RC 21816 (98173) August 3, 2000 Computer Science/Mathematics

Research Report

Variance Reduction Techniques for Value-at-Risk With Heavy-Tailed Risk Factors

Paul Glasserman
Columbia Business School
Columbia University
New York, New York 10027

Philip Heidelberger
IBM T. J. Watson Research Center
P. O. Box 218
Yorktown Heights, NY 10598

Perwez Shahabuddin
IEOR Department
Columbia University
New York, New York 10027

This paper will appear in the proceedings of the 2000 Winter Simulation Conference

Research Division
Almaden - Austin - Beijing - Haifa - T. J. Watson - Tokyo - Zurich

VARIANCE REDUCTION TECHNIQUES FOR VALUE-AT-RISK WITH HEAVY-TAILED RISK FACTORS

Paul Glasserman
Graduate School of Business
Columbia University
New York, NY 10027

Philip Heidelberger
IBM Research Division
T.J. Watson Research Center
Yorktown Heights, NY 10598

Perwez Shahabuddin
IEOR Dept.
Columbia University
New York, NY 10027

ABSTRACT

The calculation of value-at-risk (VAR) for large portfolios of complex instruments is among the most demanding and widespread computational challenges facing the financial industry. Current methods for calculating VAR include comparatively fast numerical approximations—especially the linear and quadratic (delta-gamma) approximations—and more robust but more computationally demanding Monte Carlo simulation. The linear and delta-gamma methods typically rely on an assumption that the underlying market risk factors have a Gaussian distribution over the VAR horizon. But there is ample empirical evidence that market data is more accurately described by heavy-tailed distributions. Capturing heavy tails in VAR calculations has to date required highly time-consuming Monte Carlo simulation. We describe two methods for computationally efficient calculation of VAR in the presence of heavy-tailed risk factors, specifically when risk factors have a multivariate t distribution. The first method uses transform inversion to develop a fast numerical algorithm for computing the distribution of the heavy-tailed delta-gamma approximation. For greater accuracy, the second method uses the numerical approximation to guide in the design of an effective Monte Carlo variance reduction technique; the algorithm combines importance sampling and stratified sampling. This method can produce enormous speed-ups compared with standard Monte Carlo.

1 INTRODUCTION

The calculation of value-at-risk (VAR) for large portfolios of complex instruments is among the most demanding and widespread computational challenges facing the financial industry. The VAR is defined to be an extreme quantile (typically the 99'th percentile) of the distribution of portfolio losses over a time horizon of fixed duration (typically one day or two weeks). Current methods for calculating VAR include comparatively fast numerical approximations—especially the linear and delta-gamma (quadratic) approximations—and more robust but more computationally demanding Monte Carlo simulation. The linear and delta-gamma methods typically rely on an assumption that the underlying market risk factors have a normal (Gaussian) distribution over the VAR horizon. But there is ample empirical evidence that market data is more accurately described by *heavy-tailed* distributions under which occasional very large market moves are more likely than a normal distribution would predict. Capturing heavy tails in VAR calculation has to date required highly time-consuming Monte Carlo simulation.

This paper gives an overview of methods for computationally efficient calculation of VAR in the presence of heavy-tailed risk factors. These methods are described more fully in Glasserman, Heidelberger and Shahabuddin (2000b) (henceforth GHS). Theorems related to these methods are stated here; their proofs are given in GHS (2000b). The methods model market risk factors through a *multivariate t distribution*, which has both heavy tails and empirical support. Our key mathematical result is a transform analysis of a quadratic form in multivariate t random variables. Using this result, we develop two computational methods. The first uses Fourier transform inversion to develop a heavy-tailed delta-gamma approximation; this method is extremely fast, but like any delta-gamma method is only as accurate as the quadratic approximation. For greater accuracy, we therefore develop an efficient Monte Carlo method; this method uses our heavy-tailed delta-gamma approximation as a basis for variance reduction. Specifically, we use the numerical

approximation to design a combination of importance sampling and stratified sampling of market scenarios that can produce enormous speed-ups compared with standard Monte Carlo. Under certain conditions, this Monte Carlo algorithm possesses the highly desirable "bounded relative error" property (in the sense Shahabuddin (1994)) for estimating the relevant rare event properties. It also easily adapts to the estimation of a related useful quantity called the conditional excess (sometimes also called the conditional VAR; see, e.g., Bassi, Embrechts and Kafetzaki(1998) for a discussion).

2 Heavy Tails and Market Data

The multivariate normal distribution is the most widely model of changes in market prices and rates, in large part because of its many convenient mathematical properties. Even in GARCH and related models the innovations are commonly assumed normal, and the increments of any diffusion process are approximately normal over a sufficiently short time horizon. In contrast, virtually all empirical studies report systematic deviations from normality in market data. (An early reference is Blattberg and Gonedes (1974); recent studies find the same patterns.) One of the most pervasive features observed across equity, foreign exchange, and interest rate markets is *excess kurtosis*. This means that, compared to a normal distribution with the same mean and standard deviation, the true distribution assigns greater probability to extreme market moves. Clearly, extreme moves are of paramount importance in risk management and should be modeled accurately in the calculation of VAR.

To contrast the normal and t distributions, it is useful to consider first the univariate case. If Z is a standard normal random variable (mean 0, standard deviation 1), its tail is described by

$$P(Z > x) \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}, \quad x \to \infty,$$

in the sense that the ratio of the two sides approaches 1 as x increases. In contrast, if X has a t distribution with degrees-of-freedom parameter ν, then

$$P(X > x) \sim \text{constant} \times x^{-\nu}. \tag{1}$$

Thus, the two distributions have fundamentally different tail behavior. The power-law decay of the t distribution's tail is far slower than the decay in the normal tail, confirming that large moves will have much greater frequency in a t-based model. Equation (1) further shows that the parameter ν controls the heaviness of the t distribution's tail, with smaller values producing heavier tails. Empirical evidence suggests that ν in the range of 4–6 is appropriate for market returns over short time horizons. As $\nu \to \infty$, the t distribution converges to the normal distribution, so the normal may be viewed as a special, limiting case of the t family.

Modeling the joint distribution of multiple risk factors requires multivariate versions of these distributions. For simplicity we assume a mean of 0; both the normal and t distributions can be translated to produce a nonzero mean. In this case, a multivariate normal density is completely determined by its covariance matrix Σ or, equivalently, by the standard deviations of its components and the correlations between them. The fact that a multivariate normal density is summarized by its standard deviations and correlations is very convenient in modeling market data; this rather special feature is shared by the multivariate t (see, e.g., Anderson (1984)), which has density

$$f(x) = \frac{\Gamma(\frac{1}{2}(n+\nu))}{(\nu\pi)^{n/2}\Gamma(\frac{1}{2}\nu)|\Sigma|^{1/2}} \left(1 + \frac{1}{\nu} x'\Sigma^{-1}x\right)^{-\frac{1}{2}(n+\nu)} \tag{2}$$

for $x \in \Re^n$. If $(X_1, \ldots, X_n)$ has this density with $\nu > 2$, then its covariance matrix is $\nu\Sigma/(\nu-2)$; the factor $\nu/(\nu-2)$ may be viewed as reflecting the heavier tails of the t marginals. A further important distinction between the t distribution and the normal is that uncorrelated normal random variables are mutually independent, whereas the components of a multivariate t are in general dependent even if they are uncorrelated. In modeling market data, this makes it possible to capture a situation in which two risk factors exhibit little dependence in ordinary market conditions but tend to move together in extreme conditions.

In GHS (2000b) we also work with an extension of (2) that allows different marginals to have different degrees of freedom. This is useful in modeling the joint distribution of returns with varying degrees of heaviness in their tails.

3 Quadratic Approximation: Heavy-Tailed Delta-Gamma

Calculating VAR entails finding the distribution of losses over the VAR horizon (e.g., one day or two weeks). Let the random variable L denote the loss on a portfolio over a fixed horizon; finding VAR means finding a point x_p for which $P(L > x_p) = p$ with, e.g., $p = 1\%$.

The delta-gamma method (e.g., Britten-Jones and Schaefer (1999), Rouvinez (1997), and Wilson (1999)) is based on making a quadratic approximation to L of the form

$$L \approx c + b'X + X'AX \quad (3)$$
$$\equiv c + Q. \quad (4)$$

Here, X is a vector of changes in underlying market prices over the VAR horizon, c is a constant, b is a vector, and A is a matrix. In the usual delta-gamma method, X is assumed normal (Duffie and Pan (1999) use a Poisson mixture of normals) and b and A are obtained from the first- and second-order sensitivities of the instruments in the portfolio with respect to the underlying market prices. In other words, the "deltas" and "gammas" of individual instruments are combined to get the overall sensitivities of the portfolio and (3) is a Taylor approximation to the loss. We will take the slightly more general view that some approximation of the form (3) is available.

Our first goal is to approximate the distribution of the portfolio loss L by the distribution of the quadratic $c + Q$ in (4) when X has a multivariate t distribution. Through an orthogonal transformation of the coordinate axes, we may without loss of generality assume that both A in (3) and Σ in (2) have been diagonalized. In this case, we have

$$Q = \sum_i b_i X_i + \sum_i \lambda_i X_i^2,$$

where the X_i are *uncorrelated* risk factors; changes in the actual risk factors are then linear combinations of changes in these risk factors. In the normal case, the analysis proceeds as in Rouvinez (1997) by finding the characteristic function of Q. Because uncorrelated normals are independent, the characteristic function of Q factors into a product of one-dimensional characteristic functions and is thus easy to find. This method does not extend to the multivariate t because the X_i will not in general be independent, even if they are uncorrelated. Also, due to the heavy tail, the characteristic function of each X_i has a complicated form. Hence the charateristic function of Q is intractable and a different approach is needed. To this end, note that X_i can be generated as $X_i = Z_i/\sqrt{Y/\nu}$ where the Z_i's are independent normals with mean 0 and variance Σ_{ii} and Y is a chi-square random variable with ν degrees of freedom, independent of Z. Next, for any x define the random variable

$$Q_x = (Y/\nu)(Q - x)$$

and let $F_x(y) = P(Q_x \leq y)$. One of our key mathematical results is the following:

Theorem 1. $P(Q \leq x) = F_x(0)$ and F_x is the distribution with characteristic function $\hat{F}_x(\omega) = \phi_x(\omega\sqrt{-1})$,

$$\phi_x(\theta) = \frac{1}{(1+\alpha(\theta))^{\nu/2}} \prod_i \frac{1}{\sqrt{1-2\theta\lambda_i}} \quad (5)$$

where

$$\alpha(\theta) = \frac{2\theta x}{\nu} - \frac{1}{\nu}\sum_i \frac{\theta^2 b_i^2}{1-2\theta\lambda_i}. \quad (6)$$

Thus, we have an indirect way of computing the distribution of the quadratic approximation. This result leads to the following method to approximate $P(L \leq x)$ using $P(c + Q \leq x)$:

1. set $x_c = x - c$ with c the constant in (3);
2. numerically invert the Fourier transform $\hat{F}_{x_c}$;
3. evaluate $F_{x_c}(0) = P(c + Q \leq x) \approx P(L \leq x)$.

The procedure can be repeated for multiple values of x in order to approximate the complete distribution of L and find VAR. This method thus combines much of the computational convenience of the traditional normal-based delta-gamma method with greater empirical validity of the multivariate t distribution.

Figure 1 illustrates the potential danger of using a normal-based delta-gamma approximation in a world with heavy-tailed risk factors. Compared with the exact portfolio (horizontal axis), the normal-based delta-gamma approximation (vertical axis, left panel) severely underestimates the magnitudes of large losses. The t-based approximation (right) shows a much better fit. This relationship is also useful in accelerating Monte Carlo, as we discuss next.

4 Fast Monte Carlo Estimation of VAR

The transform inversion method presented above is very fast, but it is only as accurate as the underlying quadratic approximation. The first four columns of Table 1 illustrate the effectiveness of the method in approximating loss probabilities near 1% for a set of option portfolios. The parameters of these test portfolios are detailed in GHS (2000a). The third column gives the $P(L > x)$ estimated using simulation to within 2% accuracy (more precisely the 99% confidence interval half-width is always within 2% of the simulation estimate of the probability). The fourth column gives the

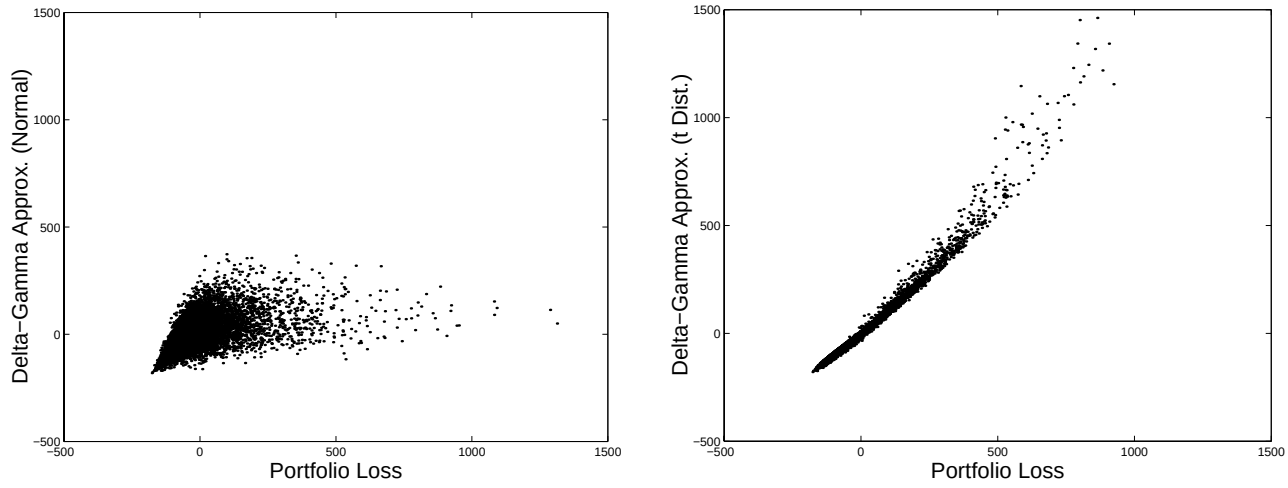

Figure 1: Comparison of delta-gamma approximations and actual portfolio losses when risk factors are multivariate t with $\nu = 5$, based on 10,000 scenarios. The portfolio is the one labeled "0.1yr ATM" in GHS (2000a) and contains 150 options. The approximation in the left panel is based on assuming normally distributed risk factors with volatilities scaled up by $\nu/(\nu-2)$; notice that it greatly underestimates losses in scenarios where the true loss is large. The t-based approximation gives a much better fit.

quadratic approximations. Note that the quadratic approximation, though fairly good in some cases, is as much as 80% to 90% off in other ones. Also, without the (near) exact estimates of $P(L > x)$ there is no way of judging their accuracy for each case, which may be a problem in practice. More accurate estimation of VAR (together with estimates of the error) requires precise revaluation of a portfolio in each market scenario and this can generally only be achieved through Monte Carlo simulation. Because portfolio valuation can be very time consuming, the number of scenarios generated must be kept fairly small.

In GHS (1999c,2000a) we developed a method for strategically sampling scenarios in order to obtain precise estimates of loss probabilities using far fewer scenarios than would be required using standard Monte Carlo sampling. The method in GHS (1999c,2000a) assumes that the market risk factors are normally distributed and uses the delta-gamma approximation to guide the sampling of scenarios. Here, we address the problem of efficient Monte Carlo when the risk factors are modeled by a multivariate t distribution. We use the t based delta-gamma approximation analyzed above as a basis for a combination of *importance sampling* and *stratified sampling*.

The heavy-tailed property of the t distribution has fundamental implications for the design of variance reduction techniques in a simulation. In GHS (1999a, 1999b, 1999c, 2000a), we have demonstrated that importance sampling based on an idea known as *exponential twisting* can be extremely effective in simulations driven by normal random variables. This approach is not, however, applicable to heavy-tailed distributions because it requires the existence of a moment generating function, which is incompatible with the power-law behavior illustrated in (1).

We circumvent this difficulty by applying the earlier mentioned representation result: if X has the multivariate t distribution in (2), then X has the same distribution as $Z/\sqrt{Y/\nu}$, where Z is multivariate normal with covariance matrix Σ and Y is a chi-square random variable with ν degrees of freedom, independent of Z. The advantage of this representation is that Z and Y each permit exponential twisting, though X does not. Our key result for Monte Carlo simulation is the following *exact* relation between the portfolio loss L and the quadratic approximation Q.

Theorem 2. For any x and y, and any θ for which $\phi_{x_c}(\theta) < \infty$, the portfolio loss distribution satisfies

$$P(L > x) = E_\theta \left[e^{-\theta Q_{x_c}} \phi_{x_c}(\theta) I(L > x) \right] \qquad (7)$$

where $I(\cdot)$ denotes the indicator function, $x_c = x - c$, ϕ_x is as in (5), and E_θ denotes expectation under which Z_i are conditionally normal given Y with mean $\theta \Sigma_{ii} b_i \sqrt{Y/\nu}/(1 - 2\theta\lambda_i)$ and variance $\Sigma_{ii}/(1 - 2\theta\lambda_i)$, Y has a gamma distribution with shape parameter $\nu/2$

Portfolio	x	$P(L>x)$	$P(c+Q>x)$	Variance Ratios IS	IS-Strat
0.5yr ATM	311	1.02%	1.17%	53	333
0.1yr ATM	469	0.97%	1.56%	46	134
Delta hedged	617	1.07%	1.69%	42	112
0.25yr OTM	355	1.02%	1.17%	53	242
0.25yr ITM	355	1.02%	1.17%	53	242
Large λ_1	1474	1.10%	1.58%	21	70
Linear λ	3464	1.11%	1.75%	37	100
100, $\rho = 0.0$	4993	1.06%	1.88%	58	346
100, $\rho = 0.2$	5195	1.12%	1.99%	36	158
Index	2019	1.04%	1.22%	26	93

Table 1: Comparison of variance reduction methods based on delta-gamma approximations for heavy-tailed risk factors, using test portfolios from GHS (2000a). The number of risk factors ranges from 10 to 100; the number of options per portfolio ranges from 150 to 2000. All cases use $\nu = 5$. Variance ratios are estimated from 40,000 replications; the stratified estimator uses 40 strata and 1000 samples per stratum. Variance ratios are estimates of the computational speed-up relative to standard Monte Carlo.

and scale parameter $2/(1+2\alpha(\theta))$, and the risk factors are given by $X_i = Z_i/\sqrt{Y/\nu}$.

Ordinary Monte Carlo would estimate the left side of (7) by randomly sampling scenarios from the multivariate t distribution and calculating the fraction of these scenarios in which $L > x$. Using our importance sampling method, we instead estimate the right side of (7) by changing the distribution of the Z and Y in the representation of the risk factors as $X = Z/\sqrt{Y/\nu}$ and evaluating the expression inside the expectation in each scenario. This change of distribution samples large portfolio losses far more frequently than does ordinary Monte Carlo, and this leads to more precise estimates of VAR. We choose the scalar $\theta = \theta_x$ by solving the equation $d\phi_{x_c}(\theta)/d\theta = 0$; this choice puts the average loss near x.

Let $m_2(\theta, x)$ denote the second moment of the importance sampling estimate defined on the right hand side of (7). The next theorem states that if the delta-gamma approximation is exact (and if $\lambda_i > 0$ for all i), then the method satisfies the bounded relative error property. As described in Shahabuddin (1994), this means that only a fixed number of samples are required to estimate $P(L > x)$ to within a specified relative error, no matter how large x is (equivalently, no matter how small $P(L > x)$ is). With standard simulation, the required sample size to obtain a specified relative error grows without bound as $P(L > x) \to 0$.

Theorem 3. If $L = c + Q$ and $\lambda_i > 0$ for all i, then for all sufficiently large x there exist positive constants c_1, c_2 and c_3, such that

$$c_1 x^{-\nu/2} \leq P(L > x) \leq c_2 x^{-\nu/2}, \tag{8}$$

$$m_2(\theta_x, x) \leq c_3 x^{-\nu}. \tag{9}$$

Thus with standard simulation, the second moment, which is simply $P(L > x)$, is of order $x^{-\nu/2}$ whereas the importance sampling estimator has second moment of order $x^{-\nu}$, which is the best possible exponent for any unbiased estimator.

If $L = c + Q$ and $\lambda_i > 0$ for *some* i, then the upper bounds on $P(L > x)$ and $m_2(\theta_x, x)$ in (8) and (9) remain true; in this case, there is a constant c_4 such that

$$m_2(\theta_x, x) \leq c_4 P(L > x) x^{-\nu/2} \tag{10}$$

for large x. Thus the second moment of the importance sampling estimate is greatly reduced in this more general situation as well, although it may not have bounded relative error since the lower bound in (8) has not been established.

To obtain still more variance reduction, we *stratify* the random variable Q_{x_c} appearing in the exponent in (7). For example, to generate N scenarios, we construct N intervals, each having probability $1/N$ for Q_{x_c} and sample one value from each interval. Construction of the bins relies on the transform inversion developed in Section 3 above. This eliminates much of the variance due to the exponent in (7). Moreover, since $c+Q$ is the quadratic approximation to L, stratification can also reduce the variance in the portfolio loss itself.

The last two columns of Table 1 illustrate the effectiveness of the method in estimating the loss probabilities for the portfolios in that table. In particular,

they report the ratio of the variance using standard Monte Carlo relative to using our importance sampling method (IS) and our combination of importance sampling and stratification (IS-Strat). These variance ratios show how many times larger the number of scenarios using standard Monte Carlo has to be to achieve the same precision as the corresponding variance reduction technique. They are thus estimates of the computational speed-up resulting from our methods, with larger ratios indicating greater speed-ups. These examples indicate the potential for enormous speed-ups from our methods. Additional experimental results, that include cases where the marginals of the multivariate t have different degrees of freedom, are reported in GHS (2000b).

If L is the loss then the conditional excess is defined to be $E(L|L > y)$ where y is some fixed constant (maybe the VAR). Both the IS and IS-Strat can easily be adapted to the estimation of this quantity. GHS (2000b) gives theoretical efficiency results and presents experiments.

ACKNOWLEDGMENTS

This research has been supported in part by NSF NYI Award DMI 9457189 and NSF Career Award DMI 9625297.

REFERENCES

Anderson, T.W. 1984. *An Introduction to Multivariate Statistical Analysis*. Second Edition. New York: Wiley.

Blattberg, R.C., and N.J. Gonedes. 1974. A comparison of the stable and student distributions as statistical models for stock prices, *Journal of Business*, 47:244–280.

Bassi, F., P. Embrechts, and M. Kafetzaki. 1998. Risk management and quantile estimation. In *A Practical Guide to Heavy Tails*, ed. R.J. Adler, R. Feldman, and M. Taqqu, 111–130, Boston, MA: Birkhauser.

Britten-Jones, M., and S.M. Schaefer. 1999. Non-linear value-at-risk, *European Finance Review*, 2:161–187.

Duffie, D., and J. Pan. 1999. Analytical value-at-risk with jumps and credit risk, working paper, Graduate School of Business, Stanford University.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 1999a. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance* 9: 117–152.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 1999b. Importance sampling in the Heath-Jarrow-Morton framework. *Journal of Derivatives*, 7:32–50.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 1999c. Variance reduction techniques for estimating value-at-risk. IBM Research Report RC 21577, Yorktown Heights, NY. To appear in *Management Science*.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000a. Importance sampling and stratification for value-at-risk. In *Computational Finance 1999 (Proceedings of the Sixth International Conference on Computational Finance)*, ed. Y. S. Abu-Mostafa, B. LeBaron, A. W. Lo, and A. S. Weigend, 7-24, Cambridge, MA: MIT Press.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000b. Portfolio value-at-risk with heavy-tailed risk factors. IBM Research Report, Yorktown Heights, NY.

Rouvinez, C. 1997. Going Greek with VAR, *Risk*, 10 (Feb):57–65.

Shahabuddin, P. 1994. Importance sampling for the simulation of highly reliable Markovian systems, *Management Science* 40:333–352.

Wilson, T. 1999. Value at risk. In *Risk Management and Analysis, Vol 1*, ed. C. Alexander, 61–124, Chichester, England: Wiley.

AUTHOR BIOGRAPHIES

PAUL GLASSERMAN is a Professor in the Management Science division of the Columbia University Graduate School of Business. Prior to joining the Columbia faculty he was a Member of Technical Staff in the Operations Research department of AT&T Bell Laboratories in Holmdel, NJ. He holds a Ph.D. from Harvard University and an A.B. from Princeton University. His email and web addresses are <pg20@columbia.edu> and <www.gsb.columbia.edu/faculty/directory/gla_pa.html>.

PHILIP HEIDELBERGER received a B.A. in mathematics from Oberlin College in 1974 and a Ph.D. in Operations Research from Stanford University in 1978. He has been a Research Staff Member at the IBM T.J. Watson Research Center since 1978. His email and web addresses are <philiph@us.ibm.com> and <www.research.ibm.com/people/b/berger/>.

PERWEZ SHAHABUDDIN is an Associate Professor in the Industrial Eng. and Operations Research Dept. at Columbia University. From 1990

to 1995, he was a Research Staff Member at the IBM T.J. Watson Research Center. He received his B.Tech in Mechanical Eng. from I.I.T./Delhi, in 1984, worked at Engineers India Limited, India, from 1984 to 1985, and then obtained a M.S. in Statistics and a Ph.D. in Operations Research from Stanford University in 1987 and 1990, respectively. His email and web addresses are <perwez@ieor.columbia.edu> and <www.ieor.columbia.edu/~perwez/>.