution function*

$$S(x) \stackrel{\text{def}}{=} \text{Prob}(X > x). \tag{22}$$

Alternatively, we can use the cumulative distribution function

$$F(x) \stackrel{\text{def}}{=} \text{Prob}(X \le x) = 1 - S(x). \tag{23}$$

Once we know this probability distribution, what is a reasonable price of this financial instrument?

First approximation to pricing under risk: a straightforward application of the traditional probability approach. A seemingly natural approach to pricing can be obtained from a straightforward application of the traditional probability approach; see, e.g., [26]. If we invest in similar financial instruments, with a similar probability distribution, again and again, then on average, due to the Law of Large Numbers, our average gain per investment

$$\frac{X_1 + \ldots + X_n}{n} \tag{24}$$

will be equal to the mean (expected) value $E[X]$ of the corresponding random variable X. Thus, from the traditional probability approach, it seems reasonable to use this mean value as the fair price of the financial instrument.

In terms of the decumulative distribution function $S(x)$, this mean $E[X]$ has the form

$$E[X] = \int_L^\infty \text{Prob}(X > x)\, dx + L = \int_L^\infty S(x)\, dx + L, \tag{25}$$

where L is the smallest possible price. Usually, we take $L = 0$, so the formula (25) takes the simplified form

$$E[X] = \int_0^\infty \text{Prob}(X > x)\, dx = \int_0^\infty S(x)\, dx. \tag{26}$$

Comment. Strictly speaking, according to decision theory [9, 10, 16, 20, 25], we must take the mean not of the future price itself, but of the *utility* corresponding to the future price. This difference is important because people's utilities are non-linearly related to their gains and losses. So, strictly speaking, instead of the distribution of the prices, we should start with the corresponding distribution of utilities. For simplicity, in the future, we will simply talk about future prices (gains, etc.), but all our considerations are applicable to utilities as well.

Need to go beyond the first approximation. According to the above straightforward application of the traditional probability approach, the fair price of the financial instrument should be equal to the expected value of the future price. In real life, however, most people prefer to get, e.g., a \$1 to a situation in which, with probability 1/2, we get \$2 and with probability 1/2, we get 0.

One may claim that this preference is caused by the non-linearity of the corresponding utility function, but the same preference can be observed if we use utils (units of utility) instead of dollars:

- On the one hand, theoretically, getting 1 util with certainty should be equivalent to getting 2 units with probability 1/2 and 0 utils with probability 1/2.
- However, in practice, in this choice, most people prefer 1 util with certainty.

In other words, risk decreases the fair price of a financial instrument: the more risk, the smaller the price.

For insurance instruments, a similar effect of risk moves the price in the opposite direction: we need to pay extra insurance premium for risk: the larger the risk, the larger this premium and thus, the larger the resulting price.

How can we describe the resulting fair pricing under risk?

Wang transform: an efficient empirical method for pricing under risk. In his papers [30, 31], S. S. Wang proposed to replace the expression (26) with a modified formula

$$\int_0^\infty g(S(x))\,dx, \tag{27}$$

where the *Wang transform* $g(y)$ is defined by the formula

$$g(y) = \Phi(\Phi^{-1}(y) + \alpha), \tag{28}$$

in which α is a constant, and $\Phi(y)$ is a cumulative distribution function of some appropriate distribution – e.g., of the standard normal distribution (with 0 mean and standard deviation 1).

For financial instruments, when the presence of risk decreases the price, we take $\alpha \leq 0$. For estimating insurance premiums, when the presence of risk increases the price, we should take $\alpha \geq 0$.

This method turned out to be highly efficient in practice. As a result, this method (and its modifications and extensions) is now one of the main methods for pricing under risk; see, e.g., [12].

Difference between subjective and objective probabilities. A partial explanation of the Wang's formula comes from the difference between subjective and objective probabilities. Indeed, how can we explain the fact that, in spite of the seeming naturalness of the mean, people's pricing under risk is usually different from the mean $E[X]$?

A natural explanation comes from the observation that the mean $E[X]$ is based on the "objective" probabilities (frequencies) of different events, while people use "subjective" estimates of these probabilities when making decisions. For a long time, researchers thought that subjective probabilities are approximately equal to the objective ones, and often they are equal. However, in 1979, a classical paper by D. Kahneman and A. Tversky ([15], reproduced in [29]) showed that our "subjective" probability of different events is, in general, different from the actual ("objective") probabilities (frequencies) of these events. This difference is especially drastic for events with low probabilities and probabilities close to 1 – i.e., for the events that are most important when estimating risk.

Specifically, Kahnemann and Tversky has shown that there is a one-to-one correspondence $g(y)$ between objective and subjective probabilities, so that once we know the (objective) probability $\mathrm{Prob}(E)$ of an event E, the subjective probability $\mathrm{Prob}_{\mathrm{subj}}(E)$ of this event E is (approximately) equal to

$$\mathrm{Prob}_{\mathrm{subj}}(E) = g(\mathrm{Prob}(E)). \tag{29}$$

This idea was further explored by other researchers; see, e.g., [2] and references therein.

Resulting justification of the Wang-type formula. Let us apply the Kahneman-Tversky idea to the events $X > x$ corresponding to different values x. For such an event E, its objective probability is equal to the decumulative distribution function $S(x)$. Thus, the corresponding subjective probability of this event is equal to

$$\text{Prob}_{\text{subj}}(X > x) = g(S(x)). \tag{30}$$

If we compute the "subjective" mean $E_{\text{subj}}[X]$ of X based on these subjective probabilities, we get

$$E_{\text{subj}}[X] = \int_0^\infty \text{Prob}_{\text{subj}}(X > x)\, dx = \int_0^\infty g(S(x))\, dx. \tag{31}$$

This is exactly the expression (27) for pricing under risk.

Remaining problem. From the purely mathematical viewpoint, we can use different functions $g(y)$, not necessarily functions of Wang's type. As of now, empirically, Wang's formula is among of the most efficient ones.

A natural question is: which features of the Wang's formula follow from the general principles – and can therefore be applied when the financial markets change – and which features may reflect the specifics of the past financial markets.

The existing justifications of Wang's measures. There exist several justifications of the Wang transform method, justifications that show that this method is uniquely determined by some reasonable properties. The first justifications was proposed by Wang himself, in [32]; several other justifications are described in [17].

The main limitations of the existing justifications of the Wang transform is that these justifications are too complicated, too mathematical, and not very intuitive – while our intent is to find justifications based on the economics-related first principles.

What we do in this section. In this section, we provide a new justifications of the Wang transform based on the notion of symmetry, a notion which naturally appears in the economics context.

Our new justification: main idea. Our first justification comes from the fact that the Wang transform is not a single function: we have a family of transforms $g_\alpha(y)$ corresponding to different values α. Different values of α correspond to different degree of dependence on risk.

These transforms form a *transformation group* in the sense that the following three conditions are satisfied:

- First, the identity transform $g(y) = y$ is a particular case of the Wang transform: specifically, it corresponds to $\alpha = 0$.

- Second, the inverse operation to a Wang transform $g_\alpha(y)$ is also a Wang transform: $g_\alpha^{-1}(y) = g_\beta(y)$ for some β in the sense that

$$g_\beta(g_\alpha(y)) = g_\beta(g_\alpha(y)) = y \tag{32}$$

for all y; specifically, this is true for $\beta = -\alpha$.

- Third, the composition of Wang transforms is also a Wang transform: if we first apply a transform $g_\alpha(y)$, and then a transform $g_\beta(y)$, then this is equivalent to applying the Wang transform with a single value γ:

$$g_\beta(g_\alpha(y)) = g_\gamma(y) \tag{33}$$

for all y; specifically, this is true for $\gamma = \alpha + \beta$.

It is reasonable to require that we have a family of transformation that forms a 1-parametric group. It is also reasonable to require that the transformations are continuous, and that the dependence on the parameter is also continuous, i.e., that this is a *Lie group*; see, e.g., [4]. It is known that every 1-dimensional Lie group (i.e., a Lie group described by a single parameter) is (at least locally) isomorphic to the additive group of real numbers. In precise terms, this means that instead of using the original values of the parameter v describing different elements of this group, we can use a re-scaled parameter $\alpha = h(v)$ for some non-linear function h, and in this new parameter scale:

- the identity element corresponds to $\alpha = 0$;
- the inverse element to an element with parameter α is an element with parameter $-\alpha$: $g_{-\alpha} = g_\alpha^{-1}$; and
- the composition of elements with parameters α and β is an element with the parameter $\alpha + \beta$: $g_\alpha \circ g_\beta = g_{\alpha+\beta}$.

Let us denote $\Phi(x) \stackrel{\text{def}}{=} g_x(0.5)$. Then, for every $y = \Phi(x)$ and for every real number α, we get

$$g_\alpha(y) = g_\alpha(\Phi(x)) = g_\alpha(g_x(0.5)) = (g_\alpha \circ g_x)(0.5). \tag{34}$$

Because of our choice of parameters, we have

$$g_\alpha \circ g_x = g_{x+\alpha} \tag{35}$$

and thus, the formula (34) takes the form

$$g_\alpha(y) = g_{x+\alpha}(0.5). \tag{36}$$

By definition of the function $\Phi(x)$, this means that

$$g_\alpha(y) = \Phi(x + \alpha). \tag{37}$$

Here, $y = \Phi(x)$, hence $x = \Phi^{-1}(y)$ and thus,

$$g_\alpha(y) = \Phi(\Phi^{-1}(y) + \alpha). \tag{38}$$

This is exactly the Wang transform formula (28).

Towards a yet more natural justification. The above justification was based on the assumption that we have a 1-parametric family of transformations, i.e., the family depending on a single parameter α. It is indeed true that in the Wang transform formula (28), there is exactly one parameter, but this 1-parametric character of the family of transforms sounds more like a mathematical assumption that an economically meaningful fact. To get a more natural justification, it is therefore desirable to provide a more intuitive explanation for this 1-parametric character.

Indeed, such an explanation comes from the fact that in many important economic situations, the decision making is done not by a single individual, but rather by a group of individuals, with slightly different objectives and slightly different attitudes to risk.

Different attitude to risk can be represented by different distortion transformations $g(y)$ in the formula (27). In the individual decision making, the decision maker estimates all the related probabilities. Thus, the corresponding distortion transformation is applied to the objective probabilities – and result in the (slightly distorted) subjective probabilities of the decision maker.

In group decision making, especially in a group decision making in a complex situation, it is very difficult for each individual to estimate all the related probabilities. A natural strategy for a group is to divide this complex estimation task between different participants, so that different participants estimate different probabilities. Then, the participants communicate their estimates to one another, and make a decision based on all these estimates.

This group process leads to an additional distortion of probabilities. Namely, first, the participants who is asked to estimate the corresponding probabilities somewhat distorts these probabilities – corresponding to his or her level of risk aversion. Other participants, however, view these probabilities as reasonable approximations to the objective ones – and thus, also somewhat distort these probabilities in accordance with their own degree of risk aversion. Thus, in group decision making, the objective probabilities are distorted twice:

- first, the original estimator e transforms the objective probabilities y into somewhat distorted values $y' = g_e(y)$;
- after that, another decision maker d transforms the values y' into somewhat different subjective probabilities $y'' = g_d(y') = g_d(g_e(y))$.

In other words, the distortion $y'' = g(y)$ corresponding to the interaction between two decision makers is equal to the *composition* of the distortion functions g_d and g_e corresponding to them individually: $g(y) = g_d(g_e(y))$, i.e., using the mathematical notation $\circ$ for the composition, $g = g_d \circ g_e$.

In our main justification, we have already emphasized the importance of the composition, so at first glance, one may think that we do not gain anything new by considering group decision making. However, the group-decision aspect opens new possibilities. For example, it is known that often, group decisions are reasonable stable, even when the roles of different individuals within a group slightly change. In other words, whether the participant e collects the original estimates and passes them to the participant d, or, vice versa, d collects the original estimates and passes them to e, the final decisions are (largely) the same.

In terms of composition, the difference between these two situations is that in the first situation, we have the composition $g_d \circ g_e$, while in the second situation, we have a different composition $g_e \circ g_d$. Thus, the above invariance means that the order of the composition does not matter: $g_d \circ g_e = g_e \circ g_d$. In mathematical terms, we can say that the corresponding transformations *commute*, and the transformation group is thus *commutative.* Now, the following simple mathematical result shows that commutative groups are indeed 1-parametric.

Proposition. *Let X be a set, and let G be a commutative group of transformations $g : X \to X$ with the property that for every two values $x \in X$ and $x' \in X$ there exists a transformation $g_{x,x'} \in G$ that transforms x into x': $g_{x,x'}(x) = x'$. Then, for every $x \in X$ and $x' \in X$, there exists only one transformation $g \in G$ for which $g(x) = x'$.*

Discission. In the economic case, we have a group of transformations $g : (0,1) \to (0,1)$, with $X = (0,1)$.

For every objective probability $x \in (0,1)$ of the positive event, we can have extremely risk-averse individuals for whom the corresponding subjective probabilities $g(x)$ are close to 0. We can also have extremely risk-prone individuals for which the corresponding subjective probability of the positive event is close to 1. Of course, we can also have all the values in between.

Thus, for the group of distortion transformations, it is very reasonable to assume that for every objective probability $x \in (0,1)$, the quantity $g(x)$ can take all possible values from the interval $(0,1)$.

Under this reasonable assumption, the above Proposition implies that for every $x \in (0,1)$, the transformation $g \in G$ is uniquely determined by a single number – the value $g(x)$. Thus, the family of distortion transformations is indeed 1-parametric.

Comment. In mathematics, groups that can transform every element x into every other element are called *transitive*, and groups for which, for every x and x', there exists exactly one transformation mapping x to x', are called *simply transitive.* In these terms, the above Proposition states that every commutative transitive group is simply transitive.

Proof of the Proposition. Let $x_1 \in X$ and $x_2 \in X$, and let f and g be transformations for which $f(x_1) = g(x_1) = x_2$. We need to prove that $f = g$, i.e., that for every $x \in X$, we have $f(x) = g(x)$.

Indeed, let us pick any value $x \in X$, and let us prove that $f(x) = g(x)$. Since the group G is transitive, there exists a transformation h for which $h(x_1) = x$. The group G is also commutative, so we have $f \circ h = h \circ f$; in particular, for

x_1, we have $f(h(x_1)) = h(f(x_1))$. We know that $f(x_1) = x_2$. By our choice of h, we have $h(x_1) = x$. Thus, we conclude that $f(x) = h(x_2)$.

Similarly, we have $g \circ h = h \circ g$, hence $g(h(x_1)) = h(g(x_1))$. We know that $g(x_1) = x_2$. By our choice of h, we have $h(x_1) = x$. Thus, we conclude that $g(x) = h(x_2)$.

Therefore, both $f(x)$ and $g(x)$ are equal to the exact same value $h(x_2)$ and are, hence, equal to each other. The proposition is proven.

Comment. A similar argument can explain why "modifiers" in fuzzy logic, i.e., transformations $g : (0, 1) \to (0, 1)$ that describe how linguistic modifiers like "very", "almost", etc., also usually form a 1-parametric family; see, e.g., [18, 23].

Additional idea. Symmetry groups can be used to justify not only the Wang transform formula (28), they can also explain the original formula (27).

This explanation comes from the fact that in many cases, an investor is most interested in his ability to have enough cash after a certain period of time. For example, a person who invests as part of his retirement plan, would like to have, at the moment of his or her retirement, enough money to spend on the basics during the retirement.

To describe this idea in precise mathematical terms, let us denote the corresponding minimal return by x. Then, what is most important to the investor is the probability that this return will be attained, i.e., the probability $\text{Prob}(X \geq x)$. This is, in effect, the decumulative distribution function $S(x)$. Thus, from this viewpoint, the utility of the investment to each investor can be characterized by the corresponding value $S(x)$.

Different investors have different thresholds x. So, for different investors, the same financial instrument is characterized by the values $S(x)$ corresponding to different values x. We need to somehow combine these utility values into a single numerical criterion – the price describing this particular financial instrument.

How can we combine different utility values? For example, if we have two investors with utilities a and b, what combination $a * b$ of these values should we use? It is reasonable to assume that this combination does not depend on the order in which we combine these utilities, i.e., in mathematical terms, that the corresponding combination function $a * b$ is commutative ($a * b = b * a$) and associative ($a * (b * c) = (a * b) * c$). It is also reasonable to assume that for each "positive" utility value u, there is a corresponding negative value v for which $u * v = 0$. In other words, it is reasonable to require that the set of real numbers (possible utility values) with the operation $*$ form a *group*.

We have already mentioned that every 1-dimensional Lie group is (at least locally) isomorphic to the group of real numbers. This isomorphism means that there exists a mapping $g(y)$ from the original utility scale to the new scale in which, for every two real numbers a and b, we have

$$g(a * b) = g(a) + g(b), \tag{39}$$

and thus,

$$a * b = g^{-1}(g(a) + g(b)). \quad (40)$$

Similarly, for n utilities $u_1, \ldots, u_n$, we have

$$u_1 * \ldots * u_n = g^{-1}(g(u_1) + \ldots + g(u_n)). \quad (41)$$

If we have n investors, with thresholds $x_1, \ldots, x_n$, then their utilities are equal to $u_1 = S(x_1)$, $\ldots$, $u_n = S(x_n)$ and therefore, the combined utility is equal to

$$u_1 * \ldots * u_n = g^{-1}(g(S(x_1) + \ldots + g(S(x_n))). \quad (42)$$

In real life, we have a large number of investors, with different values x_i, so it is reasonable to approximate the sum by the corresponding integral. As a result, the combined utility is equal to $u = g^{-1}(\int g(S(x))\,dx)$. To price the financial instrument, we must compare it with a simple investment X_v that provides a certain value v with guarantee. In other words, we must find the value v for which

$$g^{-1}\left(\int g(S(x))\,dx\right) = g^{-1}\left(\int g(S_v(x))\,dx\right), \quad (43)$$

where S_v is the decumulative distribution function corresponding to the guaranteed-return investment v. By applying the function $g(y)$ to both sides of the equality (43), we conclude that

$$\int g(S(x))\,dx = \int g(S_v(x))\,dx. \quad (44)$$

For the guaranteed investment, for every x, we have either $S_v(x) = 0$ or $S_v(x) = 1$ and thus, $g(S_v(x)) = S_v(x)$ for all x. So, for this investment, the integral $\int g(S_v(x))\,dx$ is thus equal to $\int S_v(x)\,dx$, i.e., to the mean return of this investment – which is exactly v. Thus,

$$\int g(S_v(x))\,dx = v. \quad (45)$$

So, according to the formula (44), the fair price v of an investment is determined by the formula

$$v = \int g(S(x))\,dx. \quad (46)$$

This is exactly the formula (27).

Acknowledgments.

This work was supported in part by NSF grant HRD-0734825 and by Grant 1 T36 GM078000-01 from the National Institutes of Health.

References

[1] J. Aczel, *Lectures on Functional Equations and Their Applications*, Dover Publ., New York, 2006.

[2] J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Springer Verlag, New York, 2001.

[3] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 1986, Vol. 31, pp. 307–327.

[4] C. Chevalley, *Theory of Lie Groups*, Princeton University Press, Princeton, New Jersey, 1999.

[5] J.-M. Dufour, L. Khalaf, J.-T. Bernard, and I. Genest, Simulation-based finite-sample tests for heteroskedasticity and ARCH effects, Journal of Econometrics, 2004, Vol. 122, No. 2, pp. 317–347.

[6] W. Enders, *Applied Econometrics Time Series*, John-Wiley & Sons, 2003.

[7] R. F. Engle, Autoregressive conditional heteroscedasticity with estimates of variance of united kingdom inflation, *Econometrica*, 1982, Vol. 50, pp. 987–1008.

[8] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.

[9] P.C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.

[10] P.C. Fishburn, *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, MD, 1988.

[11] H. Föllmer and A. Schied, *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter, Berlin, New York, 2004.

[12] P. Glasserman, "Calculating portfolio credit risk", In: *Handbooks in Operations Research and Management Science*, Volume 15, Elsevier, 2007, pp. 437–470.

[13] L. R. Glosten, R. Jagannathan, and D. E. Runkle, David E., On the relation between the expected value and the volatility of the nominal excess returns on stocks, *Journal of Finance*, 1993, Vol. 48, No. 5, pp. 1779–1801.

[14] D. Gujarati and D. Porter, *Basic Econometrics*, McGraw-Hill, 2008.

[15] D. Kahneman and A. Tversky, "Prospect theory: an analysis of decision under risk", *Econometrica*, 1979, Vol. 47, pp. 263–292.

[16] R. L. Keeney and H. Raiffa, *Decisions with Multiple Objectives*, John Wiley and Sons, New York, 1976.

[17] M. Kijima and Y. Muromachi, "An extension of the Wang transform derived from Bühlmanns economic premium principle for insurance risk", *Insurance: Mathematics and Economics*, 2008, Vol. 42, No. 3, pp. 887–896.

[18] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic*, Prentice Hall, New Jersey, 1995.

[19] V. Kreinovich, H. T. Nguyen, and S. Sriboonchitta, A new justification of wang transform operator in financial risk analysis, *International Journal of Intelligent Technologies and Applied Statistics* (to appear).

[20] R. D. Luce and H. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

[21] D. B. Nelson, Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, 1991, Vol. 59, pp. 347–370.

[22] H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.

[23] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, CRC Press, Boca Raton, Florida, 2006.

[24] J. Pexider, Notiz uber Funktionaltheoreme, Monatsch. Math. Phys., 1903, Vol. 14, pp. 293–301.

[25] H. Raiffa, *Decision Analysis*, Addison-Wesley, Reading, Massachusetts, 1970.

[26] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC Press, Boca Raton, Florida, 2007.

[27] S. Sriboonchitta and V. Kreinovich, Asymmetric heteroskedasticity models: a new justification, *International Journal of Intelligent Technologies and Applied Statistics* (to appear).

[28] S. Sriboonchitta, W. K. Wong, S. Dhompongsa, and H. T. Nguyen, *Stochastic Dominance and Applications to Finance, Risk and Economics*, Chapman and Hall/CRC Press, Boca Raton, Florida, to appear.

[29] A. Tversky, *Preference, Belief, and Similarity: Selected Writings*, MIT Press, Cambbridge, Massachusetts, 2003.

[30] S. S. Wang, "A class of distortion operators for pricing financial and insurance risks", *Journal of Risk and Insurance*, 2000, Vol. 67, pp. 15–36.

[31] S. S. Wang, "A universal framework for pricing financial and insurance risks", *Astin Bulletin*, 2002, Vol. 32, pp. 213–234.

[32] S. S. Wang, "Equilibrium pricing transforms: New results using Bühlmanns 1980 economic model", *Astin Bulletin*, 2003, Vol. 33, pp. 57–73.

[33] J.-M. Zakoian, Threshold heteroskedastic models, *Journal of Economic Dynamics and Control*, 1994, Vol. 18, No. 5, pp. 931–955.