Choosing the Right Solution of IRR Equation to Measure Investment Success

We consider the problem of multiple mathematical solutions of the IRR equation, which is a mathematical base for numerous applications in the financial industry, such as investment performance measurement, all yields related valuations for bonds, spot interest rates, forward rates, and lots of other applications. Previously, this problem has been studied mostly from the mathematical perspective, and no satisfactory resolution has been found. Our research takes into account both mathematical and business aspects of the problem. We discovered and convincingly proved that the largest root of the IRR equation, which accordingly produces the largest rate of return, is the most adequate solution of the IRR equation, both from mathematical and business perspectives, which should be used in practical computations of rate of return based on the IRR equation. Based on our study, we introduced the "Rule of the largest root" for choosing the right solution of the IRR equation, which effectively solves the problem of multiple roots of the IRR equation. Solving this long-standing problem, which is of very high practical and theoretical importance in finance, opens lots of new opportunities for developing new robust financial instruments and advanced analytical methods.

Yuri Shestopaloff, Ph.D
is Director of Technology, Research & Development at SegmentSoft Inc. He is a Doctor of Sciences and has a Ph.D. in developing mathematical methods for data interpretation. He develops mathematical algorithms for the financial industry and other areas, designs and develops middle- and large-scale software applications and systems, including performance measurement and trading systems for major financial institutions. He has one hundred professional publications and is the author of fourteen books.

Alexander Shestopaloff
is presently a Ph.D. student at the Department of Statistics, University of Toronto. He is a recipient of provincial and federal governments' scholarships. Previously, he published several articles on the subject, one of which received an award from The Journal of Performance Measurement. *His other areas of interest include machine learning, neural networks, theory of probability, and teaching statistics.*

AMBIGUITY OF IRR EQUATION'S SOLUTIONS

Let us consider the following example. An investor invested \$1 in a penny stock and made withdrawals (W) and deposits (D) at yearly intervals as shown in the table below.

The compounding rate of return for such a portfolio is defined by the following equation (It is called the IRR equation, where IRR stands for "internal rate of return"):

$$(1 + R)^3 - 3.28(1 + R)^2 + 3.591+ R) = 1.31$$

Often, such equations are rewritten in the following mathematically equivalent form:

$$1 - \frac{3.28}{(1+R)} + \frac{3.59}{(1+R)^2} = \frac{1.31}{(1+R)^3}$$

Table 1

Time, years	0	Year 1	Year 2	Year 3
Market value, \$	1	5	4	1.31
Cash transaction, \$	1	3.28 W	3.59 D	

The problem with this equation is that it has multiple solutions: $R_1 = 6\%$, $R_2 = 1\text{-}\%$, $R_3 = 12\%$. Mathematically, all solutions look possible. However, which one does the investor have to use to evaluate performance of his portfolio? In fact, similar situations occur not only in the investment performance measurement business but in many others, as well as in more numerous financial applications in which IRR equation or its derivatives are used. Specialists often do not realize that formulas they are using are nothing but the same classic IRR equation rewritten for a particular task. Many financial analysts will be surprised to know that the IRR equation is a mathematical base for a very large range of financial instruments and methods. For instance, bond related analytics are based on the IRR equation, such as yield to maturity, bond equivalent yield, effective annual yield, yield curve, bond value, yield to call, bond stripping and pricing of coupon bonds, etc. Spot rates (the current interest rate appropriate for discounting a cash flow of some given maturity) and forward interest rates are computed via the IRR equation. All mortgage and annuity mathematics are based on a particular simplified version of the IRR equation. Expectation hypothesis of interest rates (a theory that forward interest rates are unbiased estimates of expected future interest rates) relies on the IRR equation as well. Calculations for currency swap agreements involve a particular form of IRR, too (Bodie *et al*. (2005)). NPV, MIRR are derivatives of IRR equation, which is discussed in detail in Shestopaloff and Shestopaloff. (2011a). However, the importance of the IRR equation does not stop here, since many practical financial methods are based on calculation of rates of return, such as portfolio analysis and composition, risk measurement, and performance attribution. Examples of interesting and thought provoking approaches and commentaries can be found in the articles of Waring (2006), Kozhemiakin (2006), Hood (2005), Kritzman *et al.* (2006), and Busse *et al.*(2010), which emphasize the importance of objective valuation of rates of return as a foundation of risk measurement, attribution analysis, and investment performance measurement. Calculation of much accounting data requires usage of rates of return or interest rates, such as the cost of debt and its optimal structure considered in Binsbergen *et al.* (2010).

The rate of return is indirectly linked to many financial instruments; for instance, risk measurement values which relates to a company's debt capacity (Rampini and Viswanathan (2010)). A good review of historical developments and presently used methods for computing rates of return of investment portfolios and quantitative parameters based on them is given in Spaulding (2005) and Spaulding *et al*. (2009). There is interesting research in Osborne (2010) that studies the relationship of multiple roots of IRR equation with the NPV method.

Compared to the IRR equation and associated methods such as NPV and MIRR, both the Modified Dietz equation and TWRR have one traditional advantage: they always produce the same solution. The IRR equation and its siblings generally have multiple solutions. In practice, for large portfolios with relatively small transactions, the problem of ambiguity of solution of the IRR equation does not arise often, but such possibility always exists and it should be under control. When volatility of the involved financial instruments increases and the relative amount of transactions increases, the probability of meaningful multiple roots substantially rises too. In either case, from a practical perspective, the problem of multiple roots of the IRR equation should be addressed. From a theoretical perspective, solving this problem would provide many interesting and useful insights into the nature of compounded rates of return.

Ambiguity of IRR equation solutions is one of the major factors that impede the wider acceptance of IRR and associated methods. The community agrees that IRR represents a fairly objective measure of rate of return, while in many of the financial applications listed above, for instance, in mortgage and annuities related mathematics, there are no alternatives. IRR equation is the only available mathematical vehicle. So the area of application of the results of this study is very wide and can be beneficially used by financial analysts in all of the considered financial applications. Besides, introducing robust criteria, which uniquely defines the solution of the IRR equation when it has multiple roots, will be of great benefit to software developers of financial systems.

In this article, we propose methodological approaches and practical methods for determining how to unambiguously choose the correct solution of the IRR equation when it has more than one mathematical root. These approaches can be easily extended to methods associated with IRR, such as NPV, MIRR, mortgage and annuity equations, bonds related mathematics, etc. Some researchers also study the complex roots of the IRR equation and their practical meaning, as well as other representations of returns; for instance, matrices, including matrices with complex elements, (Pierru (2010)). Our considerations are applicable to complex roots as well, although below we only consider real solutions of the IRR equation. Presently, the business meaning of complex roots has several interpretations, sometimes contradictory, and there is no consensus on how to use complex roots of the IRR equation in financial applications. On the other hand, in this study, the use of the complex solution with the largest modulus is consistent with our considerations and even may complement some of them; for instance, the notion of continuity of solutions when the parameters of the IRR equation change continuously. So complex roots should not be discarded, but lots of further studies have to be done in order to prove their usefulness and efficiency. In this regard, the model we consider using the real roots of the IRR equation to measure return on investment is also imperfect. However, as research showed, it can be considered to be an objective description of investment phenomena in a wide domain of applications.

IRR EQUATION AND ITS PROPERTIES

Mathematically, the IRR equation is a generalized polynomial equation, which is sometimes called a power equation (Rahman (2002)), (Shestopaloff (2010), Shestopaloff and Shestopaloff (2011a) Shestopaloff (2011b), (Chestopalov and Beliaev (2004)). Shestopaloff (2010, 2011b) shows that the maximum possible number of positive solutions of such an equation is equal to the number of sign changes of the equation's coefficients or less by a multiple of two when this equation is written in the descending order of powers (we assume that the free term has degree zero). For instance, for the IRR equation, it means that when we do not have withdrawals, then all cash flows are positive, and the only sign change occurs on the free term (summand in the equation which is a known number; for instance, in the equation (x + 2) = 0, number 2 is a free term). In such cases, the IRR equation always has a unique positive solution for the value of (1 + R), where *R* is the rate of return. Given the fact that $R \geq -1$, it means that this value covers the whole domain of possible rates of return and we should not consider negative solutions of the IRR equation with respect to the value of (1 + R).

Mathematically, the IRR equation can be written in different forms. Derivation and complete analysis of different forms of the IRR equation can be found in Shestopaloff (2009) and Shestopaloff and Shestopaloff (2011a). The main difference is due to which reference time we choose. For instance, if we discount the cash flows to the beginning of the investment period, then we compute the *present value P*. This is done as follows.

$$P = B + \sum_{j=1}^{j=N} \frac{C_j}{(1+R)^{t_j}} \quad . \qquad (1)$$

Here, C_j is a cash flow; t_j is the time period between the beginning of the investment period and the moment when the cash flow occurred; *T* is the total period, *B* is the beginning market value; *N* is the number of cash flows.

Note that the ending market value *E* and the present value are related as follows.

$$P = \frac{E}{(1+R)^T} \quad . \qquad (2)$$

If we substitute (2) into (1) and $(1 + R)^T$, multiply both parts of (1) by , then after transformations we obtain the following convenient formula.

$$E = \sum_{j=0}^{j=N} C_j (1+R)^{T_j} \quad . \qquad (3)$$

Here, $T_j = T - t_j$ is the real-valued time period. The cash flow remains in the portfolio; that is, from the moment when it appears in the portfolio *until the end* of the total period *T*; $C_0 = B$ (the beginning market value of the portfolio). Note that all periods have to be measured in the *same* units of time. Cash inflow (adding cash to a portfolio) is positive; cash outflow (withdrawal from a portfolio) is negative. The convenience of formula (3) is that it treats the beginning market value *B* in the same way as a regular cash flow, done at the beginning of the investment period. In many instances, this form is more suitable for mathematical transformations and business interpretations.

The rate of return R computed on the basis of (3) relates to *one unit of time*. So the total rate of return R_T for the whole investment period *T* should be calculated as $R_T = (1 + R)^T - 1$. Often, however, it is assumed that the total period T = 1.

We can see that the rate of return R in (3) is not multiplied by the period length. Also, the power in (3) denotes the period length, but not the number of periods, although we know that the IRR equation does a compounding. Why is this? In (3), we assume that the rate of return corresponds to one unit of time, and the number of periods T_j is the same as the number of units of time in T_j(note that the number of periods can be fractional). It means that the realized gains are implicitly compounded over a number of periods equal to the number of units of time in the considered time interval T_j. This issue is often overlooked in applications. If we want to have more flexibility with regard to the length of the time period, and choosing the unit of time, then this default implicit setting may be overridden, and we can do compounding over an arbitrary period of length τ_j for each cash flow. This way, we obtain the following general form of IRR equation.

$$E = \sum_{j=0}^{j=N} C_j (1 + R\tau_j)^{\frac{T_j}{\tau_j}} \quad . \qquad (4)$$

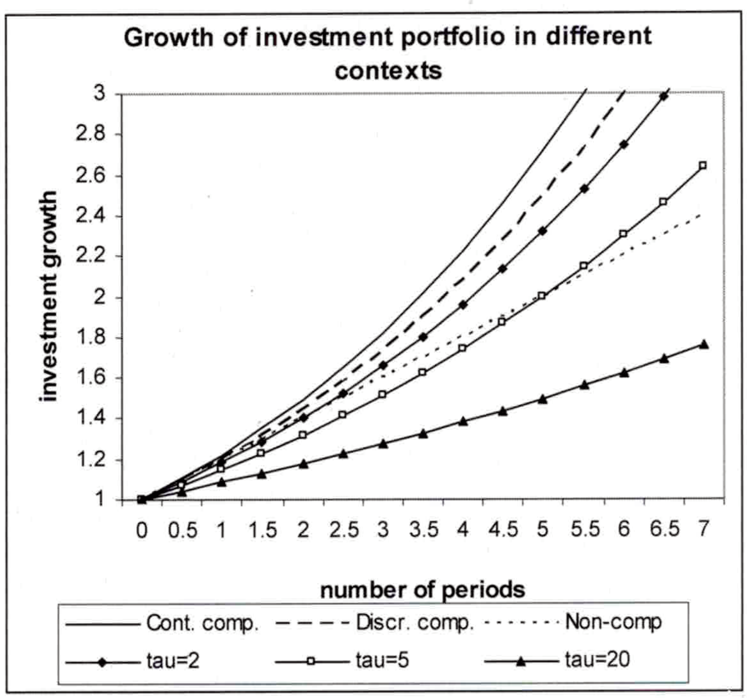
This generalized IRR equation (GIRR) first introduced and comprehensively studied in Shestopaloff and Shestopaloff (2011a) presents an interesting development. On one hand, it adds a great deal of flexibility to how we can compute the rate of return based on additional information. For instance, if the actual compounding is done infrequently, we can increase the length of the compounding period τ_j in order to make our valuations more objective. On the other hand, the general form of IRR equation (4) *continuously* covers the whole spectrum of investment contexts, including noncompounding and continuous compounding. These important notions of investment contexts were introduced in Shestopaloff (2008) and further developed in Shestopaloff (2009). When τ_j goes to zero, (4) transforms into the following IRR equation with continuous compounding.

$$E = \sum_{j=0}^{j=N} C_j e^{RT_j} \quad . \qquad (5)$$

The comprehensive study of properties of generalized IRR equation can be found in Shestopaloff and Shestopaloff (2011a). On the other hand, if we assume in (4) that the values of τ_j go to $\tau_j = T_j$, then (4) *continuously* transforms to the *noncompounding* scenario. Formula (4) transforms into

$$E = \sum_{j=0}^{j=N} C_j (1 + RT_j) \quad . \qquad (5)$$

Figure 1



However, (4) has even more to offer. If we continue our exploration to the opposite side, that is when $\tau_j \to \propto$, then we also *continuously* arrive at the following limit that is proved in Shestopaloff and Shestopaloff (2011a).

$$E = \sum_{j=0}^{j=N} C_j \quad . \tag{6}$$

Condition (6) means zero rate of return (or zero interest rate), which we can call the "noninterest context" when we always obtain the same amount that we invested. So (4) describes the *continuous* range of investment contexts from the noninterest context to continuous compounding with noncompounding context as an intermediate investment scenario. Figure 1 presents different scenarios of investment growth based on (4) for different values of τ. We show a case of continuous compounding that practically coincides with the appropriate graph of discrete compounding when $\tau = 0.01$.

PROPERTIES OF SOLUTIONS OF IRR EQUATION

Let us consider a particular IRR equation. The following IRR equation may have a maximum of three positive solutions, because it has three sign changes in its coefficients: $1.0 \to (-2.5)$

$(-2.5) \to 1.05$, $0.97 \to (-0.2)$.

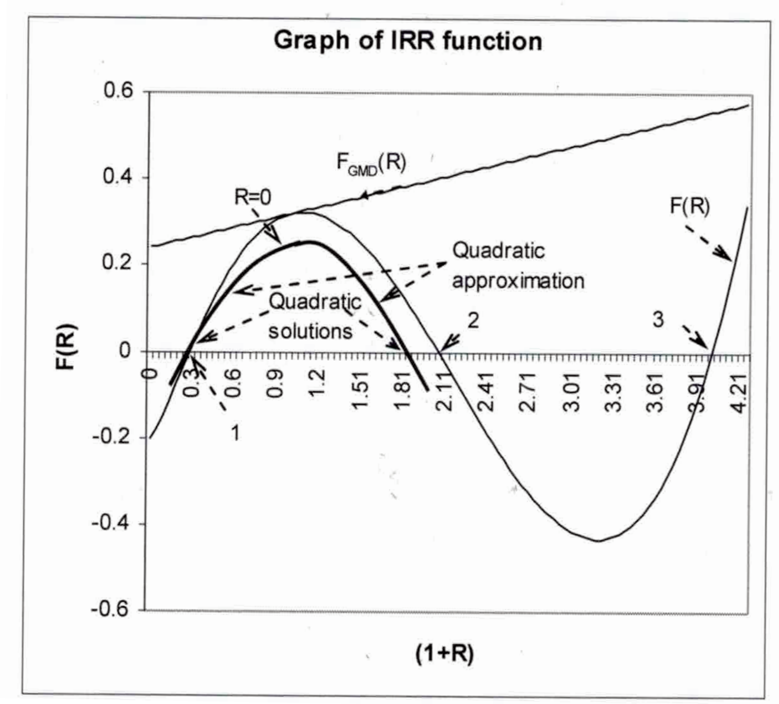
$$(1+R)^{2.64} - 2.5(1+R)^{2.24} + 1.05 \times (1+R)^{1.73} + 0.97 \times (1+R)^{1.26} - 0.2 = 0 \quad . \tag{7}$$

The graph of the appropriate IRR function *F(R)* is shown in Figure 2.

$$F(R) = (1+R)^{2.64} - 2.5(1+R)^{2.24} + 1.05 \times (1+R)^{1.73} + 0.97 \times (1+R)^{1.26} - 0.2 \quad . \tag{8}$$

The interpretation of the IRR equation (7) is as follows. The beginning market value is \$1.0 (which is the coefficient of the term $(1+R)^{2.64}$). Then, we made a withdrawal of \$2.5 and added (deposited) \$1.05 and \$0.97. The ending market value of the portfolio is \$0.2. The whole period has a length of 2.64 units of time, and the intermediate cash flows occur accordingly at times 2.24, 1.73, and 1.26, measured from the end of investment period. Once we obtain the rate of return R for one unit of time, we can recalculate it for the whole period *T*=2.64

Figure 2



Graph of IRR Function (8) and the Three Solutions (1, 2, 3) of the IRR Equation (7).

using the following formula.

$$R_T = (1+R)^T - 1 \quad . \qquad (9)$$

Our portfolio grew rapidly between the start and first cash transaction because we were able to withdraw \$2.5, while we invested only \$1.0 at the beginning. Actually, relatively large (compared to the market value of portfolio) cash flows with opposite signs are the reason why multiple solutions of the IRR equation appear.

Note that the units in which we measure period lengths do not influence the value of the rate of return for the whole period. This can be proved as follows. Let us denote z = (1 + R); *A* is some positive real number, such that $y = z^A$. Then we can rewrite (3).

$$E = \sum_{j=0}^{j=N} C_j z^{T_j} = \sum_{j=0}^{j=N} C_j \left(z^A\right)^{\frac{T_j}{A}} = \sum_{j=0}^{j=N} C_j (y)^{\frac{T_j}{A}} \quad . \qquad (10)$$

Let us assume that z_0 and y_0 are solutions of IRR equations

$$E = \sum_{j=0}^{j=N} C_j z^{T_j}$$

and

$$E = \sum_{j=0}^{j=N} C_j (y)^{T_j / A}$$

Then the appropriate total rates of return, according to (9), can be found as follows.

$$R_{Tz} = (1+R)^T - 1 = z^T - 1 \quad . \qquad (11)$$

$$R_{Ty} = (1+R)^T - 1 = y^{\frac{T}{A}} - 1 = \left(z^A\right)^{\frac{T}{A}} - 1 = z^T - 1 \quad . \qquad (12)$$

We can see that the rates of return produced by (11) and (12) for the total period are identical. For example, let us take $(1 + R)^{2.64} = (1 + R_a)$. This means that the new unit of time becomes equal to 2.64 previous units of time, so that the new variable R_a denotes the rate of return per 2.64 previous units of time. In the case of (7), this is the whole period. Then we can rewrite the IRR function (8) as follows.

$$F(R_a) = (1+R_a) - 2.5(1+R_a)^{0.848} + 1.05 \times$$
$$(1+R_a)^{0.66} + 0.97 \times (1+R_a)^{0.477} - 0.2 \quad . \qquad (13)$$

Once we know R_a, we can find the rate of return for an arbitrary period length using (9).

The graph of IRR function in Figure 2 presents the typical features of IRR functions. Their value at the beginning point R = -1 is negative (equal to ending market value with opposite sign); they have zero, one or more oscillations and always go to plus infinity when the rate of return goes to plus infinity. All of its roots always lie in an interval with finite length. In Figure 2, as examples for illustrating our considerations, we use the IRR functions (8) and (13). We note that any other possible IRR function differs from these examples only in its number of oscillations.

FINDING SOLUTION OF IRR EQUATION

Which of the three solutions presented in Figure 2 is the correct one from a business perspective? One possible way of solving this problem is to use approximate solutions of the IRR equation that produce less ambiguous results, and then choose the root of the IRR equation which is closest to such solutions. Let us explore this possibility. We have the following candidate methods for this task:

- Simple rate of return
- Generalized Modified Dietz (GMD) method
- Zero quadratic approximation
- Quadratic approximation (needs approximate value of rate of return)

Let us apply these methods to our case. Simple rate of return takes into account only the beginning and ending market values of a portfolio, so that

$$R_S = (E - B) / B = (0.2 - 1.0) / 1.0 = -0.8$$

Quadratic approximation methods are presented in Shestopaloff (2009). These methods generally produce two solutions, but it is still easier to choose between two values than between three or more. Later we will introduce the so-called “Rule of Ascending Curve,” which excludes one of the approximate roots found by quadratic approximation methods. This allows us to unambiguously select the solution that has a valid business meaning in the case of quadratic approximation methods.

Generalized Modified Dietz (GMD) formula from Shestopaloff (2009) produces R_{GMD} = -4.066. The graph of the GMD function, which is a straight line, is shown in Figure 2.

$$F_{GMD}(R) = \sum_{j=0}^{j=N} C_j (1 + RT_j) - E$$

Note that this line is *tangent* to the IRR function F(R) at the point R = 0, as shown in Figure 2. This graphically illustrates the result earlier obtained in Shestopaloff (2009) that GMD and Modified Dietz (MD) methods represent linear approximations of the IRR equation by a Taylor series at the point R = 0. The solution R_{GMD} = -4.066 is the point of intersection of this line with the abscissa. This is an invalid result because by definition, the rate of return cannot be less than (-1.0). Shestopaloff (2009) gives a detailed account of this property of the GMD and Modified Dietz (MD) methods, namely their ability to produce invalid values of rates of return; *i.e.*, less than (-1.0). We can see that in certain situations the tangent can be parallel to the abscissa, which would correspond to solutions infinitely larger in absolute value.

The idea of quadratic approximation is demonstrated in Figure 2. We approximate part of the IRR function F(R), defined by formula (13) by a quadratic parabola (in Figure 2, this parabola is shown in bold, and the parabola is oriented upside down). In the case of the Zero Quadratic Approximation (ZQA) method, the graph of this parabola is a tangent to IRR function F(R) at the point R=0 (this is where the word "zero" in the name of this method comes from). In Figure 2, this is the left branch of the aforementioned upside down parabola that is tangential to IRR function at point . In the case of a general quadratic approximation, we draw a similar parabola, with the only difference that such a parabola is tangential to IRR function F(R) at an arbitrary point R =R_0.

The approximating parabola may intersect the abscissa at two points, touch the abscissa at one point, or not intersect the abscissa at all. Accordingly, the quadratic approximation may produce two, one, or no solutions. The ZQA quadratic equation for an arbitrary period length, which is derived similarly to formula (5.21) from Shestopaloff (2009), is presented below.

$$R^2\left[\sum_{j=0}^{j=N} \frac{1}{2} C_j T_j (T_j - 1)\right] + R\left[\sum_{j=0}^{j=N} C_j T_j\right] + \left[\sum_{j=0}^{j=N} C_j - E\right] = 0 \quad . \qquad (14)$$

Here, we assume B = C_0. Solving (14) for the example above, we find R_{ZQA1} = -0.7356, R_{ZQA2} = 0.8978. Note that the closest solution of the IRR equation (14) is the first solution R_{IRR1} = 0.7219. The second solution of this IRR equation is R_{IRR2} = 1.055. Analyzing the rates of return we have thus obtained, we may be inclined to say that the true rate of return is negative.

However, what we said above is still a qualified guess rather than a solid algorithm. The simple rate of return is not reliable when cash flows are comparable to the market value of our portfolio. ZQA uses an arbitrary tangency point of R = 0. In order to further reduce the ambiguity and enable us to make a better founded final decision, we should introduce additional approximations. One is a rate of return that resembles the simple rate of return but accounts for the cash flows. We call it "cash based rate of return," and it is defined as follows (Shestopaloff and Shestopaloff (2011a)).

$$R_{CB} = \frac{E + \sum_{i+1}^{i=I} |W_i|}{B + \sum_{j=1}^{j=J} D_j} - 1 \quad . \qquad (15)$$

Here, the letter *W* denotes cash withdrawals, *D* – cash deposits, so that if the total number of cash transactions is *N*, then N = I + J. Vertical lines in the numerator of (15) denote absolute values. The idea behind (15) is this. We compare how much money we deposited into our portfolio to how much we withdrew. We have to use absolute values because cash withdrawals are represented by negative cash flows.

Substituting into (15) the values of our portfolio, we find R_{CB} = -0.106. We can further use a quadratic approximation (QA) at this point, in which case we obtain the value of RQA = -07236. We can then use this value as an initial value for iterative algorithms, such as Newton-Raphson, linear iterative, or quadratic iterative algorithms from Shestopaloff (2009). All of these algorithms produce the value of rate of return R =

Table 2: Rates of Return Defined by Different Methods		
Method	First Solution	Second solution
Cash based (CB) RR	-0.106	N/A
GMD	-0.106	N/A
QA (CB)	-0.7236	0.841
ZQA	-0.7356	0.8978
Simple IRR	-0.8	N/A
IRR	-0.7219	1.055

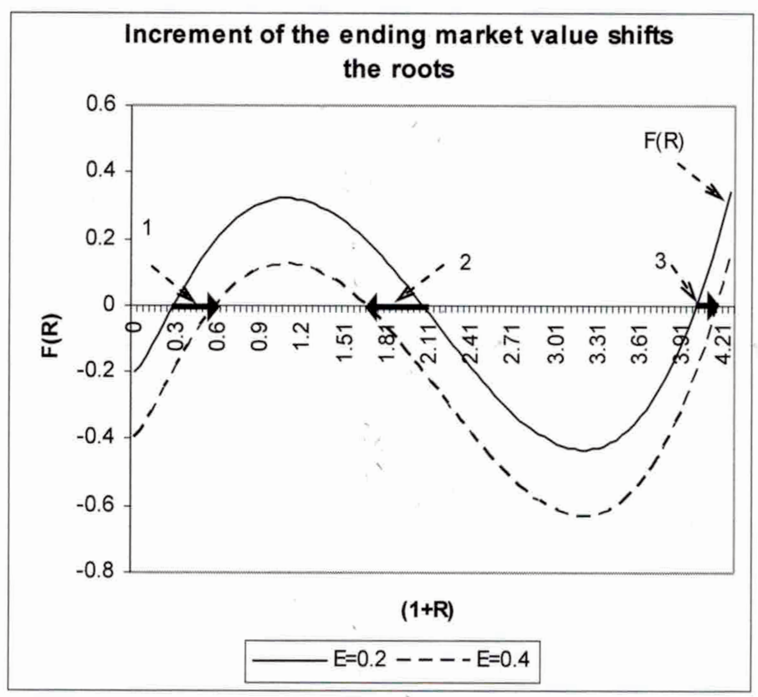
-0.7219. Results are summarized in Table 2.

THE RULE OF ASCENDING CURVE

The next criterion, which alone decreases the number of candidate solutions by about half, is based on the following mathematically elegant consideration. Let the ending market value increase while keeping the beginning market value and all cash flows and transaction times the same. It is common sense that in this case the rate of return should increase as well because the final gains are higher. In geometrical terms, referring to Figure 3, the increase of the ending market value means "sinking" of the graph of IRR function F(R), while all coordinate axes remain fixed. For instance, when E changes from 0.2 to 0.4 in Figure 3, the values of solutions change in the directions indicated by bold arrows.

We can see that only solutions 1 and 3 satisfy this condition, that is, they are the ones that increase their values

Figure 3



Increment of the ending market value shifts the roots. Incrementing the portfolio's ending market value increases the value of the first solution, decreases the value of the second solution and increases the value of the third solution.

because they are located on the *ascending* parts of the IRR function. Solution 1 corresponds to R1 = -0.7219. Solution 3 corresponds to the value of R3 ≈ 3.166. Solution 2 should be excluded because it does not satisfy the rule of ascending curve. Obviously, the same consideration holds true for IRR functions that have a greater number of oscillations, which means that solutions located on the descending parts of the oscillations have to be excluded. We call the discovered criterion "The Rule of Ascending Curve."

How do we know when a solution resides on the ascending part of the IRR function? The first derivative of the IRR function F(R) should be positive for such a solution. In mathematical terms, this means

$$\sum_{j=0}^{j=N} C_j T_j (1+R)^{T_j - 1} > 0 \quad . \qquad (16)$$

Although the remaining candidate rates of return are very different, we still lack some solid criterion that would allow us to uniquely select the correct solution. From the business perspective, we lost some money, but not as much as the first negative solution R_1 = -0.7219 implies. On the other hand, the $1 that we invested from the very beginning grew so rapidly that in 0.4 units of time (2.64 - 2.24) = 0.4 we were able to withdraw $2.5, which means that our investment during this period had a rate of return of at least 625% in the noncompounding context. Afterward, we were losing money, but not as fast, so that the final rate of return became about 317%, which looks acceptable. So the third solution is not a bad candidate for the correct solution. However, this is still an assumption, and we need additional criteria to make a final choice.

THE RULE OF CONTINUITY

Let us introduce a criterion based on the continuity of rate of return as a function of the ending market value. We will consider only real, not complex, solutions of the IRR equation. Note that the Rule of Ascending Curve, probably for the first time in the literature, takes into account the *business meaning* of the solutions of the IRR equation. Previously, all studies considered the IRR equation from a mathematical perspective, except for the restrictions that required a positive beginning market value and nonnegative ending market value. In this subsection, we introduce another root selection criterion that is also based on the business meaning of the IRR equation.

In order to understand it, let us rewrite (8) and consider varying the ending market value.

$$F(R) = (1+R)^{2.64} - 2.5(1+R)^{2.24} + 1.05 \times (1+R)^{1.73} + 0.97 \times (1+R)^{1.26} - E \quad . \qquad (17)$$

Let us assume that the correct value of rate of return is solution 1. If the ending market value increases, then this would mean that the function F(R) in Figure 3 sinks deeper below the abscissa, and solution 1 increases accordingly. At some point (when $E_M \approx$ \$0.52324), the function F(R) will be tangent to the abscissa. The value of the rate of return at this tangency point is $R_{1M} \approx$ 0.0828, and solutions 1 and 2 coincide. Now if we increase the ending market value by even a very small amount, the IRR equation will have a single solution of about $R_3 \approx 3.245$. So an infinitesimally small change in the ending market value creates a large jump in the value of the rate of return; that is from $R_{1M} \approx 0.0828$ to $R_3 \approx 3.245$. Such jumps in the value of the rate of return are poorly compatible with business logic, which requires that small changes in the portfolio's ending market value have to cause small changes in the value of the rate of return. In other words, continuous changes in the ending market value should be accompanied by a continuous change of the rate of return. The only solution that satisfies this requirement is solution 3. In this case, the ending market value can change from zero to infinity, and the rate of return will accordingly change *continuously* and *mono tonically*. All other solutions of the IRR equation, at some value of the ending market value, produce a finite jump in the value of the rate of return at some point when the ending market value changes by an infinitesimally small amount. This property follows from the fact that the IRR function is a generalized polynomial, so that it cannot have vertical asymptotes and consequently its extrema (maximums and minimums) are *finite*.

We need to be sure that the IRR function always has an ascending part that intersects (or begins from) the abscissa and then goes to plus infinity. This is guaranteed by the dominance of the first term when the rate of return increases, in (17) in particular and in general in (3) and (4), when the rate of return goes to infinity. This term is always positive because it is associated with the beginning market value. The rigorous mathematical proof of this property of the IRR function can be found in Shestopaloff (2009) and Shestopaloff (2010, 2011b).

In order to understand the continuity properly and prove that the solution of the IRR equation continuously changes when its coefficients continuously change, we consider continuous change of the ending market value. Note that the same continuous change of the IRR solution will be observed when we change other parameters of the IRR equation, such as the values of all cash flows and periods. Minor changes of these parameters should be accompanied by appropriately small changes of the value of the rate of return.

In strict mathematical terms, such a continuity of solutions is formulated as follows: if we increase the value of the cash deposit C_j by some small value of δC_j, then the rate of return R_c has to decrease by an appropriate small value of δR_{Cj}, so that the following limit should hold true for any cash deposit.

$$\lim \delta R_{C_j \, \delta C_j \to 0} = 0 \quad . \quad (18)$$

Similarly, if we increase the value of some cash withdrawal by a small amount, then the rate of return should increase by an appropriately small amount and a limit similar to (18) should hold true. If we change the time of some cash transaction by a small value - for instance, make it earlier - then the rate of return should decrease by an appropriately small amount and no jumps or breaks in the value of the rate of return should occur. If we consider solutions of the IRR equation from this perspective, then all solutions, except the largest one, have *finite* domains which these solutions change continuously when the IRR equation's parameters vary. This is caused by the presence of *oscillations* in the IRR function. Unlike all other solutions, the interval in which the largest solution changes continuously is unbounded. So the largest root of the IRR equation is the best choice from the point of view of continuity.

However, although it is the best among other solutions, this root also does not satisfy the continuity requirement completely. For instance, if we begin to increase the beginning market value of $1 in (8), then the minimum on the right eventually intersects the abscissa and the value of the largest real root jumps from solution 3 to solution 1, in which case solution 1 becomes the largest. So we made some progress, but the verdict is not final yet, and we will continue our discussion. Let us take a closer look at our candidate solutions using some properties of the generalized IRR equation (4).

CONTINUITY OF SOLUTION OF THE GIRR EQUATION

As we learned, GIRR equation (4) allows for continuous change of the investment context when the equation's parameter τ changes from zero to infinity. So the equation's solutions change continuously as well. Figure 4 shows how changing τ effects the values of solutions of the equation corresponding to the IRR function (13) when it is presented in the form (4). We can observe interesting behavior of the first root when τ is increased to be greater than one; the root remains negative and asymptotically approaches zero, and this is not an optical illusion. This root has the following property. Its absolute value is less than one, so that the term $(1 + R\tau)$ is also less than one, or it may even be negative. However, this term is meaningful only when $(1 + R\tau) \geq 0$, so we consider the case when $(1 + R\tau)$ is nonnegative but less than one. As it is known, if such values are raised to a positive power greater than one, the result is *less* than the base, while for the bases that are greater than one, the result is *greater* than the base. The opposite effect is present when the power is positive and less than one. In this case, a base that is less than one produces a greater number, while a base that is greater than one produces a smaller number.

In our case, the value of the rate of return corresponding to the first solution is approximately $R = -1/\tau$, so that the value of $(1 + R\tau)$ is small, nonnegative, and less than one. Since T_j / τ is less than one but positive,

$$(1+R\tau)^{\frac{T_j}{\tau}} > (1+R\tau)$$

though usually not by too much; for instance, the square root of 0.01 is equal to 0.1 (which is a larger number), while the square root of 100 is equal to 10 (which is a smaller number). So the first root essentially represents the case when, in our IRR equation, we sum up small numbers. Growth of τ should "push" the root closer to the value of $R = -1\tau$. When τ grows, the power decreases, which increases the value of

$$(1+R\tau)^{\frac{T_j}{\tau}}$$

The base also decreases, which leads to a smaller value of (1 + Rτ), and this is what we can see in Figure 4. As a consequence, $R \to (-1/\tau)$ when $\tau \to \infty$. Smaller values of τ "liberate" the solution and allow for more divergence from the value of (-1/τ). However, in this case, we can begin to obtain values of rate of return less than (-1.0), which is prohibited by the definition of the rate of return. The fact that such a negative solution may originate in scenarios with obviously positive rates of return jeopardizes the business sense of the smallest solution, and makes it less preferable than the largest solution. Of course, the largest solution can be negative, although in such situations it often becomes the only solution of the IRR equation. So from the business perspective, the largest solution is more preferable than the smallest one.

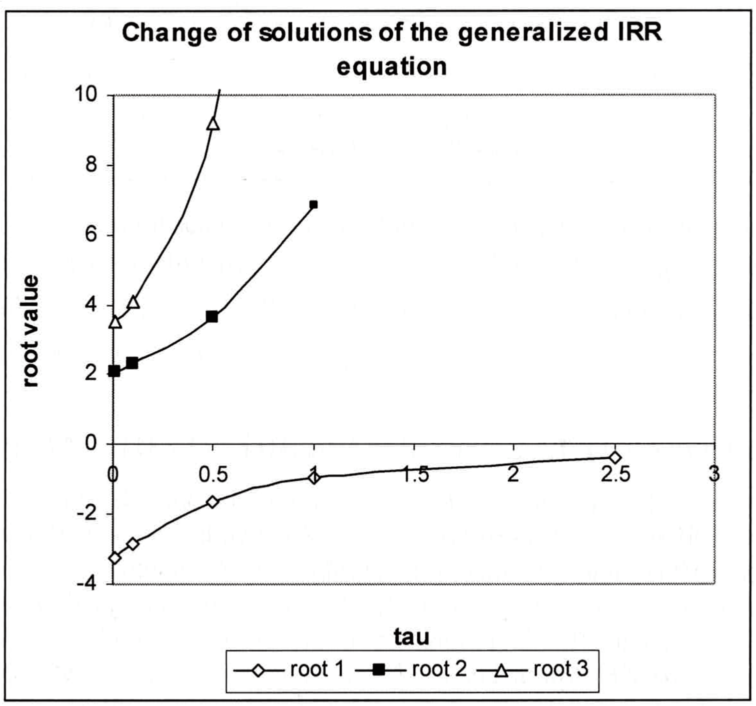
The next question is, how do we know when the smallest solution corresponds to a sum of small numbers and consequently has the convergence features we have just discovered? A good indicator is when we have multiple solutions and the value of the smallest one is negative. We may refer to this as "Smallest Root Convergence Criterion."

In fact, we do not have many scenarios of how the roots of the IRR function can originate. The IRR function can have a spike at the beginning which can or cannot reach the abscissa, illustrated by scenarios 1 and 2 in Figure 5 (values of *C* and *T* relate to the second term in (13)). The IRR function can also begin to increase monotonically (scenario 3 in Figure 5) and thus intersect the abscissa only once. It can also first decrease without any spikes and then return and intersect the abscissa at one point. Certainly, several spikes are possible, and so more intermediate solutions can be present. However, we have previously discarded these solutions in favor of the largest or the smallest solution.

PROPERTIES OF THE IRR EQUATION AT THE SPECIAL POINT

We know that the rate of return R = -1 corresponds to the case when the investment is completely lost; that is the ending market value becomes zero and the IRR function begins at zero at the point R = -1. Note that as the ending market value becomes sufficiently small, the smallest solution will go to this point *continuously*. In other words, the smallest solution of the IRR equation

Figure 4



Dependence of Solutions of the Generalized IRR Equation on the Value of τ

will become R = -1. However, if we take as an example the IRR equation (7), we can see that we actually invested \$3.02, did a withdrawal of \$2.5, and so lost only \$0.52, but not the entire investment, as the value of R = -1 would imply. The same considerations hold true for *any* IRR equation that has more than one root and whose ending market value is equal to zero.

There is more to say here. Multiple solutions occur only if we do some withdrawals; otherwise, we cannot obtain negative coefficients in the IRR equation, whose presence is required to have multiple roots. Moreover, these withdrawals have to be quite significant, comparable in value to the beginning market value, in order for the IRR function to "sink" below the abscissa and produce more than one solution. However, it does not make sense to say that we have lost all of our invested money once we made a withdrawal. This fact undermines the possibility that the first solution is a correct one because selecting it leads to a contradiction from the business perspective.

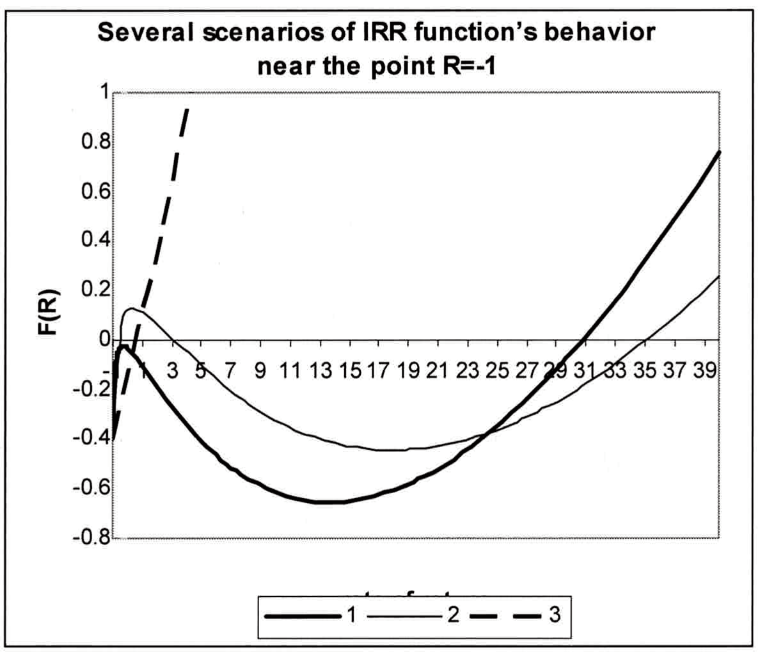
Thus, if the IRR equation has multiple solutions, then relatively large cash withdrawals from the portfolio must have been made, and consequently we cannot lose the *whole* investment in principle. However, when the ending market value goes to zero, the smallest rate of return goes to R = -1, which means that we lost *all* of our invested money. As we have just said, this cannot be true because of the presence of withdrawals.

On the other hand, if we accept the hypothesis that the largest root is correct, then we do not have such a contradiction. So we have found one more reason why we should give preference to the largest solution.

INTERMEDIATE MARKET VALUE OF A PORTFOLIO

The investment context can vary during the same investment period, (Shestopaloff and Shestopaloff (2011a)). However, because we use one rate of return for the whole period, it is reasonable to expect that at any given moment the current value of the investment portfolio makes sense from a business perspective. In particular, if we have several solutions for the rate of return, such that the intermediate market value of the portfolio becomes negative for some solutions, while there is a solution for which it remains positive, then the solution that produces positive intermediate market values should be preferred over solutions producing negative

Figure 5



Several scenarios of IRR function's behavior near the point R = 1. Curve 1 – the spike does not reach the abscissa, C_2 = -2.65, T_2 = 0.83 ; 2 – the spike intersects the abscissa, C_2 = -2.65, T_2 = 0.748; 3 – function has no spike and increases monotonically, C_2 = -2.5T_2 = 0.85.

intermediate market values.

In order for the intermediate market values to remain positive for some solution, we may impose the following restriction, which implies a nonnegative market value of the portfolio at the moment of each cash transaction. This restriction is that for all J = 0,1,..., N and for some solution of IRR equation, $R_k > -1$, k=0,1,…,*K*, the following condition should be fulfilled.

$$E_J(R_k) = \sum_{j=0}^{J} C_j (1+R_k)^{T_j - T_J} \geq 0 \quad . \qquad (19)$$

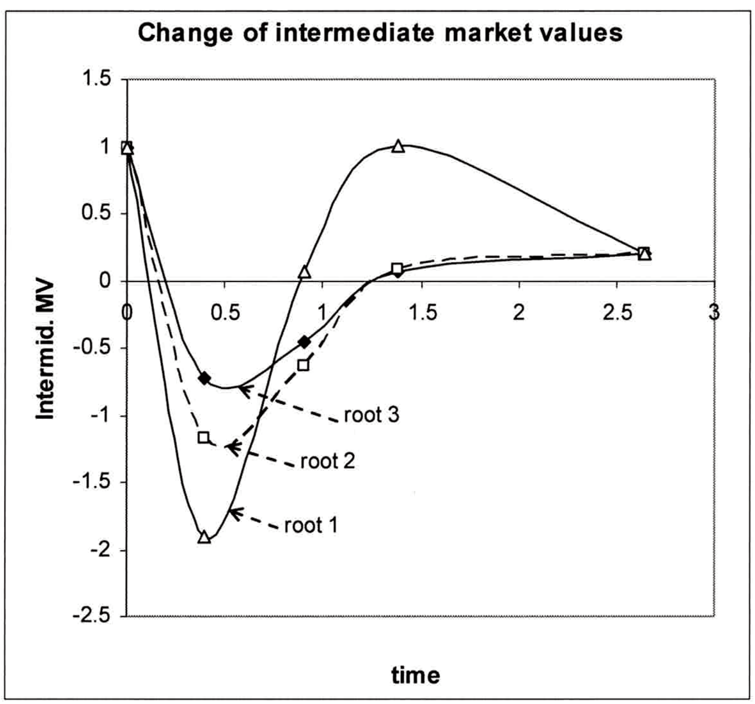
It is possible to create such an artificial portfolio that (19) will not be fulfilled even for the largest root. Such is the case of the IRR equation (7). Graphs of intermediate market values for all three roots of this equation are presented in Figure 6. However, such situations are extreme, and we never encountered them in practical applications. On the other hand, we often encounter situations when the intermediate market values for all roots, except for the largest one, become negative at some point.

It is more likely that (19) is fulfilled for larger solutions. One reason is that the larger the rate of return is, the more dominant the first term becomes, and this term is always positive, as we discussed earlier.

Another advantage of selecting the largest solution is that the intermediate values fluctuate substantially less than when we use the smallest solution. This is an empirical observation, although there is some mathematical rationale for this effect related to the specifics of exponentiation of small (less than one in absolute value) numbers. So, we have one more factor that favors the selection of the largest solution.

This completes our research with regard to which solution of the IRR equation should be accepted from business and mathematical perspectives. This involved the

Figure 6



Change of Intermediate Market Values for the Three Roots of Equation (7)

introduction of several criteria, such as the Rule of Ascending Curve, the Rule of Continuity, and others. All of them favored the largest solution.

PRACTICAL COMPUTATION OF THE IRR EQUATION SOLUTION

Another issue is the practical computation of solutions. In general, the IRR equation, as well as NPV and MIRR equations, can only be solved numerically, usually by using iterative methods. So the choice of initial value for iterative methods is important. There are several possible approaches to finding the correct solution of the IRR equation when the IRR equation has multiple solutions.

We can use some mathematical advances in this area, adjusting them to our purpose. In particular, the work Davenport and Mignotte (1990) present methods for finding the largest root of a polynomial. The problem is that the IRR equation in general is a generalized polynomial whose powers can be real positive numbers, not only integers, which is the case considered in the referred to work. However, we can approximate the IRR equation by a polynomial, using a substitution that approximates the powers by rational numbers. Then, through a series of transformations described in detail in Shestopaloff (2010, 2011) we obtain a polynomial equation that approximates the original IRR equation with high accuracy. Then, the methods in Davenport and Mignotte (1990) can be applied directly. Once we obtain the approximate value of the largest root, we can use this value for iterative or other computational algorithms in order to find the precise value of the rate of return.

In certain instances, we can use the GMD method to find the initial value of the rate of return that can be used in some computational procedure that converges to the largest root of the IRR equation. Some conditions have to be fulfilled in order to guarantee that the procedure converges to the largest root. Certain useful considerations on this account can be found in Shestopaloff (2009).

Another approach would be to find a value of the rate of return on the rightmost ascending part of the IRR curve, when the first term of the IRR equation exceeds by several times, in absolute value, the contribution of other terms. Then this value can be used in some computational algorithm, such as an iterative one, to descend to the largest root. In this regard, the considered quadratic iterative (QI) algorithm is beneficial because it provides very speedy convergence to the solution. The number of required iterations can be several times less than in the case of linear iterative algorithms such as Newton-Raphson. This approach is simple and practical.

Other mathematical and computational methods can also be employed in order to find the largest root. This is a matter of taking into account the problem's specifics, software implementation requirements, available computational resources, and choosing the most adequate approach. In this regard, the book Burden (2005) is a valuable resource for understanding how to choose an efficient computational algorithm.

We tested the discussed approaches, except the ones from Davenport and Mignotte (1990), on simulated and real portfolios, and found that they are very reliable.

NOTE ABOUT COMPLEX SOLUTIONS OF THE IRR EQUATION

Presently, using complex roots of the IRR equation is considered rather exotic. On the other hand, in our considerations, if we substitute real solutions by *moduli* of complex roots of the IRR equation, we will also come to similar results. Namely, if we need to choose some complex root, then, by and large, the complex root that has the largest modulus makes the most sense from a business perspective because it is consistent with all criteria that we introduced above for selecting the correct real solution of the IRR equation. Certainly, the time when complex rates of return will be in use may never come. However, if this ever happens, then the Rule of Continuity that we introduced above could be even more important because in this case the root with the largest modulus will cover the *whole* domain of possible rates of return, and it cannot undergo any jumps.

Another good thing about using complex roots is the natural inclusion of real solutions into the realm of complex solutions, as a particular case. The simplest and probably the most noncontradictory way of doing this is using the moduli of complex numbers. However, the issue requires more research and it is difficult to say up front what will be the best approach.

The notion of context should be an important vehicle in studies of business meaning and applications of complex roots. Complex solutions have to reside in a certain specific context. However, these new contexts have to be connected to known investment contexts and be able to undergo continuous transformation into other contexts.

So studying complex solutions of the IRR equation and its applications should not be forbidden or neglected, but it should also be understood that the introduction of such advanced concepts into the industry, which by its nature has to be reasonably conservative, requires diligent and comprehensive research and lots of qualified discussions.

CONCLUSION

There are several related areas which we covered in this article. The first is the problem of finding an adequate and accurate value of the rate of return. Using the notion of conceptual context of an investment, we presented the structure and interrelationships of IRR methods and their areas of applicability. We introduced the generalized IRR equation, which has many interesting and useful features. The importance of this generalization is that it unifies all available different methods for calculating rates of return and defines their investment contexts. This equation also allows *continuous* transformations from one method and its associated context to another by changing one of its parameters.

We introduced algorithms and criteria for selecting the correct, from the business and mathematical perspective, solution of the IRR equation in the case when it has multiple roots. In this regard, the introduction of the Rule of Ascending Curve, the Rule of Continuity, and others is very illustrative since these criteria naturally reflect the *business* specifics of the IRR equation and eventually allow deriving the *Rule of the Largest Root*, which we believe solves the problem of selecting the correct root of the IRR equation.

REFERENCES

Bodie, Zui, Kane, Alex, Marcus, Alan J., Perrakis, Stylianos, Ryan, Peter , Investments, McGraw Hill Ryerson, Fifth Canadian Edition, 2005.

Burden, Richard L., Faires, J. Douglas, Numerical Analysis, Thomson Brooks/Cole, 8th edition, 2005.

Busse, Jeffrey A., Goyal, Amit, Wahal, Sunil, "Performance and persistence in institutional investment management," *The Journal of Finance*, Vol. 65: 765-790, 2010.

Binsbergen, Jules H., Graham, John R., Yang, Jie, 2010, "The cost of debt," *The Journal of Finance*, Vol. 65: 2089-2136.

Chestopalov, Iourii, Beliaev, Sergei, "A simplified method for calculating the money-weighted rate of return," *The Journal of Performance Measurement*, Vol. 9, No. 2: 14-23, 2004.

Davenport, J. H., Mignotte, Maurice, "On finding the largest root of a polynomial," *Mathematical Modeling and Numerical analysis*, Vol. 24: 693-696, 1990.

Hood, L. Randolph, "Determinants of portfolio performance: 20 years later," *Financial Analysts Journal*, Vol. 61, No. 5: 6-8, 2005.

Kirschman, Jeanette R., Dietz Peter O.,"Evaluating portfolio performance" in: Managing Investment Portfolios: A Dynamic Process, Warren, Gorham & Lamont, 2nd edition, 1990.

Kozhemiakin, Alex, Waring, M. Borton, Siegel Laurence B., "The myth of the absolute-return investor: (Comment and Response)," *Financial Analysts Journal*, Vol. 62, No. 6: 10-12, 2006.

Kritzman, Mark, Hood, L. Randolph, Brinson, Gary P., "Determinants of portfolio performance: 20 years later," *Financial Analysts Journal*, Vol. 62, No. 1: 10-13, 2006.

Osborne, Michael, "A resolution to the NPV - IRR debate?" (on-line article at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=522722), 2010.

Rahman, Qazi I., Schmeisser, Gerhard ,*Analytical Theory of Polynomials*, Clarendon Press, Oxford, 2002.

Pierru, Axel, "The Simple meaning of complex rates of return," *The Engineering Economist*, Vol. 55, No. 2: 105-117, 2010.

Rampini, Adriano A., Viswanathan, S., "Collateral, risk management, and the distribution of debt capacity," *The Journal of Finance*, Vol. 65: 2293–2322, 2010.

Shestopaloff, Yuri K., Shestopaloff, Alex Yu., "A hierarchy of methods for calculating rates of return," *The Journal of Performance Measurement*, Vol. 12, No. 4: 39-52, 2007.

Shestopaloff, Yuri K., "The role of conceptual context in the problem of finding rate of return," *The Journal of Performance Measurement*, Vol. 13, No. 1: 37-50, 2008.

Shestopaloff, Yuri K., *Science of Inexact Mathematics. Investment Performance Measurement, Mortgages and Annuities, Computing Algorithms, Attribution, Risk Valuation*, Toronto: AKVY Press, 2009.

Shestopaloff, Yuri K., *Properties and Interrelationships of Polynomial, Exponential, Logarithmic and Power Functions with Applications to Modeling Natural Nhenomena* Toronto: AKVY Press, 2010.

Shestopaloff,Yuri K., Shestopaloff Alex, Yu., *Solving the Puzzle of IRR Equation. Choosing the Right Solution to Measure Investment Success*, Toronto: AKVY Press, 2011a.

Shestopaloff, Yuri K., "Properties of sums of some elementary functions and their application to computational and modeling problems," *Journal of Computational Mathematics and Mathematical Physics*, Vol. 51, No. 5: 699-712, 2011b.

Shestopaloff, Yuri K., Wolfgang Marty, "Properties of the IRR equation with regard to ambiguity of calculating the rate of return and a maximum number of solutions," *The Journal of Performance Measurement*, Vol. 15, No. 3: p. 8-22, 2011c.

Spaulding, David, Investment performance: A User's Guide, Somerset: TSG Publishing, 2005.

Spaulding, David, Tzitzouris, James A., Brinson, Gary P., Classics in Investment Performance Measurement, Somerset: TSG Publishing, 2009.