Measuring Welfare Effects in Models with Random Coefficients

Erik Meijer* and Jan Rouwendal†

July 20, 2000

SOM-theme F Interactions between consumers and firms

Abstract

In economic research, it is often important to express the marginal value of a variable in monetary terms. This marginal monetary value is the ratio of two partial derivatives of the conditional indirect utility function, which reduces to the ratio of two coefficients if the utility function is linear. Based on the overwhelming evidence of taste differences among people, random coefficient models have become increasingly more popular in recent years. In random coefficient models, the marginal monetary value is the ratio of two random coefficients and is thus random itself. In this paper, we study the distribution of this ratio and particularly the consequences of different distributional assumptions about the coefficients. It is shown both analytically and empirically that important characteristics of the distribution of the marginal monetary value may be sensitive to the distributional assumptions about the random coefficients. The median, however, is much less sensitive than the mean.

The authors would like to thank Ton Steerneman for stimulating discussions and helpful comments.

* Department of Econometrics, University of Groningen, P.O. Box 800, 9700 AV Groningen, THE NETHERLANDS. E-mail: e.meijer@eco.rug.nl.

† Department of Economics, Wageningen Agricultural University, and MuConsult.

1 Introduction

In applied economics one frequently wants to express the value a consumer attaches to (a change in the value of) a particular variable, such as a product attribute or the availability of a public good, in monetary terms. The computation of equivalent and compensating variations is perhaps the best known example of such a procedure. Often attention is focused on the effect of relatively small changes in the variable. In that case the monetary valuation can be approximated by the ratio of two partial derivatives of the utility function: the one in the numerator refers to the particular variable that changes, and the one in the denominator to price or income (or a closely related variable such as expenditures). Although this procedure is simple and routinely carried out, for instance for the evaluation of travel time in transport economic research, it may run into difficulties when the partial derivatives are random variables. These problems are usually ignored by researchers, presumably because they are not aware of them. It is the purpose of this paper to clarify the difficulties and to propose some possibilities of dealing appropriately with them.

It is widely recognized that individuals differ in their evaluation of product attributes and income. Some of this variation can be related to differences in observed characteristics of these individuals. However, in practice there usually remains a residual amount of heterogeneity, which is often substantial. This can be attributed to unobserved heterogeneity among these individuals, which means that it can only be incorporated in the analysis by using random variables. If the random variables are introduced in the utility function as an additive term, this has no effect on the welfare calculations. These calculations use the partial derivatives of the utility function, which remain deterministic. However, it was recognized at least since the end of the seventies that it might be of considerable importance to allow for the possibility that the coefficients in the structural relationship are also random variables. For instance, Hausman and Wise (1978) showed that the restrictive 'independence of irrelevant alternatives' (IIA) property of the multinomial logit model could be avoided by allowing for variations in tastes for the attributes of the choice alternatives among respondents. In their approach, the (conditional indirect) utility function is linear in the parameters, which are assumed to be normally distributed random variables.

Assume that utility is a function of a vector of variables x, with the first component, x_1, a monetary variable, such as price or income. Throughout, we assume that x is nonstochastic. If some elements of x are stochastic, this means that we condition on its realized values. Suppose that we want to express the value the consumer attaches to a change in the j-th variable, Δx_j, in monetary terms. That is, we want to find the change in x_1 that would bring utility back to its initial value (i.e., before the change in x_j occurred). As long as the changes are small, we may write as an approximation:

$$\Delta u = \frac{\partial u}{\partial x_1}\Delta x_1 + \frac{\partial u}{\partial x_j}\Delta x_j \; . \tag{1}$$

If we set Δu equal to zero, this equation can be solved for Δx_1:

$$\Delta x_1 = -\left(\frac{\partial u}{\partial x_j}\bigg/ \frac{\partial u}{\partial x_1}\right)\Delta x_j \; . \tag{2}$$

If one allows for variation in tastes across individuals the partial derivatives $\partial u/\partial x_m$ will in general also become random variables. For instance, Hausman and Wise (1978), use a linear utility function and the partial derivatives are then equal to the random coefficients, which they assume to be normally distributed. However, the ratio of two normal variates is not always well behaved and may be Cauchy distributed which implies that an expected value does not exist. This suggests that even though the parameters of the utility function are successfully estimated and have the expected sign, magnitude, et cetera, it may still be impossible to find a reasonable value of the average monetary valuation that we are interested in for the purposes of welfare economic analysis. This paper addresses this issue.

In section 2, the main issues are discussed on the basis of a simple linear utility function. In section 3, the distribution of the marginal monetary value in a linear utility model with random coefficients is discussed for a number of convenient choices for the distributions of the random coefficients. Section 4 devotes attention to some specification problems in the leading case of a discrete choice model and section 5 gives some results for nonlinear utility models and models with observed heterogeneity. In section 6, the theoretical results are applied to an empirical data set, for which the consequences of different distributional assumptions are studied. Finally, section 7 gives the discussion.

2 Linear utility function with random coefficients

Consider a random utility function that is both linear in variables and linear in parameters, and assume for simplicity that there are only two variables. The utility function can be written as

$$u = \beta x_1 + \gamma x_2 + \varepsilon,$$

where x_1 is a monetary variable, such as price or income, x_2 is another variable of interest, ε is a random error term, and β and γ are random coefficients. The generalization to more variables is straightforward and the extension to nonlinear specifications will be briefly discussed in section 5.

From (2), we have that the marginal monetary value of the attribute x_2 is $-\gamma/\beta$. In a model with fixed coefficients, this value can be easily computed and in many applications, it is routinely reported, for example, the value of time in transportation economic studies. In our case, however, β and γ are random variables, which implies that the coefficient ratio γ/β is also a random variable. Of course, this is in line with every day experience: some people attach more value to travel time than others, some people are willing to pay more for an accessory than others. Generally, the researcher will be interested in the distribution of the ratio, or some properties of that distribution, such as its mean or median, or the proportion of people for which this ratio exceeds some value.

Evidently, the distribution of the coefficient ratio follows from the joint distribution of the coefficients. If β and γ are continuously distributed with joint density function $f(\beta, \gamma)$, then the coefficient ratio $r = \gamma/\beta$ has density function

$$g(r) = \int_{-\infty}^{+\infty} |\beta| \, f(\beta, r\beta) \, \mathrm{d}\beta \tag{3}$$

(cf. Spanos, 1986, p. 103). Traditionally, the normal distribution has been by far the most frequently used distribution for random coefficients. The distribution of the ratio of two normally distributed variables has been discussed by several authors. Its mean and higher moments do not exist and for some parameter values, the density is bimodal. This poses severe problems in presenting and interpreting the results.

One may note, however, that the problems appear not so much to be related to the fact that the partial derivative that refers to the monetary variable is a random variable, but to the possibility that it can be close to zero (i.e. that its density function is positive around zero). This is an unavoidable consequence of the assumption that the parameters of the utility function are normally distributed. This suggests, of course, that the problem should be interpreted as one of model specification.

Economic theory assumes that individuals prefer more income to less and prefer lower prices to higher. This implies that a specification for the density function of the monetary variable should be used that has positive support only on the positive half of the real line, or the negative half for prices. In the following, we assume that the sign of the monetary variable is chosen such that its coefficient should be positive according to economic theory. Moreover, the same holds for other variables as well. People prefer less travel time to more, first class to second class, and so forth. On the other hand, some people prefer to live in the city and others in the country, so not all parameters are restricted in sign. Ideally, the specification of the distributions of the random coefficients allows sign restrictions for some coefficients, but not for others. In the next section, some convenient distributions are studied.

3 Distribution of the coefficient ratio

In this section, the distribution of the coefficient ratio γ/β will be discussed for a number of convenient distributions of the random coefficients. The formula (3) can be used for distributions not discussed here. In many cases, however, it may not be possible to obtain an analytical expression of the integral in this formula. In such cases is will usually be relatively easy to approximate the integral numerically, for example, by quadrature methods, or to simulate from the density.

Normal denominator

The coefficient in the denominator will typically be the coefficient of income, price, or some related variable. Therefore, as argued in the previous section, the sign of this variable should be restricted to be negative (for a price variable) or positive (for an income variable or if the sign of a price variable is reversed). A normal distribution is therefore theoretically inappropriate for this coefficient. Nevertheless, the normal distribution has been used most frequently to model random coefficients. The reasons for this can possibly be traced back to some form of central limit theorem, the popularity of the normal distribution in statistics and econometrics for all sorts of problems, the relatively simple form of the likelihood function for linear models with normally distributed random coefficients, the relatively straightforward way in which explanatory variables can be introduced (cf. section 5), and so forth.

Furthermore, if the mean of the random coefficient is far from zero, relative to its standard deviation, the normal distribution may be a good approximation to other distributions, such as the gamma distribution or the lognormal distribution, and convenience may lead to a preference for the normal distribution as an approximation of the true distribution of the random coefficient.

Given this rationale, the only relevant case to consider is the case in which the numerator is also normally distributed. The distribution of the ratio of two normal random variables has been studied by several authors, mainly in the context of the distribution of the ratio of two means. Its mean, variance, and higher moments do not exist and for some values of the parameters, the distribution is bimodal. The density function is given by Hinkley (1969) and Marsaglia (1965) gives some plots of the possible shapes of the density function.

If β and γ are jointly normally distributed with means μ_1 and μ_2, respectively, variances σ_1^2 and σ_2^2, respectively, and correlation ρ, and if $\mu_1 \gg \sigma_1$, then the distribution of the coefficient ratio $r \equiv \gamma/\beta$ can be approximated by

$$\Pr(r \leq w) \approx \Phi \left\{ \frac{\mu_1 w - \mu_2}{\sqrt{\sigma_1^2 w^2 - 2\sigma_1\sigma_2\rho w + \sigma_2^2}} \right\},$$

where $\Phi(\cdot)$ is the standard normal distribution function (Hinkley, 1969).

Lognormal denominator

It was argued above that the coefficient in the denominator should typically be restricted to be negative (for price), or positive (for income or if the sign of the price variable is reversed). Therefore, it seems appropriate to consider distributions that have only negative or only positive support. Distributions that have only positive support are more common in the literature, and thus we will assume that the sign of the variable is such that the coefficient should be positive.

A distribution that has only positive support and that is very convenient as well, is the lognormal distribution. This is also a well-known distribution in economics, where it is frequently used to approximate income distributions (see, e.g., Aitchison & Brown, 1957). It is obtained from the normal distribution in the following way: If $z \sim \mathcal{N}(\mu, \sigma^2)$, then $x \equiv \exp(z)$ is lognormally distributed: $x \sim \mathrm{LN}(\mu, \sigma^2)$. Its mean is $\exp(\mu + \frac{1}{2}\sigma^2)$ and its variance is $\exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$. It tends to be heavily skewed with a thick tail. The median is $\exp(\mu)$, which may in some cases be the most appropriate characteristic to report. This distribution has been proposed as a convenient distribution for random coefficients in discrete choice models by Revelt and Train (1998) and Train (1998). Kim, Blattberg, and Rossi (1995) compared a lognormal distribution of the negative of the coefficient of log(price) with a semi-nonparametric alternative and found strong support for the lognormal distribution.

From its definition, it is immediately clear that the ratio of two lognormally distributed coefficients is also lognormally distributed: Let $\beta = \exp(\eta_1)$ and $\gamma = \exp(\eta_2)$, where η_1 and η_2 are jointly normally distributed with means μ_1 and μ_2, respectively, variances σ_1^2 and σ_2^2, respectively, and correlation ρ. Then

$$r \equiv \frac{\gamma}{\beta} = \frac{\exp(\eta_2)}{\exp(\eta_1)} = \exp(\eta_2 - \eta_1) = \exp(\tilde{\eta}),$$

where $\tilde{\eta} \equiv \eta_2 - \eta_1$, which is normally distributed with mean $\mu_2 - \mu_1$ and variance $\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$. Hence, it follows that

$$r \sim \mathrm{LN}(\mu_2 - \mu_1, \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2).$$

Consequently, the mean of r is

$$\mathrm{E}(r) = \exp\left[(\mu_2 - \mu_1) + \tfrac{1}{2}(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)\right],$$

its variance is

$$\begin{aligned}\mathrm{Var}(r) = \exp&\left[2(\mu_2 - \mu_1) + 2(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)\right] \\ &- \exp\left[2(\mu_2 - \mu_1) + (\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)\right],\end{aligned}$$

and its median is $\exp(\mu_2 - \mu_1)$.

If the coefficient in the numerator follows a normal distribution, the distribution of the ratio is not as easily expressible. There does not seem to be a closed form expression of its density. The density function can, however, be easily approximated with high precision by using Gaussian quadrature (see, e.g., Press, Teukolsky, Vetterling, & Flannery, 1992, pp. 140–155). The moments of this distribution are easily derived. In particular, let $\beta = \exp(\eta_1)$ and $\gamma = \eta_2$, where η_1 and η_2 are jointly normally distributed with means μ_1 and μ_2, respectively, variances σ_1^2 and σ_2^2, respectively, and correlation ρ. Let $r \equiv \gamma/\beta$. Then the mean and variance of r are

$$\begin{aligned}\mathrm{E}(r) &= \exp(-\mu_1 + \tfrac{1}{2}\sigma_1^2)[\mu_2 - \rho\sigma_1\sigma_2]; \\ \mathrm{Var}(r) &= \exp(-2\mu_1 + 2\sigma_1^2)[\mu_2^2 + \sigma_2^2 - 2\rho^2\sigma_2^2\mu_1 + 4\rho^2\sigma_1^2\sigma_2^2 - 4\rho\sigma_1\sigma_2\mu_2] \\ &\quad - \exp(-2\mu_1 + \sigma_1^2)[\mu_2 - \rho\sigma_1\sigma_2]^2.\end{aligned}$$

Furthermore, simulation from this distribution is also straightforward.

Gamma denominator

Another distribution that has only positive support and that is very convenient is the gamma distribution. A standard gamma variate x with shape parameter $\alpha > 0$ has probability density function

$$g(x;\alpha) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}$$

and cumulative distribution function

$$G(x;\alpha) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t}\,\mathrm{d}t. \tag{4}$$

A two-parameter gamma variate y with shape parameter α and scale parameter τ can be obtained from x as $y = \tau x$. Note, however, that frequently, $\lambda \equiv 1/\tau$ is defined as the scale parameter.

The mean of the two-parameter gamma distribution is $\tau\alpha$ and its variance is $\tau^2\alpha$. The gamma distribution is skewed with a somewhat lighter tail than the lognormal distribution (its kurtosis is smaller for a given skewness).

The gamma distribution is frequently used to model heterogeneity in duration models (e.g., Lancaster, 1990, pp. 65–70) and count data models (e.g., Cameron & Trivedi, 1998, pp. 100–101). In those cases, however, the model contains a factor $\exp(\beta' x)\nu$, where ν is gamma distributed and β is a vector of nonrandom coefficients. Thus, the heterogeneity is additive in the form $\exp(\beta' x + \varepsilon)$, where $\exp(\varepsilon)$ is gamma distributed, whereas in our case the coefficients β themselves are gamma distributed.

A difficulty with the gamma distribution is that there is no natural way in which dependent gamma variates can be defined. It is very likely that different random taste parameters are correlated, and both negative and positive correlations are conceivable. An intuitively appealing way to define multivariate dependent gamma variates is given by Moran (1969). He defines a bivariate gamma distribution as follows. Let z_1 and z_2 be correlated normally distributed random variables with mean zero and variance one, then define two gamma variates x_1 and x_2 as

$$x_1 \equiv \tau_1 G^{-1}\left[\Phi(z_1);\alpha_1\right]; \tag{5a}$$

$$x_2 \equiv \tau_2 G^{-1}\left[\Phi(z_2);\alpha_2\right], \tag{5b}$$

where $G(\cdot\ ;\cdot)$ is defined in (4) and $\Phi(\cdot)$ denotes the standard normal distribution function. Clearly, x_1 and x_2 are dependent and their marginal distributions are two-parameter gamma. Note that this idea can be applied to obtain dependent multivariate distributions with arbitrary marginal distributions (see, e.g., Meijerink, 1996, who developed nonlinear structural equation models as univariate nonlinear transformations from normality). More specifically, if β is assumed to be gamma distributed and γ is assumed to be normally distributed, and possibly β and γ are dependent, they may be assumed to be generated according to

$$\beta \equiv \tau_1 G^{-1}\left[\Phi(z_1);\alpha_1\right]; \tag{6a}$$

$$\gamma \equiv \mu_2 + \sigma_2 z_2. \tag{6b}$$

If β and γ are assumed to be generated according to (5), the density function of the ratio $r \equiv \gamma/\beta$ is not expressible in closed form. The k-th moment of r exists if $k < \alpha_1$, the shape parameter of the distribution of β. The moments are not expressible in closed form, but can be easily approximated with great precision by Gaussian quadrature or simulation. In the special case $\rho = 0$, i.e., r is the ratio of two *independent* gamma variates, the density of r is

$$g(r) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{\psi^{\alpha_2} r^{\alpha_2 - 1}}{(1 + \psi r)^{\alpha_1 + \alpha_2}}, \tag{7}$$

where $\psi \equiv \tau_1/\tau_2$. Hogg and Klugman (1983) call this the *generalized Pareto* distribution. The well-known class of F-distributions is a subset of this class of distributions. If $\rho = 0$, the moments of r are

$$\mathrm{E}(r^k) = \frac{\Gamma(\alpha_1 - k)\Gamma(\alpha_2 + k)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\psi^k},$$

provided $k < \alpha_1$, the shape parameter of the distribution of β. If $k > \alpha_1$, this moment does not exist. In particular, if $\alpha_1 > 2$, the mean and variance of r are

$$\mathrm{E}(r) = \frac{\alpha_2}{(\alpha_1 - 1)\psi};$$
$$\mathrm{Var}(r) = \frac{\alpha_2(\alpha_1 + \alpha_2 - 1)}{(\alpha_1 - 1)^2(\alpha_1 - 2)\psi^2}.$$

If β and γ are assumed to be generated according to (6), the density function of the ratio $r \equiv \gamma/\beta$ is not expressible in closed form. The k-th moment of r exists if $k < \alpha_1$, the shape parameter of the distribution of β. The moments are not expressible in closed form, but can be easily approximated with great precision by Gaussian quadrature or simulation. In the special case $\rho = 0$, i.e., r is the ratio of *independent* normal and gamma variates, the moments of r are easily found from $\mathrm{E}(r^k) = \mathrm{E}(\gamma^k)\,\mathrm{E}(\beta^{-k})$, provided $k < \alpha_1$, the shape parameter of the distribution of β. If $k > \alpha_1$, this moment does not exist. In particular, if $\alpha_1 > 2$, the mean and variance of r are

$$\mathrm{E}(r) = \frac{\mu_2}{\alpha_1 - 1};$$
$$\mathrm{Var}(r) = \frac{\mu_2^2 + (\alpha_1 - 1)\sigma_2^2}{(\alpha_1 - 1)^2(\alpha_1 - 2)}.$$

Latent class approach

The latent class approach is frequently used in marketing applications (e.g., Kamakura & Russell, 1989; Wedel & DeSarbo, 1994). It assumes that there are M types (classes or segments) of consumers, each with its own (nonrandom) taste parameters (β_m, γ_m), $m = 1, \ldots, M$. The proportion of the m-th class in the population is π_m. It is not known which individuals belong to which class. Hence, the distribution of (β, γ) is discrete and $(\beta, \gamma) = (\beta_m, \gamma_m)$ with probability π_m. Consequently, the distribution of $r \equiv \gamma/\beta$ is also discrete, $r = \gamma_m/\beta_m$ with probability π_m, and the mean of r and other characteristics of its distribution can be easily computed using standard formulas for discrete distributions. Note that the randomness of the taste parameters originates from the assumption that the researcher does not know to which class a consumer belongs. If the class is known, the taste parameters follow deterministically.

Latent class methods are extremely useful to divide the market into a number of relatively homogeneous segments, for which specific marketing strategies can be developed (e.g., Wedel & Kamakura, 1998). If the number of classes M is allowed to increase with sample size, this approach is a nonparametric approach to estimation of the distribution of the taste parameters β and γ. In finite samples, however, the estimated discrete distributions tend to be quite coarse and may not cover possibly interesting small niches among the consumers, see, e.g., Allenby, Arora, and Ginter (1998). See also Wedel et al. (1999) for an extensive discussion about discrete and continuous representations of heterogeneity.

To overcome the drawbacks of the standard latent class methods, Allenby et al. (1998) proposed a mixture distribution in which there is random (normal) taste variation *within* each class. The distributions of the random coefficients are generally smooth, whereas the latent class approach is still a special case (with variances equal to zero), as well as normally distributed coefficients (with only one class). Hence, it is easy to compare the results with these special cases. In their examples, Allenby et al. find that the coefficients are generally smoothly nonnormally distributed. In the general case, r is a mixture of ratios of normally distributed variables, which leads to very complicated formulas. We will not discuss these further.

Other distributions

Obviously, many other distributions may be used to model the distributions of the random coefficients. This applies not only to the coefficients that are restricted in sign, but also to the other coefficients. Examples of distributions for positive coefficients are the Weibull distribution and the inverse Gaussian distribution and possible distributions for coefficients that are not restricted in sign are the Gumbel distribution and the t-distribution. For all these distributions, the theoretical and empirical consequences can be studied analogously to the distributions studied above.

If one wants to avoid arbitrary distributional assumptions, the latent class approach gives a nonparametric alternative. As mentioned above, however, the estimated distributions tend to be quite coarse, whereas taste heterogeneity may be expected to be smooth. A useful approach that results in smooth estimates of the density functions is the *semi-nonparametric* approach of Gallant and Nychka (1987), see also Gallant and Tauchen (1989). In this approach, the density function of a general random vector x that is estimated is $f(x) = [P_K(x)]^2 f_0(x)$, where $P_K(x)$ is a polynomial of degree K of the elements of x and $f_0(x)$ is a base density, for example, the density function of a multivariate standard normally distributed vector. In order for f to be a density function, the coefficients of the polynomial P should satisfy some restrictions. Gallant and Nychka showed that, if the degree K of the polynomial is allowed to increase with sample size, the density f is a consistent estimator of (nearly) any smooth density. Sign restrictions are most easily incorporated by taking the exponential of the corresponding element of the vector, as with the lognormal distribution above. A detailed discussion of the application of this approach to random parameter models is given by Davidian and Gallant (1993).

The seminonparametric approach is especially useful as alternative hypothesis to test the adequacy of a given parametric choice density f_0, which corresponds to $K = 0$. For the alternative hypothesis, K is set to a higher value and nonrejection of the null hypothesis implies that the parametric density gives an adequate description of the distribution of the random coefficients (Kim et al., 1995).

Most choices of the distributions of the random coefficients lead to analytically intractable densities of the ratio r with distributional characteristics (moments, quantiles) that have no closed-form expression. Usually, however, it is easy to obtain an arbitrary close approximation by Gaussian quadrature or by simulation from the density. Moreover, as the model will frequently have to be estimated by simulation methods (e.g., Gouriéroux & Monfort, 1991), drawings from the distribution of the random coefficients will be readily available and these can be used to estimate the required characteristics of the distribution of the ratio. One problem is, however, that the moments may not exist, and hence the estimated moments are useless. Therefore, it has to be proved first whether certain moments exist. Fortunately, this is usually relatively easy.

4 Specification issues in a discrete choice model

An important application of random utility models with random coefficients is in discrete choice models. The specific characteristics of discrete choice models, however, invoke additional specification issues. Consider first a binary choice context. An economic agent chooses one option (e.g., buys one product or chooses one behavioral option) from a set of 2 alternatives. It is assumed that the agent attaches a utility u_j to each alternative $j = 1, 2$, and that the alternative with the highest utility is chosen. This choice is observed by the researcher.

The utilities of option 1 and option 2 are

$$u_1 = \beta x_{11} + \gamma x_{21} + \tilde{\varepsilon}_1;$$
$$u_2 = \beta x_{12} + \gamma x_{22} + \tilde{\varepsilon}_2.$$

The researcher only observes whether $u_1 > u_2$ or $u_1 < u_2$, or equivalently, whether

$$u = \beta x_1 + \gamma x_2 + \tilde{\varepsilon}$$

is positive or negative, where $u \equiv u_2 - u_1$, $x_1 \equiv x_{12} - x_{11}$, $x_2 \equiv x_{22} - x_{21}$, and $\tilde{\varepsilon} \equiv \tilde{\varepsilon}_2 - \tilde{\varepsilon}_1$. Write $\tilde{\varepsilon} = \sigma\varepsilon$, where σ is a scale parameter and ε is a random variable with known variance. For example, if $\tilde{\varepsilon}$ is assumed to be normally distributed, ε is conveniently chosen to be standard normally distributed (with variance one), or if $\tilde{\varepsilon}$ is assumed to be logistically distributed, ε is conveniently chosen to be standard logistically distributed (with variance $\pi^2/3$). Hence, the utility function can be written as

$$u = \beta x_1 + \gamma x_2 + \sigma\varepsilon. \tag{8}$$

Because it can only be observed whether $u > 0$ or $u < 0$, the parameters on the right hand side of (8) may be multiplied by an arbitrary positive number. This leads to an equivalent model. Therefore, in applications of binary choice models, σ is usually chosen as one.

It turns out, however, that in random coefficients models, the parameters of the distributions of β and γ frequently diverge to (plus or minus) infinity, especially with so-called *stated preference* or *conjoint choice* data. These are data for which respondents were asked to state their most preferred alternative (product, transportation mode, environmental situation) from a set of hypothetical alternatives that are described by a number of characteristics (attributes).

Revelt and Train (1998), based on a suggestion by Ruud (1996), assumed that the coefficient (β in our case) of the monetary variable is nonrandom, which solved the technical problems. Moreover, this specification is very convenient, because the problems associated with the ratio of two random variables disappear if it can be assumed that the one in the denominator is a (nonzero) constant instead of a random variable. In that case the characteristics of the distribution of the numerator essentially determine all the characteristics of the distribution of the ratio.

This 'solution' is, however, not very satisfactory. A priori it appears at least as likely that the coefficients in the utility function that refer to the monetary variable are random variables, as that those referring to any other variable are. From (8), it can be seen that, if $\sigma \downarrow 0$, β and γ diverge to infinity if the model is scaled to $\sigma = 1$. It seems likely that in stated preference contexts, σ may be close to zero, because the alternatives are completely specified by the given attributes. Hence, differences in choices are mostly due to differences in tastes and not to unobserved characteristics and completely random disturbances, as captured in ε.

If $\sigma = 0$, the model is

$$\begin{aligned} u &= \beta x_1 + \gamma x_2 \\ &= \beta_0 x_1 + \gamma x_2 + (\beta - \beta_0) x_1 \\ &= \beta_0 x_1 + \gamma x_2 + \varepsilon^*, \end{aligned}$$

where $\beta_0 \equiv \mathrm{E}(\beta)$ and $\varepsilon^* \equiv (\beta - \beta_0)x_1$. This equation is of the same form as (8), with the additional restriction that β_0 is nonrandom. In this case, however, the error term is clearly heteroskedastic and if β and γ are not independent, the distributions of γ and ε^* are not independent as well, whereas it is usually assumed that ε^* is i.i.d. across different individuals and independent of γ. Hence, the model with a constant monetary coefficient is different from the model (8). It is well-known that if heteroskedasticity is neglected in logit and probit models, the resulting estimators are inconsistent. In our view, it is better to use (8) as the primary model specification, scaled with $\sigma = 1$. If the estimated parameters of the distribution of β and γ are very large, the model can be reestimated with a different scaling (e.g., with the mean of β equal to plus or minus one), with an estimation algorithm that allows σ to be equal to zero. Then, it can be tested whether $\sigma = 0$.

If the number of alternatives from which the economic agent chooses is larger than two, we arrive at multinomial (or possibly ordered) discrete choice models. The above discussion still largely applies, except that now, different variances of the error terms for different alternatives are identified, except that there remains always at least one arbitrary scaling. Furthermore, the error terms for different alternatives may be correlated. These issues invoke some technical complications, but do not alter the general discussion in this paper.

5 Nonlinearity and observed heterogeneity

The discussion thus far concentrated mainly on utility functions that are both linear in coefficients and linear in variables. Frequently, however, such a bilinear specification is not satisfactory on theoretical and/or empirical grounds. Economic theory and data frequently suggest nonlinear specifications. In such cases, the general ideas of the previous sections can be used to derive the distribution of the marginal monetary value of a variable. The specific distributional results given above, however, will not hold anymore. The formula of interest is (2). General results can not be obtained, but we will now give some illustrative examples.

Kim et al. (1995) derived the utility function

$$u = \beta \log(x_1) + \gamma x_2 + \varepsilon,$$

where x_1 is price and x_2 is quality. For this model, the marginal monetary value is $-(\gamma/\beta)x_1$ and the discussion of the previous sections applies, except that the distributions should be multiplied (for the individual in question) by the observed constant x_1. If x_1 is a price that is common to all individuals, the distribution of the marginal monetary value in the population is the same as the distribution for one individual. If x_1 is income or some other monetary variable that varies over individuals, the distribution of the marginal monetary value in the population is obtained by aggregating over the population distribution of x_1.

If the utility function is

$$u = \beta x_1 + \gamma_1 x_2 + \gamma_2 x_2^2 + \varepsilon,$$

the marginal monetary value is $-(\gamma_1/\beta) - (\gamma_2/\beta)x_2$. Depending on the assumed joint distribution of β, γ_1, and γ_2, this may or may not lead to an easy expression given x_2. The distribution of the marginal monetary value in the population is again obtained by aggregating over x_2.

Apart from *unobserved* heterogeneity operationalized as random coefficients, there is in many cases also a large amount of *observed* heterogeneity in the form of characteristics of the individuals, typically demographic variables. In such cases, it is customary to explain part of the heterogeneity by these characteristics. For example, large cars will be (more) preferred by large households and small cars will be (more) preferred by people who live in the city. Traditionally, such effects are modeled as

$$\beta_j = w'\delta_j + \zeta_j, \tag{9}$$

where β_j is the j-th (random) coefficient of the utility function, w is a vector of explanatory variables (such as person characteristics), δ_j is a vector of (nonrandom) parameters, and ζ_j is a random residual. The effect of the explanatory variables is that of shifting the mean of the distribution of β: this is now $w'\delta_j$, which differs between people, because their characteristics w are different, instead of a constant. This specification is particularly useful if the random coefficients are normally distributed. In that case, $\beta_j \sim \mathcal{N}(w'\delta_j, \sigma^2)$ and the previous observations in section 3 still hold, except that the means are now (deterministic) functions of the exogenous variables.

For other distributions, this specification is not so natural. For example, the specification (9) will generally not guarantee positiveness of the coefficients. For nonnormal distributions, other parametrizations are more natural. With the lognormal distribution, for example, a natural parametrization is

$$\beta_j = \exp(w'\delta_j + \zeta_j) = \exp(w'\delta_j)\exp(\zeta_j), \tag{10}$$

$\beta_j \sim \mathrm{LN}(w'\delta_j, \sigma^2)$ and the previous observations in section 3 still hold, except that the parameters are now (deterministic) functions of the exogenous variables. However, the effect of the explanatory variables is now multiplicative instead of additive.

Apparently, the choice of parametric distribution has also implications for the most natural functional form of the observed heterogeneity in the model. It is, however, possible to specify different functional forms for the same distributions. This is analogous to the problem of the choice of link function in generalized linear models (McCullagh & Nelder, 1989), where the canonical link function is the "natural" link function for a given distribution, but other link functions are still possible.

6 Empirical application

In this section we report estimation results of a number of model specifications with random coefficients. The data we use are stated preferences of 235 respondents, which were obtained in 1987 by Hague Consulting Group for the national Dutch Railways (NS)[1], from experiments in which two options for traveling by train had to be compared. Each respondent made a sequence of choices among two possibilities for traveling by train that differed in some or all of the following attributes: fare, journey time, number of rail-to-rail transfers (interchanges), and comfort level. The sample was composed of persons who had recently traveled from the city of Nijmegen (located in the eastern part of the Netherlands) to Amsterdam, Rotterdam, or The Hague, which are in the Randstad (the core region of the Netherlands located in the western part of the country). The number of decisions differed over the respondents and was on average equal to 12.5. A total number of 2929 choices were registered. These data have also been used in the first case study reported in Ben-Akiva, Bolduc, and Bradley (1993).

Our basic specification is the logit model, which is obtained from an extension of (8) to more variables by normalizing $\sigma = 1$ and by assuming that ε is standard logistically distributed. As is well known, this gives the following formula for the probability that

[1]We thank Theo van der Star of NS Reizigers/MOA Consult for permission to use these data.

alternative 1 is chosen:

$$p_1 = \frac{\exp(\Delta V)}{1 + \exp(\Delta V)},$$

with

$$\Delta V = \beta' \Delta x.$$

In these equations, p_1 denotes the probability that alternative 1 is chosen, ΔV the difference in the systematic part of utility attached to both alternatives (i.e. the systematic part of utility of alternative 1 minus that of alternative 2), which is further specified as a linear function of the differences in the attribute levels mentioned above, Δx. The coefficients β have to be estimated and can be used to compute the implied value of travel time by train. This value of time is β_2/β_1. The probability that alternative 2 is chosen is, of course, $1 - p_1$. The data that we use refer to individuals i, $i = 1, \ldots, 235$. Each individual makes a sequence of K_i choices. We let Δx_{ik} denote the value of Δx for the k-th choice of individual i. We allow for the possibility that the parameters β are random variables, with a distribution that can be characterized by parameters θ. If we let y_{ik} denote an indicator variable that equals 1 if alternative 1 is chosen and 0 otherwise, we can denote the likelihood of the observed sequence of choices made by respondent i as

$$\begin{aligned} \mathcal{L}_i(\theta) &= \int_{\mathcal{D}} \prod_{k=1}^{K_i} \frac{\exp\left(y_{ik}\beta'\Delta x_{ik}\right)}{1 + \exp\left(\beta'\Delta x_{ik}\right)} \, \mathrm{d}F(\beta;\theta) \\ &= \mathrm{E}\left[\prod_{k=1}^{K_i} \frac{\exp\left(y_{ik}\beta'\Delta x_{ik}\right)}{1 + \exp\left(\beta'\Delta x_{ik}\right)} \, ; \theta\right], \end{aligned} \tag{11}$$

where $F(\beta;\theta)$ is the distribution function of β given the parameter vector θ and $\mathcal{D}$ is the domain of β. Because the integral is difficult to compute in a number of cases we want to consider, we approximate the expectation in (11) by simulation from the distribution of β. We do so by taking a large number R of drawings from the distribution $F(\beta;\theta)$. The expectation in (11) is then replaced by the average of the R drawings. Denoting the value of β in the r-th drawing as $\beta_r(\theta)$, thereby recognizing its dependence on the parameters θ, we can write the simulated value of the likelihood as:

$$\mathcal{L}_i^*(\theta) = \frac{1}{R} \sum_{r=1}^{R} \left[\prod_{k=1}^{K_i} \frac{\exp\left[y_{ik}\beta_r(\theta)'\Delta x_{ik}\right]}{1 + \exp\left[\beta_r(\theta)'\Delta x_{ik}\right]}\right]. \tag{12}$$

The simulated loglikelihood of the sample equals the sum of the logarithms of the individual simulated likelihoods given by (12). The parameters θ are estimated by minimizing the value of the simulated loglikelihood of the sample. This estimation method is called *simulated maximum likelihood*. It is asymptotically equivalent to maximum likelihood if the number of replications R is allowed to grow with the sample size N such that $N/R \to 0$ (e.g., Gouriéroux & Monfort, 1991).

Note that our formulation of the likelihood function takes into account explicitly that a number of choices are made by each individual and is different from the likelihood function that would result when all observations were treated as being independent.

Specifications of the distributions of the coefficients

The basic specification of the model refers to the situation in which all parameters are fixed scalars. This standard model will be referred to in what follows as model I. This is the only model to be considered in which there is no need to take into account explicitly the fact that each individual makes a sequence of choices. All the alternatives to the basic model that will be described below relax the assumption of a single value for the coefficients. Three of them do so by using parameterized specifications of the random coefficients. We estimated models in which some or all of the parameters were random variables with a

– normal (model II),

– lognormal (III), or

– gamma density function (IV).

For these three models with different distribution functions we estimated four variants,

(a) cost parameter fixed, all other parameters random, diagonal covariance matrix,

(b) cost parameter fixed, all other parameters random, unrestricted covariance matrix,

(c) all parameters random, diagonal covariance matrix,

(d) all parameters random, unrestricted covariance matrix.

For all these variants our estimator used simulation. For each random parameter we used $R = 250$ independent drawings from the four-variate standard normal distribution as our basic drawings and transformed them to drawings from lognormally or gamma distributed variates when necessary. This procedure was intended to make the results for the various specifications as comparable as possible.

Let z_r, $r = 1, \ldots, R$, be these basic drawings. Then, for the normal model (II), the random coefficients $\beta_r(\theta)$ are computed as

$$\beta_r(\theta) = \mu + Lz_r,$$

where μ is the mean vector of β and L is a lower triangular matrix, the Cholesky root of the covariance matrix Σ of β (i.e., $\Sigma = LL'$). The parameter vector θ consists of the elements of μ and L. The variants (*a*)–(*d*) are obtained by imposing restrictions on L: In (*a*), L is diagonal with its first element zero; in (*b*), the first column of L is zero but the diagonal and subdiagonal elements of the remaining columns are unrestricted; in (*c*), L is diagonal without further restrictions; and in (*d*), L is lower triangular without further restrictions.

For the lognormal model (III), the random coefficients were analogously computed as

$$\beta_r(\theta) = \exp(\mu + Lz_r).$$

For the gamma model (IV), the computation is slightly more complicated, because the transformation is not straightforward. The general formula is (5), with slightly different notation, and extended to a four-variate distribution. The steps of the computation are as follows. First, the uncorrelated standard normal random variables z_r are transformed to correlated normal random variables with mean zero and variance one, according to the formula

$$\tilde{z}_r = DLz_r,$$

where L is a lower triangular matrix that is normalized by setting its diagonal elements to one and D is a diagonal matrix with elements $D_{jj} = [(LL')_{jj}]^{-1/2}$, so that the covariance matrix $\Sigma = DLL'D$ of $\tilde{z}_r$ has diagonal elements equal to one. Second, the random coefficients are computed by solving the univariate equations

$$G[\beta_{rj}(\theta)/\tau_j; \alpha_j] = \Phi(\tilde{z}_{rj}),$$

where Φ denotes the standard normal distribution function and G is defined in (4). Evidently, for the variants (*a*) and (*c*) with independent coefficients, the first step can be omitted and only the second step is performed. For the gamma model, the parameters are the unrestricted subdiagonal elements of L and the parameters τ_j and α_j of the gamma distribution.

Finally, we estimated latent class models (V). In these models the population is treated as a mixture of a finite number of homogeneous classes. For each of these classes a standard logit model is valid. Alternatively, as discussed above, one can regard a latent class model as giving a nonparametric alternative to the three parametric distributions of models II, III, and IV by approximating the unknown distribution of the random coefficients β by a finite number of mass points. For this model, simulation is not needed. The likelihood function $\mathcal{L}_i(\theta)$ given in (11) reduces to a weighted average of the likelihoods of the M latent classes:

$$\mathcal{L}_i(\theta) = \sum_{m=1}^{M} \pi_m \left[\prod_{k=1}^{K_i} \frac{\exp\left(y_{ik}\beta_m' \Delta x_{ik}\right)}{1 + \exp\left(\beta_m' \Delta x_{ik}\right)} \right],$$

where β_m is the vector of parameters of the m-th latent class and π_m is the fraction of the population belonging to that class. The values of the π_m's have to be estimated as well. This can be done by the computationally convenient reparameterization

$$\pi_m = \frac{\exp(\lambda_m)}{\sum_{n=1}^{M} \exp(\lambda_n)},$$

where one of the λ's (the M-th) is set equal to 0, and the others are estimated. Note that also in this case it has been taken into account explicitly that each respondent made a number of choices.

Results

The first striking result is that all likelihood ratio tests for nested models are highly significant ($p < 0.0001$). The values of the likelihood ratio test statistics are given in Table 1. This means that the hypothesis that the cost coefficient is nonrandom and the hypothesis that the different coefficients are independent are rejected. Closer scrutiny of the detailed estimation results gives an indication that the most flexible models (variant *d*) may be a little overparametrized, but this will not be studied in detail, as it is outside the scope of this paper to find the optimal model. (Similarly, the observed heterogeneity in the form of the sex, age, and trip purpose of the respondents are not taken into account.)

Note that the likelihood ratio test statistics are much larger than those found on the same data by Ben-Akiva et al. (1993), who estimated slightly different models, with only one (lognormally distributed) random coefficient. This difference occurs partly because they treat the repeated choices of the same respondent as independent, whereas we treat them as dependent, see (11). Estimation of their model specification with our dependence structure gives likelihood ratio statistics of 65 and 93, respectively, for their Model 1 and Model 2, with one degree of freedom, compared to 3.0 and 5.7 in Ben-Akiva et al. (1993). See appendix A for a detailed discussion of these differences.

A summary of fit statistics (loglikelihood, AIC, and CAIC) is given in table 12 in the appendix. In what follows, we focus on the consequences of the different parametrization for (characteristics of) the value of time (VOT) distribution. Detailed estimation results are documented in the appendix.

Table 1 *Likelihood ratio tests for nested models.*

Test	*Degrees of Freedom*	*Model*		
		Normal (II)	*Lognormal* (III)	*Gamma* (IV)
(a) – Model I	3	384	606	606
(b) – (a)	3	28	116	94
(c) – (a)	1	308	206	206
(d) – (b)	4	322	120	140
(d) – (c)	6	42	30	28

Note. All tests significant at $p < 0.0001$.

Standard logit

In the standard logit model (model I) the coefficients β are interpreted as constants. Table 2 reports the results of estimating this specification of the model. All coefficients have the expected negative sign and are statistically different from zero. The implied value of time is $-0.0287/-0.1483 = 0.19$ Dutch guilder per minute or 11.59 Dutch guilder per hour.

Table 2 *Results for the basic model.*

Variable	*Parameter*	*Standard error*	*Abs. t-value*
Cost	-0.1483	0.0068	21.9
Travel time	-0.0287	0.0026	10.9
Interchanges	-0.3263	0.0591	5.5
Comfort level	-0.9457	0.0655	14.4

Loglikelihood: -1724.15.

Our a priori expectations of the signs of all four coefficients are clearly supported by the estimation results of the standard logit model. There is therefore no need to consider specifications in which some of the coefficients are normal (because they are allowed to have either a positive or negative sign) whereas others are lognormal or gamma (because they are expected to have a definite sign).

Normally distributed parameters

When the cost parameter is fixed and the travel time parameter is distributed normally, the value of time is also normally distributed. The distribution of the value of time when the cost parameter is also normally distributed is discussed in section 3, and for these cases the expected value and variance do not exist. However, the density functions of the implied values of time can be plotted for all four cases. Figure 1 shows the results. There is little difference between the distribution of the value of time that results from estimating the variants in which the cost parameter is fixed (i.e. variants II*a* and II*b*). In these cases the expected value of time is 0.26 and 0.27 Dutch guilder, respectively. This is considerably (more than 30%) higher than the point estimate provided by the standard logit model I. The models II*a* and II*b* also indicate a substantial amount of heterogeneity in the valuation of time: the estimated standard deviations are 0.29 and 0.31, respectively. This implies, among other things, that almost 20% of the valuations are on the negative part of the real line. Whether this is a property of the data or a consequence of the restrictive properties (for the present purposes) of the normal distribution remains to be seen.

It has been noted above that for the ratio of two normally distributed random variables no expected value, variance or higher moments exist. However, we can compare the modes and median values for all four variants. For (*a*) and (*b*) the modes and medians are equal to the expected values given above. For (*c*) and (*d*) considerably lower values of the modes are found: 0.13 and 0.12 respectively. This is more than 30% lower than the value of 0.19 implied by the standard logit model. The medians for (*c*) and (*d*) were estimated by simulation (250 replications) from the estimated distribution. The estimated medians were 0.16 and 0.13, respectively, which are higher than the corresponding modes, but still lower than the values for the standard logit model.

The part of the density shown in the figure (that covers the interval from -0.4 to 1.1) covers 98% for variants II*a* and II*b*, whereas for II*c* and II*d* it covers 90% and 93%, respectively. This is, of course, the pattern one would expect given that the means and variances do not exist in the latter two cases. The striking difference shown in the figure is illustrated by a comparison of the parts of the four densities on the interval (0.0–0.4). These cover 0.49, 0.48, 0.67, and 0.76 for the variants (*a*)–(*d*), respectively. In this sense, the variants with random cost and time parameters clearly imply less heterogeneity in the valuation of time than the other two. It is, moreover, remarkable that negative valuations of time are less common for the variants in which both cost and time parameters are stochastic, and especially so for the clearly asymmetric distribution that results when the full covariance matrix is estimated (variant *d*).

Table 3 shows the correlation matrices that are implied by the estimation results of variants (*b*) and (*d*). The table shows that the correlation coefficients that appear in both matrices are markedly different. This suggests that conclusions about relationships between, for instance, the valuation of time and of comfort level are heavily dependent on the way the cost coefficient is treated. For the data used here, fixing the cost coefficient results in an estimated positive correlation between the coefficients of time and number of interchanges, whereas treating the cost coefficient as random results in a large negative correlation.

We have used the approximation formula derived by Hinkley (1969) and given in section 3 in order to see whether this gives a useful approximation to the densities of the VOT implied by variants (*c*) and (*d*). This was not the case, which is not too surprising given the fact that the approximation was derived for the case in which the expected value of the cost coefficient should be much larger than its standard deviation.

Table 3 *Correlation matrices of variants IIb and IId.*

	Variant (b)				*Variant (d)*			
	Cost	Time	Inter-changes	Comfort level	Cost	Time	Inter-changes	Comfort level
Cost	–				1.00			
Time	–	1.00			−0.30	1.00		
Interchanges	–	0.28	1.00		−0.51	−0.56	1.00	
Comfort level	–	0.29	0.19	1.00	−0.16	0.07	−0.01	1.00

Lognormally distributed parameters

The variants of model III that we estimated correspond to those of model II. In all four variants the implied value of time is a lognormally distributed variable. Table 4 gives the modes, medians, expectations, and standard deviations of these distributions, the densities of which are shown in Figure 2.

Table 4 *Some characteristics of the density of the VOT with lognormally distributed parameters (model III).*

Variant	*Mode*	*Median*	*Expected value*	*Standard deviation*
(*a*)	0.033	0.166	0.372	0.744
(*b*)	0.019	0.164	0.483	1.341
(*c*)	0.044	0.175	0.351	0.611
(*d*)	0.039	0.168	0.349	0.638

The modes are considerably lower than in model II. The median of the value of time

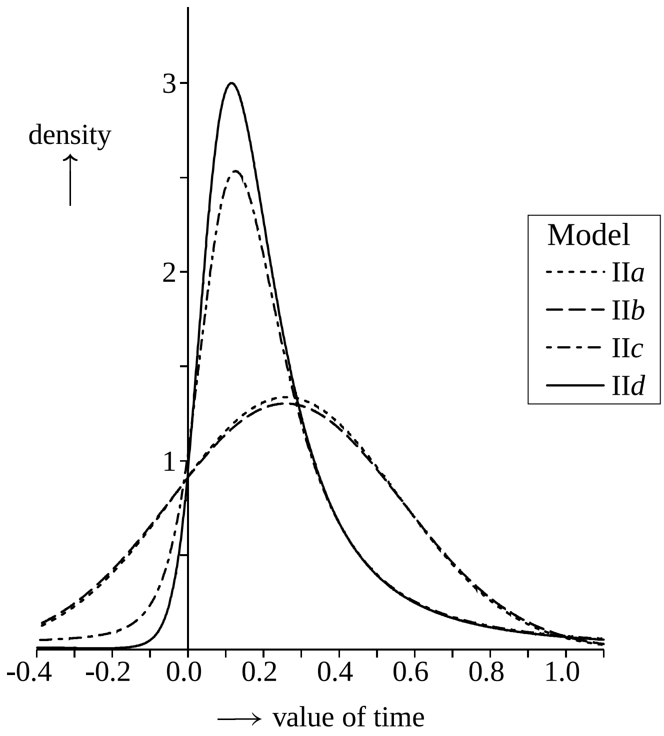

Figure 1 *Density of the value of time with normally distributed parameters.*

that is implied by model III is also somewhat lower than that of the standard model or variants (*a*) and (*b*) of model II. However, it is higher than the medians of variants (*c*) and (*d*) of model II. The expectations of the value of time implied by model III are all much higher than that of the standard logit model and variants (*a*) and (*b*) of model II. The standard deviations are also much higher than variants (*a*) and (*b*) of model II would suggest.

Figure 2 depicts the distribution of the value of time according to the four variants of the lognormal model. The differences between the variants are much smaller than those for the normal model (note that the horizontal axis refers only to the interval 0–0.4).

Gamma distributed parameters

The variants of model IV that we estimated correspond to those of models II and III. The implied value of time is a gamma distributed variable in variants (*a*) and (*b*), is

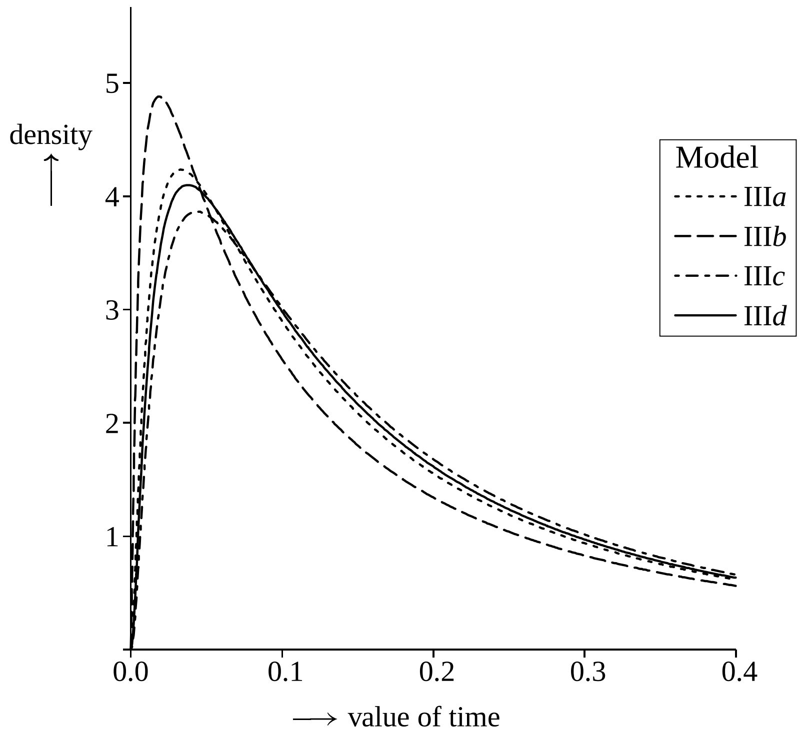

Figure 2 *Density of the value of time with lognormally distributed parameters.*

distributed with density (7) in variant (*c*), and is distributed more complicatedly (the density has no closed form expression) in variant (*d*). Table 5 gives the medians, expectations, and standard deviations of these distributions, the densities of which are shown in Figure 3. The expectations and standard deviations of variants (*a*)–(*c*) were computed analytically by using the formulas given in the text. The expectation of variant (*d*) and the medians were approximated by generating 100,000 replications from the estimated model.

The models with a random cost parameter indicate higher values of time than the models with a fixed cost parameter. Compared with the corresponding characteristics from model III, The medians, means, and standard deviations of model IV tend to be smaller for variants (*a*) and (*b*) and larger for variants (*c*) and (*d*). The differences are, however, relatively small for the medians and considerably larger for the expected values and standard deviations. For variants IV*a* and IV*b*, the medians are much smaller

Table 5 *Some characteristics of the density of the VOT with gamma distributed parameters (model IV).*

Variant	*Median*	*Expected value*	*Standard deviation*
(*a*)	0.147	0.280	0.362
(*b*)	0.158	0.360	0.520
(*c*)	0.179	0.446	does not exist
(*d*)	0.174	0.537	does not exist

than for variants II*a* and II*b*, but the means and standard deviations are (much) larger.

Figure 3 depicts the density of the value of time according to the variants of the gamma model. The differences in shape between the variants are remarkable: The densities of variants (*a*) and (*b*) are monotonically decreasing, whereas the densities of variants (*c*) and (*d*) have a single mode at approximately 0.05.

Latent class approach

We used Akaike's information criterion (AIC) to determine the optimal number of latent classes. This turned out to be 9. The solution is not very satisfactory, however. First, for 9 and 10 latent classes, we had some computational problems that had to be overcome before a solution could be found. Second, some parameters are very large in absolute value and have huge standard errors at the same time. Therefore, in practice, one may probably prefer the 7- or 8-class solutions, despite their slightly larger AIC values.

Table 6 gives the AIC and the average values of time implied by the latent class models with up to 10 classes. Note that the model with one latent class is the standard logit model (model I).

The large increase in the average value of time that is related to the introduction of a third latent class is due to the very high value of time of this new class (5.35 Dutch guilders per minute). When the number of classes is increased further, one class has always a much higher value of time than the others.

For the latent class model with 9 latent classes, the mode and median of the value of time are 0.162, the mean is 0.418 and the standard deviation is 1.034. The median is relatively close to the medians of models III and IV (with lognormal and gamma distributrions). The mean and standard deviation are, however, much larger than the means and standard deviations of most other models.

Figure 4 shows the probability mass function of this model. The general shape is similar to the density functions of the preferred variants of the other models, skewed to the right with a long tail and with a single mode. Note, however, that there is one (small) class with a large negative value of time, which is due to a nonsignificant positive cost

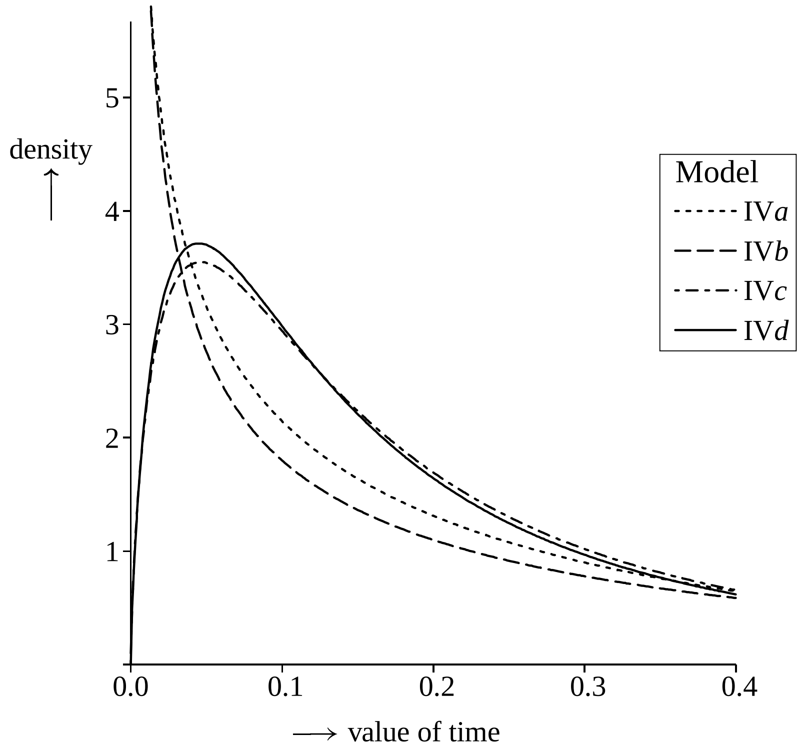

Figure 3 *Density of the value of time with gamma distributed parameters.*

coefficient, which is of course theoretically unrealistic. This illustrates the problems with this solution.

Comparison of the different distributions

The medians of most estimated distributions are relatively close, roughly between 0.16 and 0.18 Dutch guilder per minute. The medians of the normal models, however, vary considerably more, between 0.13 and 0.27. The expectations and standard deviations are much more sensitive to distributional and model specification. For some models, the expectation and standard deviation of the value of time do not exist. For others, the standard deviations do not exist, but the expectations do exist, but for most models, both expectation and standard deviation exist. Among the models in which expectation and standard deviation exist, the estimated values differ greatly.

In Figure 5 the density functions of the value of time as implied by the preferred

Table 6 *Akaike's information criterion and average value of time for the latent class models with 1–10 latent classes (model V).*

Number of classes	*AIC*	*Average value of time*
1	1.1800	0.193
2	1.0625	0.279
3	1.0105	0.659
4	0.9727	0.699
5	0.9454	0.501
6	0.9207	0.561
7	0.9166	0.572
8	0.9136	0.421
9	0.9046	0.418
10	0.9048	0.496

variants of the models II–IV are plotted jointly. The preferred variants are the most flexible versions of the various parametric distributions that have been used.

From this figure, it can be seen that the shapes of the density (or probability mass) functions of the preferred variants of the different models are qualitatively similar, skewed to the right with a long tail and a single mode, but there are considerable differences between model II on the one hand and models III and IV on the other. This is especially important if a researcher or policy maker is interested in the proportion of the population that has a value of time that exceeds a certain threshold, i.e., the proportion of the population that has a (very) small or (very) large value of time. The estimated proportions are sensitive to distributional assumptions, especially small proportions. It is, however, striking how close the densities of models III and IV are.

7 Conclusion

In this paper, we have studied the marginal monetary value of a variable in a microeconomic model. This is defined as (minus) the ratio of the partial derivatives of the conditional indirect utility function with respect to this variable and a monetary variable, respectively. If the conditional indirect utility function is linear in both variables and coefficients, it is (minus) the ratio of the two relevant coefficients.

In a model with random coefficients, this means that the marginal monetary value of a variable is also a random variable. Different assumptions about the joint distribution of the random coefficients lead to a different estimated distribution of the marginal monetary value. Traditionally, random coefficients are typically assumed to

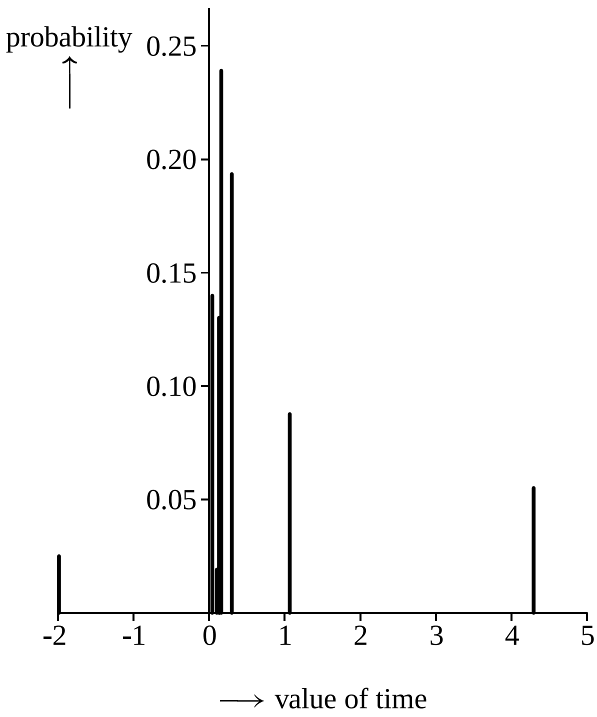

Figure 4 *Probability mass function of the value of time with latent class approach.*

be normally distributed, but in many cases, economic theory and common sense restrict the coefficients in sign. Therefore, distributions that are only defined for positive (or negative) values, such as the lognormal or gamma, may be more appropriate for some coefficients. Nonparametric estimates can be obtained by using a latent class or seminonparametric approach.

To investigate the sensitivity of (characteristics of) the distribution of the marginal monetary value to assumptions about the joint distribution, we estimated mixed logit models for a stated preference data set with different distributions of the coefficients. In the questionnaire, people were (repeatedly) asked to choose a hypothetical train trip from two alternatives. Of key interest was the distribution of the value of time, which is the ratio of the time and cost coefficients.

We used five different distributional models: (I) standard logit, fixed coefficients, (II) normally distributed coefficients, (III) lognormally distributed coefficients, (IV)

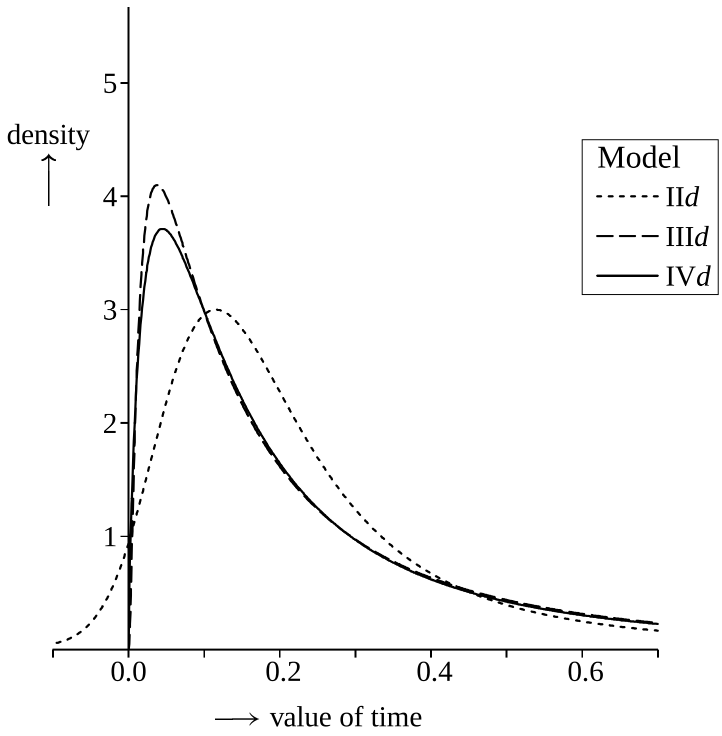

Figure 5 *Densities of the value of time with various distributions of the parameters.*

gamma distributed coefficients, (V) latent class approach, with a discrete distribution of the coefficients, with a finite number of mass points. For models (II)–(IV), we estimated variants with dependent and independent coefficients and with a fixed or random cost coefficient. For model (V), the optimal number of mass points was chosen by using Akaike's information criterion (AIC).

From the empirical results, it follows that the choice of joint distribution of random coefficients may have a large impact on the distribution of the value of time (VOT). Treating the coefficient of the monetary variable as a fixed constant, which has been advocated by several authors, gives markedly different distributions of the VOT and can therefore not be recommended. Allowing all parameters to be random may cause computational problems, however. In many cases, the means of the distributions of the random coefficients diverge to infinity in discrete choice models. As discussed in section 4, this is generally due to (almost) deterministic choices, given the values of the random coefficients. If this occurs, an algorithm and a parametrization should be chosen that allow the variance of the random error to become zero.

In our empirical example, it was also clear that random coefficients of different variables are correlated and that neglecting this correlation may also influence results and thus the conclusions drawn from these. Hence, we recommend that the random coefficients should be allowed to correlate. In some cases, economic theory or empirical evidence may suggest a simpler structure (e.g., Elrod & Keane, 1995; Haaijer, Wedel, Vriens, & Wansbeek, 1998), which may alleviate the computational burden posed by (many) correlated coefficients.

The specific distribution that is chosen also has a clear impact on the distribution of the marginal monetary values. The theoretical consequence may be that the mean and higher moments of this distribution do not exist. In practice, the shape of the distribution may also be different and questions of the form "How many people would be willing to pay an additional amount of X to get an improved alternative Y?" may be answered quite differently for different distributions of the random coefficients. On the other hand, the densities of the preferred gamma and lognormal models are very close, although their expectations are not. Estimated moments are much more sensitive to distributional specification than densities or medians.

Theoretical considerations may imply that signs of coefficients should be restricted. These, combined with empirical evidence and technical and interpretational convenience, may lead to a preferred specification of the joint distribution of the random coefficients. In view of the discussion above, it may be wise to assess the sensitivity of the main results of an empirical study to the distributional assumptions by comparing results using alternative distributions. A comparison with a seminonparametric alternative may also give more insight in the appropriateness of the chosen specification.

References

Aitchison, J., & Brown, J. A. C. (1957). *The lognormal distribution, with special reference to its uses in economics.* Cambridge, UK: Cambridge University Press.

Allenby, G. M., Arora, N., & Ginter, J. L. (1998). On the heterogeneity of demand. *Journal of Marketing Research, 35*, 384–389.

Ben-Akiva, M., Bolduc, D., & Bradley, M. (1993). Estimation of travel choice models with randomly distributed values of time. *Transportation Research Record, 1413*, 88–97.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data.* Cambridge, UK: Cambridge University Press.

Davidian, M., & Gallant, A. R. (1993). The nonlinear mixed effects model with a smooth random effects density. *Biometrika*, *80*, 475–488.

Elrod, T., & Keane, M. P. (1995). A factor-analytic probit model for representing the market structure in panel data. *Journal of Marketing Research*, *32*, 1–16.

Gallant, A. R., & Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica*, *55*, 363–390.

Gallant, A. R., & Tauchen, G. (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica*, *57*, 1091–1120.

Gouriéroux, C., & Monfort, A. (1991). Simulation based inference in models with heterogeneity. *Annales d'Économie et de Statistique*, *20/21*, 69–107.

Haaijer, R., Wedel, M., Vriens, M., & Wansbeek, T. (1998). Utility covariances and context effects in conjoint MNP models. *Marketing Science*, *17*, 236–252.

Hausman, J. A., & Wise, D. A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, *46*, 403–426.

Hinkley, D. V. (1969). On the ratio of two correlated normal variables. *Biometrika*, *56*, 635–639.

Hogg, R. V., & Klugman, S. A. (1983). On the estimation of long tailed skewed distributions with actuarial applications. *Journal of Econometrics*, *23*, 91–102.

Kamakura, W. A., & Russell, G. J. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, *26*, 379–390.

Kim, B.-D., Blattberg, R. C., & Rossi, P. E. (1995). Modeling the distribution of price sensitivity and implications for optimal retail pricing. *Journal of Business & Economic Statistics*, *13*, 291–303.

Lancaster, T. (1990). *The econometric analysis of transition data.* Cambridge, UK: Cambridge University Press.

Marsaglia, G. (1965). Ratios of normal variables and ratios of sums of uniform variables. *Journal of the American Statistical Association*, *60*, 193–204.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

Meijerink, F. (1996). *A nonlinear structural relations model.* Leiden: DSWO Press.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *An introduction to the theory of statistics* (3rd ed.). Singapore: McGraw-Hill.

Moran, P. A. P. (1969). Statistical inference with bivariate gamma distributions. *Biometrika*, *56*, 627–634.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in FORTRAN: The art of scientific computing* (2nd ed.). Cambridge, UK: Cambridge University Press.

Revelt, D., & Train, K. (1998). Mixed logit with repeated choices: Households' choices of appliance efficiency level. *The Review of Economics and Statistics*, *80*, 647–657.

Ruud, P. (1996). *Approximation and simulation of the multinomial probit model: An analysis of covariance matrix estimation* (working paper). Berkeley: Department of Economics, University of California.

Spanos, A. (1986). *Statistical foundations of econometric modelling.* Cambridge, UK: Cambridge University Press.

Train, K. E. (1998). Recreation demand models with taste differences over people. *Land Economics*, *74*, 230–239.

Wedel, M., & DeSarbo, W. S. (1994). A review of recent developments in latent class regression models. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 352–388). Oxford, UK: Blackwell.

Wedel, M., Kamakura, W., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R., Lenk, P., Neslin, S., & Poulsen, C. S. (1999). Discrete and continuous representations of unobserved heterogeneity in choice modeling. *Marketing Letters*, *10*, 219–232.

Wedel, M., & Kamakura, W. A. (1998). *Market segmentation: Conceptual and methodological foundations.* Norwell, MA: Kluwer.

Appendix: Detailed estimation results

The tables are largely self-explanatory, although the meaning of some of the symbols must be inferred from the text. The figures between parentheses are the standard errors of the parameters in the lines immediately above. Only the elements of the matrix L that are not fixed are given in the tables. This matrix was discussed in section 6. It is a lower triangular matrix, so the elements above the diagonal are fixed to zero. Moreover, for variants (a) and (c), it is diagonal, so the elements below the diagonal are also fixed to zero. In addition, for variants (a) and (b), the first column is fixed to zero. For model IV the diagonal is fixed to one.

Table 7 *Results of Model I (Basic model)*

	Cost	Travel time	Interchanges	Comfort level
Fixed	$-.1483$	$-.0287$	$-.3263$	$-.9457$
	(.0068)	(.0026)	(.0591)	(.0655)

Loglikelihood: -1724.

Table 8 *Results of Model II (Normally distributed parameters)*

	Cost	Travel time	Interchanges	Comfort level
Variant (a)				
Fixed	−.3243			
	(.0135)			
μ		−.0843	−.8743	−2.7070
		(.0098)	(.1853)	(.2454)
L		.0968		
		(.0073)		
			1.6009	
			(.1594)	
				−2.7463
				(.2145)
Loglikelihood: -1532				
Variant (b)				
Fixed	−0.3400			
	(0.0140)			
μ		−0.0880	−0.9380	−2.9229
		(0.0102)	(0.2064)	(0.2624)
L		0.1041		
		(0.0078)		
		0.5299	1.8325	
		(0.1875)	(0.1622)	
		0.8382	0.3207	−2.7687
		(0.2544)	(0.1655)	(0.2211)
Loglikelihood: -1518				

Table 8 *Results of Model II (Normally distributed parameters, continued)*

	Cost	Travel time	Interchanges	Comfort level
Variant (c)				
μ	−0.5851	−0.1188	−1.6518	−4.0576
	(0.0394)	(0.0088)	(0.2073)	(0.3105)
L	0.3851			
	(0.0290)			
		0.0830		
		(0.0087)		
			−2.1050	
			(0.1853)	
				−3.4339
				(0.2884)
Loglikelihood: -1378				
Variant (d)				
μ	−0.6346	−0.1211	−1.5285	−3.8793
	(0.0394)	(0.0083)	(0.1830)	(0.3112)
L	0.3380			
	(0.0257)			
	−0.0203	0.0637		
	(0.0064)	(0.0094)		
	−0.8246	−1.2602	−0.6253	
	(0.1959)	(0.2102)	(0.2443)	
	−0.5681	0.0814	0.6937	−3.4202
	(0.2424)	(0.2400)	(0.2955)	(0.2974)
Loglikelihood: -1357				

Table 9 *Results of Model III (Lognormally distributed parameters)*

	Cost	Travel time	Interchanges	Comfort level
Variant (a)				
Fixed (ln)	-0.8692			
	(0.0371)			
μ		-2.6645	-0.9055	1.1468
		(0.1151)	(0.2905)	(0.0840)
L		1.2692		
		(0.0724)		
			1.8279	
			(0.1676)	
				1.4195
				(0.0656)
Loglikelihood: -1421				
Variant (b)				
Fixed (ln)	-0.7148			
	(0.0342)			
μ		-2.5234	-0.5307	1.0410
		(0.1235)	(0.2172)	(0.0863)
L		1.4706		
		(0.0979)		
		0.8932	1.8832	
		(0.0925)	(0.1664)	
		0.6477	0.3490	1.1374
		(0.0580)	(0.0446)	(0.0722)
Loglikelihood: -1363				

Table 9 *Results of Model III (Lognormally distributed parameters, continued)*

	Cost	Travel time	Interchanges	Comfort level
Variant (c)				
μ	−0.4807	−2.2214	−0.0703	1.2279
	(0.0749)	(0.0953)	(0.1783)	(0.0852)
L	0.8195			
	(0.0488)			
		0.8478		
		(0.0735)		
			1.0825	
			(0.1282)	
				1.2234
				(0.0813)
Loglikelihood: -1318				
Variant (d)				
μ	−0.3074	−2.0918	0.0514	1.2201
	(0.1011)	(0.1121)	(0.1851)	(0.1104)
L	0.9312			
	(0.1068)			
	0.1741	0.9582		
	(0.1195)	(0.0743)		
	0.1833	0.1362	1.2580	
	(0.1288)	(0.1040)	(0.1333)	
	0.1993	0.4492	0.3332	1.0782
	(0.1175)	(0.0814)	(0.0692)	(0.0787)
Loglikelihood: -1303				

Table 10 *Results of Model IV (Gamma distributed parameters)*

	Cost	Travel time	Interchanges	Comfort level
Variant (a)				
Fixed	-0.4285			
	(0.0158)			
$-\tau$		-0.2004	-19.5840	-13.7209
		(0.0206)	(4.2948)	(1.5660)
α		0.5981	0.1707	0.5232
		(0.0728)	(0.0277)	(0.0482)
Loglikelihood: -1421				
Variant (b)				
Fixed	-0.4719			
	(0.0156)			
$-\tau$		-0.3545	-10.8104	-11.4205
		(0.0368)	(1.5278)	(1.5269)
α		0.4798	0.2353	0.5168
		(0.0565)	(0.0348)	(0.0537)
L		0.4670		
		(0.0940)		
		0.5720	0.3861	
		(0.1012)	(0.0719)	
Loglikelihood: -1374				

Table 10 *Results of Model IV (Gamma distributed parameters, continued)*

	Cost	Travel time	Interchanges	Comfort level
Variant (c)				
$-\tau$	-0.4417	-0.0794	-8.0434	-4.7868
	(0.0759)	(0.0146)	(1.5288)	(0.8524)
α	1.6823	1.6912	0.3255	1.0772
	(0.2221)	(0.2978)	(0.0598)	(0.1505)
Loglikelihood: -1318				
Variant (d)				
$-\tau$	-0.7182	-0.0979	-9.0201	-5.2471
	(0.1448)	(0.0182)	(1.7205)	(0.8182)
α	1.3644	1.6721	0.3772	0.9828
	(0.1852)	(0.2747)	(0.0621)	(0.1211)
L	0.0855			
	(0.1039)			
	0.1144	0.0422		
	(0.0685)	(0.0523)		
	0.2492	0.3430	0.1578	
	(0.1247)	(0.1336)	(0.1346)	
Loglikelihood: -1304				

Table 11 *Results of Model V (Latent class approach)*

Class	λ	Cost	Travel time	Interchanges	Comfort level
1	1.9201	−0.8803	−0.1162	−0.7436	−0.4800
	(0.7643)	(0.2123)	(0.0357)	(0.5802)	(0.4809)
2	2.5280	−0.3362	−0.0543	−0.6217	−4.3356
	(0.6649)	(0.0321)	(0.0085)	(0.1764)	(0.4433)
3	1.0598	−0.1033	−0.4437	0.5641	−1.8289
	(0.7069)	(0.919)	(0.1335)	(0.7841)	(1.0217)
4	1.7567	−0.2226	−0.0312	−3.8702	−2.2499
	(2.5516)	(0.0302)	(0.0133)	(0.5858)	(0.2828)
5	2.3162	−0.5016	−0.1513	−0.8494	−1.5781
	(0.6771)	(0.0505)	(0.0180)	(0.2671)	(0.2542)
6	1.5241	−0.1471	−0.1572	−2.1660	−7.5264
	(0.6968)	(0.0394)	(0.0386)	(0.4817)	(1.7894)
7	1.9921	−1.2318	−0.0572	−0.7738	−2.6373
	(0.7178)	(0.2789)	(0.0174)	(0.3722)	(0.6875)
8	0.2705	0.0440	−0.0877	−1.8181	−0.8001
	(0.8472)	(0.0694)	(0.0661)	(2.4945)	(2.1620)
9	0	−13.9303	−1.4182	−69.6924	−16.8089
	(–)	(166.28)	(8.2341)	(1098.26)	(89.3170)
Loglikelihood: -1281					

Table 12 *Summary of fit measures of the various models*

Model	Parameters	Loglikelihood	AIC	CAIC
I	4	-1724	1.1800	1.1895
II*a*	7	-1532	1.0509	1.0676
II*b*	10	-1518	1.0434	1.0672
II*c*	8	-1378	0.9464	0.9655
II*d*	14	-1357	0.9362	0.9695
III*a*	7	-1421	0.9751	0.9918
III*b*	10	-1363	0.9375	0.9614
III*c*	8	-1318	0.9054	0.9245
III*d*	14	-1303	0.8993	0.9327
IV*a*	7	-1421	0.9751	0.9918
IV*b*	10	-1374	0.9450	0.9689
IV*c*	8	-1318	0.9054	0.9245
IV*d*	14	-1304	0.9000	0.9333
V (9 classes)	44	-1281	0.9046	1.0096

A Comparison with the results of Ben-Akiva et al.

Ben-Akiva et al. (1993) estimated models with randomly distributed values of time on the same data. However, their specifications differ somewhat from those employed by us, and a brief comparison is appropriate. Ben-Akiva et al. report estimation results for three specifications in their first case study (entitled Intercity Rail SP data). Their 'Fixed VOT' specification is identical to our basic model and so are the results. The model they refer to as 'Lognormal 1' has one lognormally distributed coefficient, the one for travel time. Their 'Lognormal 2' model has three perfectly correlated lognormally distributed coefficients: those for travel time, number of transfers and comfort level. When we tried to reproduce the results reported by Ben-Akiva et al. for these two models by maximizing appropriately restricted variants of our likelihood specification (12), we reached coefficient estimates and loglikelihood values that are different from theirs. However, we were able to reproduce their results by maximizing a somewhat different specification of the likelihood, viz. one that does not take into account the fact that some of the choices are made by the same persons and should therefore be explained by the same preferences. In the approach adopted in this paper, we take the sequence of choices made by a single respondent as the basic element of our likelihood function and when we estimate by simulation, we average over the likelihoods of observing the sequence of choices of this respondent under different drawings. The alternative approach, which reflects the numerical integration procedure of Ben-Akiva et al., would be to regard each choice as basic and to average over the likelihoods of observing these choices under various drawings. This means that instead of $\mathcal{L}_i^*(\theta)$ in (12), one would use

$$\mathcal{L}_i^{**}(\theta) = \prod_{k=1}^{K_i} \left[\frac{1}{R} \sum_{r=1}^{R} \frac{\exp\left[y_{ik}\beta_r(\theta)'\Delta x_{ik}\right]}{1+\exp\left[\beta_r(\theta)'\Delta x_{ik}\right]} \right]. \tag{13}$$

The difference between this specification and (12) can be interpreted as a hypothesis of the stability of trade-offs between the various aspects of train transport of a single individual. Equation (12) assumes that these trade-offs are unknown to the researcher and therefore modelled as random, but stable in the sense that the same trade-offs direct all choices made by the same individual. The formulation (13) assumes that the trade-offs themselves are random in the sense that they differ as much among the various choices made by the same individual as they do among the choices made by different individuals. We can relate likelihood (13) to (12) by means of the following elaboration. First rewrite (13) as

$$\mathcal{L}_i^{**}(\theta) = \prod_{k=1}^{K_i} \left[\frac{1}{R} \sum_{r=1}^{R} p_{ikr} \right],$$

with p_{ikr} implicitly defined. We can write

$$\begin{aligned}\mathcal{L}_i^{**}(\theta) &= \frac{1}{R^{K_i}} \prod_{k=1}^{K_i} \sum_{r=1}^{R} p_{ikr} \\ &= \frac{1}{R^{K_i}} \sum_{\rho(r)} \prod_{k=1}^{K_i} p_{ik\,r(k)},\end{aligned}$$

where $\rho(r)$ denotes any combination $\{r(1), r(2), \ldots, r(K_i)\}$ of K_i integers at least equal to 1 and at most equal to R. The summation proceeds over all these combinations. We denote as $\mathcal{A}$ the set of combinations $\rho(r)$ with the property that all elements are equal. This allows us to split the set of all $\rho(r)$'s into two parts:

$$\begin{aligned}\mathcal{L}_i^{**}(\theta) &= \frac{1}{R^{K_i}} \sum_{\rho(r)\in\mathcal{A}} \prod_{k=1}^{K_i} p_{ik\,r(k)} + \frac{1}{R^{K_i}} \sum_{\rho(r)\notin\mathcal{A}} \prod_{k=1}^{K_i} p_{ik\,r(k)} \\ &= \frac{1}{R^{K_i}} \sum_{r=1}^{R} \prod_{k=1}^{K_i} p_{ikr} + \frac{1}{R^{K_i}} \sum_{\rho(r)\notin\mathcal{A}} \prod_{k=1}^{K_i} p_{ik\,r(k)} \\ &= \frac{1}{R^{K_i-1}} \mathcal{L}_i^{*}(\theta) + \frac{1}{R^{K_i}} \sum_{\rho(r)\notin\mathcal{A}} \prod_{k=1}^{K_i} p_{ik\,r(k)}.\end{aligned}$$

This elaboration shows that it is a priori unclear which of the two formulations will lead to the highest likelihood. Which of the two models should be preferred is therefore an empirical issue. What we find is that formulation (12) results in much higher values of the loglikelihood than (13). Table 13 shows the results for the various models. We interpret these as evidence in favour of the hypothesis that the trade-offs between the various aspects of travel by train are a characteristic of an individual, although they differ among individuals. However, note that the estimates of the various parameters are close to each other, especially for travel cost and time. This is analogous to linear regression analysis under heteroskedasticity, where OLS gives unbiased and consistent estimates, but incorrect standard errors and F-statistics. Further investigation of this issue may be worthwhile, but is outside the scope of the present paper.

Table 13 *Results of alternative model specifications with lognormally distributed parameters.*

	Cost	Travel time	Interchanges	Comfort level
Lognormal 1-independent[a]				
Fixed (ln)	−1.7843		−0.9603	0.0695
	(0.0634)		(0.1812)	(0.0852)
μ		−4.0554		
		(0.2779)		
σ		1.4424[b]		
		(0.3885)		
Loglikelihood: -1721				
Lognormal 1-dependent[a]				
Fixed (ln)	−1.7619		−0.8830	0.1051
	(0.0357)		(0.1416)	(0.0560)
μ		−4.1140		
		(0.2568)		
σ		1.3394[b]		
		(0.1618)		
Loglikelihood: -1659				

Note. See the next page for the notes to this table.

Table 13 *Results of alternative model specifications with lognormally distributed parameters (continued).*

	Cost	Travel time	Interchanges	Comfort level
Lognormal 2-independent[a]				
Fixed (ln)	-1.7159			
	(0.0605)			
μ		-3.8558	-1.4762	-0.2719
		(0.1622)	(0.2152)	(0.1468)
σ		1.4110[b]	[b]	[b]
		(0.2631)		
Loglikelihood: -1719				
Lognormal 2-dependent[a]				
Fixed (ln)	-1.6625			
	(0.0290)			
μ		-3.8802	-1.2978	-0.0663
		(0.1522)	(0.1479)	(0.1336)
σ		1.1453[b]	[b]	[b]
		(0.1028)		
Loglikelihood: -1631				

Notes.

Differences between the estimates reported here and those in Ben-Akiva et al. (1993) are due to a logarithmic transformation, division of the other parameters by that for travel costs and, in Lognormal 2, also by travel time, and their use of an hour as the unit of time. If the appropriate transformations are made, our results are very close to theirs. The largest differences occur in the parameters σ, for which no transformations are needed.

[a] "independent" denotes the specification (13) and "dependent" denotes the specification (12).

[b] The lower triangular matrix L has only one nonzero element in the case of model Lognormal 1, the second diagonal element. In model Lognormal 2 three elements are nonzero, the second diagonal element and the third and fourth element of the second column. However, these elements are restricted to have the same value. The single estimated value is referred to as σ in the table.

B The density of the ratio of two correlated normal variates

Let β and γ be jointly normally distributed with means μ_1 and μ_2, respectively, variances σ_1^2 and σ_2^2, respectively, and correlation ρ. Then, the density of the coefficient ratio $r = \gamma/\beta$ is (Hinkley, 1969)

$$
\begin{aligned}
g(r) = \frac{\sqrt{1-\rho^2}}{\pi\sigma_1\sigma_2 a^2(r)} \exp\left\{-\frac{c}{2(1-\rho^2)}\right\} \\
+ \frac{b(r)d(r)}{\sqrt{2\pi}\,\sigma_1\sigma_2 a^3(r)} \left[\Phi\left\{\frac{b(r)}{a(r)\sqrt{1-\rho^2}}\right\} - \Phi\left\{-\frac{b(r)}{a(r)\sqrt{1-\rho^2}}\right\}\right],
\end{aligned}
$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and

$$
\begin{aligned}
a(r) &= \left(\frac{r^2}{\sigma_2^2} - \frac{2\rho r}{\sigma_1\sigma_2} + \frac{1}{\sigma_1^2}\right)^{1/2}, \\
b(r) &= \frac{\mu_2 r}{\sigma_2^2} - \frac{\rho(\mu_2 + \mu_1 r)}{\sigma_1\sigma_2} + \frac{\mu_1}{\sigma_1^2}, \\
c &= \frac{\mu_2^2}{\sigma_2^2} - \frac{2\rho\mu_1\mu_2}{\sigma_1\sigma_2} + \frac{\mu_1^2}{\sigma_1^2}, \\
d(r) &= \exp\left\{\frac{b^2(r) - ca^2(r)}{2(1-\rho^2)a^2(r)}\right\}.
\end{aligned}
$$

Marsaglia (1965) gives some plots of the possible shapes of the density function. Its mean, variance, and higher moments do not exist and for some values of the parameters, the distribution is bimodal.

C The density of the ratio of a normal variate and a lognormal variate

Let $\beta = \exp(\eta_1)$ and $\gamma = \eta_2$, where η_1 and η_2 are jointly normally distributed with means μ_1 and μ_2, respectively, variances σ_1^2 and σ_2^2, respectively, and correlation ρ. From (3), the coefficient ratio $r = \gamma/\beta$ has density function

$$
g(r) = \int_{-\infty}^{+\infty} |\beta|\, f(\beta, r\beta)\ \mathrm{d}\beta,
$$

where $f(\cdot\,,\cdot)$ is the joint density function of (β, γ). This can in its turn be written as a function of the bivariate normal density function. Let $h(\cdot\,,\cdot)$ be the density of (η_1, η_2):

$$\begin{aligned} h(\eta_1, \eta_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \Bigg\{ - \frac{1}{2(1-\rho^2)} \Bigg[\left(\frac{\eta_1 - \mu_1}{\sigma_1}\right)^2 \\ - 2\rho \left(\frac{\eta_1 - \mu_1}{\sigma_1}\right) \left(\frac{\eta_2 - \mu_2}{\sigma_2}\right) + \left(\frac{\eta_2 - \mu_2}{\sigma_2}\right)^2 \Bigg] \Bigg\} . \end{aligned}$$

Then, the joint density of (β, γ) follows as

$$\begin{aligned} f(\beta, \gamma) = \left|\frac{1}{\beta}\right| \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \Bigg\{ - \frac{1}{2(1-\rho^2)} \Bigg[\left(\frac{\log(\beta) - \mu_1}{\sigma_1}\right)^2 \\ - 2\rho \left(\frac{\log(\beta) - \mu_1}{\sigma_1}\right) \left(\frac{\gamma - \mu_2}{\sigma_2}\right) + \left(\frac{\gamma - \mu_2}{\sigma_2}\right)^2 \Bigg] \Bigg\} . \end{aligned}$$

Hence, the density function of the coefficient ratio is

$$\begin{aligned} g(r) &= \int_{-\infty}^{+\infty} |\beta|\, f(\beta, r\beta)\ \mathrm{d}\beta \\ &= \int_{0}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \Bigg\{ - \frac{1}{2(1-\rho^2)} \Bigg[\left(\frac{\log(\beta) - \mu_1}{\sigma_1}\right)^2 \\ &\quad - 2\rho \left(\frac{\log(\beta) - \mu_1}{\sigma_1}\right) \left(\frac{r\beta - \mu_2}{\sigma_2}\right) + \left(\frac{r\beta - \mu_2}{\sigma_2}\right)^2 \Bigg] \Bigg\}\ \mathrm{d}\beta. \end{aligned}$$

An alternative formula can be obtained as follows. Let the cumulative distribution function of r be $G(R) \equiv \Pr(r \leq R)$. Then, $g(r)$ is obviously the derivative of G

with respect to R. The distribution function can be derived as

$$
\begin{aligned}
G(R) &\equiv \Pr(r \leq R) \\
&= \mathrm{E}_{\eta_1} \Pr(r \leq R \mid \eta_1) \\
&= \mathrm{E}_{\eta_1} \Pr(\eta_2 \leq R \exp(\eta_1) \mid \eta_1) \\
&= \int_{-\infty}^{+\infty} \Phi \left\{ \frac{R \exp(\eta_1) - \mu_2 - \rho\sigma_2 \left(\frac{\eta_1 - \mu_1}{\sigma_1} \right)}{\sigma_2 \sqrt{1-\rho^2}} \right\} \frac{1}{\sigma_1} \phi\left(\frac{\eta_1 - \mu_1}{\sigma_1} \right) \mathrm{d}\eta_1 \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sigma_1} \phi\left(\frac{\eta_1 - \mu_1}{\sigma_1} \right) \\
&\quad \times \Phi \left\{ \frac{1}{\sqrt{1-\rho^2}} \left[\frac{R \exp(\eta_1) - \mu_2}{\sigma_2} - \rho \left(\frac{\eta_1 - \mu_1}{\sigma_1} \right) \right] \right\} \mathrm{d}\eta_1, \qquad (14)
\end{aligned}
$$

where we have used

$$
\eta_2 \mid \eta_1 \sim \mathcal{N} \left(\mu_2 + \rho\sigma_2 \left(\frac{\eta_1 - \mu_1}{\sigma_1} \right),\ \sigma_2^2 (1 - \rho^2) \right).
$$

By differentiating (14), we find the alternative expression for the density of r:

$$
\begin{aligned}
g(r) = \int_{-\infty}^{+\infty} & \frac{\exp(\eta_1)}{\sigma_1 \sigma_2 \sqrt{1-\rho^2}} \phi\left(\frac{\eta_1 - \mu_1}{\sigma_1} \right) \\
& \times \phi \left\{ \frac{1}{\sqrt{1-\rho^2}} \left[\frac{r \exp(\eta_1) - \mu_2}{\sigma_2} - \rho \left(\frac{\eta_1 - \mu_1}{\sigma_1} \right) \right] \right\} \mathrm{d}\eta_1.
\end{aligned}
$$

D The moments of the ratio of a normal variate and a lognormal variate

In this appendix, we will derive the moments of the random variable $r = y/\exp(x)$, where x and y are correlated normally distributed random variables. In order to do so, we need the expectation of the term $z^k \exp(cz)$, where z is a normally distributed variable. More specifically, let $z \sim \mathcal{N}(\mu, \sigma^2)$. Then,

$$
\begin{aligned}
\mathrm{E}\left[z^k \exp(cz)\right] &= \int_{-\infty}^{+\infty} z^k \exp(cz) \frac{1}{\sqrt{2\pi}\,\sigma} \exp \left\{ -\tfrac{1}{2} \left(\frac{z-\mu}{\sigma} \right)^2 \right\} dz \\
&= \exp\left(c\mu + \tfrac{1}{2} c^2 \sigma^2\right) \int_{-\infty}^{+\infty} z^k \frac{1}{\sqrt{2\pi}\,\sigma} \exp \left\{ -\tfrac{1}{2} \left(\frac{z - \mu - c\sigma^2}{\sigma} \right)^2 \right\} dz \\
&= \exp\left(c\mu + \tfrac{1}{2} c^2 \sigma^2\right) \mathrm{E}(\tilde{z}^k),
\end{aligned}
$$

where $\tilde{z} \sim \mathcal{N}(\mu + c\sigma^2, \sigma^2)$. This derivation follows the derivation of the moment generating function of the normal distribution in Mood, Graybill, and Boes (1974, pp. 109–110). From the moments of the normal distribution, it follows that

$$\begin{aligned}
\mathrm{E}(\tilde{z}^0) &= 1, \\
\mathrm{E}(\tilde{z}^1) &= \mu + c\sigma^2, \\
\mathrm{E}(\tilde{z}^2) &= (\mu + c\sigma^2)^2 + \sigma^2, \\
\mathrm{E}(\tilde{z}^3) &= (\mu + c\sigma^2)^3 + 3(\mu + c\sigma^2)\sigma^2, \\
\mathrm{E}(\tilde{z}^4) &= (\mu + c\sigma^2)^4 + 6(\mu + c\sigma^2)^2\sigma^2 + 3\sigma^4.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathrm{E}[\exp(cz)] &= \exp\left(c\mu + \tfrac{1}{2}c^2\sigma^2\right), \\
\mathrm{E}[z^1\exp(cz)] &= \exp\left(c\mu + \tfrac{1}{2}c^2\sigma^2\right)(\mu + c\sigma^2), \\
\mathrm{E}[z^2\exp(cz)] &= \exp\left(c\mu + \tfrac{1}{2}c^2\sigma^2\right)[(\mu + c\sigma^2)^2 + \sigma^2], \\
\mathrm{E}[z^3\exp(cz)] &= \exp\left(c\mu + \tfrac{1}{2}c^2\sigma^2\right)[(\mu + c\sigma^2)^3 + 3(\mu + c\sigma^2)\sigma^2], \\
\mathrm{E}[z^4\exp(cz)] &= \exp\left(c\mu + \tfrac{1}{2}c^2\sigma^2\right)[(\mu + c\sigma^2)^4 + 6(\mu + c\sigma^2)^2\sigma^2 + 3\sigma^4].
\end{aligned}$$

Now, we are ready to derive the moments of $r = y/\exp(x)$, where

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_x\sigma_y\rho_{xy} \\ \sigma_x\sigma_y\rho_{xy} & \sigma_y^2 \end{pmatrix}\right].$$

Define $z = -x$, so $z \sim \mathcal{N}(-\mu_x, \sigma_x^2)$. Then, y can be written as

$$y = a + bz + \varepsilon,$$

where $a = \mu_y + b\mu_x$, $b = -\rho_{xy}\sigma_y/\sigma_x$, and $\varepsilon \sim \mathcal{N}(0, \sigma_y^2(1-\rho_{xy}^2))$, which is independent of z. It follows that

$$r = y\exp(z) = a\exp(z) + bz\exp(z) + \varepsilon\exp(z).$$

Thus, the moments of r can be expressed as sums of terms of the form $d\,\mathrm{E}(\varepsilon^k)\,\mathrm{E}[z^l\exp(mz)]$; $\mathrm{E}(\varepsilon^k)$ follows from the moments of the normal distribution and $\mathrm{E}[z^l\exp(mz)]$ has been derived above. The expressions for the first four moments of r

are:

$$\begin{aligned}
\mathrm{E}(r) &= \exp(-\mu_x + \tfrac{1}{2}\sigma_x^2)[\mu_y - \rho_{xy}\sigma_x\sigma_y];\\
\mathrm{E}(r^2) &= \exp(-2\mu_x + 2\sigma_x^2)[\mu_y^2 + \sigma_y^2 - 2\rho_{xy}^2\sigma_y^2\mu_x + 4\rho_{xy}^2\sigma_x^2\sigma_y^2 - 4\rho_{xy}\sigma_x\sigma_y\mu_y];\\
\mathrm{E}(r^3) &= \exp(-3\mu_x + \tfrac{9}{2}\sigma_x^2)[\mu_y(\mu_y^2 + \sigma_y^2) + 9\rho_{xy}\sigma_x\sigma_y(\mu_y^2 + \sigma_y^2) + 27\rho_{xy}^2\sigma_x^2\sigma_y^2\mu_y\\
&\quad + 27\sigma_x^3\sigma_y^3];\\
\mathrm{E}(r^4) &= \exp(-4\mu_x + 8\sigma_x^2)[\mu_y^4 + 6\mu_y^2\sigma_y^2 + 3\sigma_y^4 + 16\rho_{xy}\sigma_x\sigma_y\mu_y(\mu_y^2 + 3\sigma_y^2)\\
&\quad + 96\rho_{xy}^2\sigma_x^2\sigma_y^2(\mu_y^2 + \sigma_y^2) + 256\rho_{xy}^3\sigma_x^3\sigma_y^3(\mu_y + \rho_{xy}\sigma_x\sigma_y)].
\end{aligned}$$

Hence, the variance of r is

$$\begin{aligned}
\mathrm{Var}(r) &\equiv \mathrm{E}(r^2) - (\mathrm{E}(r))^2\\
&= \exp(-2\mu_x + 2\sigma_x^2)[\mu_y^2 + \sigma_y^2 - 2\rho_{xy}^2\sigma_y^2\mu_x + 4\rho_{xy}^2\sigma_x^2\sigma_y^2 - 4\rho_{xy}\sigma_x\sigma_y\mu_y]\\
&\quad - \exp(-2\mu_x + \sigma_x^2)[\mu_y - \rho_{xy}\sigma_x\sigma_y]^2.
\end{aligned}$$

The skewness and kurtosis can also be computed using the above formulas.

E The density of the ratio of two jointly bivariate gamma variates

Let $\beta = \tau_1 G^{-1}[\Phi(\eta_1); \alpha_1]$ and $\gamma = \tau_2 G^{-1}[\Phi(\eta_2); \alpha_2]$, where η_1 and η_2 are jointly normally distributed with means zero, variances one, and correlation ρ. Let $h(\cdot, \cdot)$ be the density of (η_1, η_2):

$$\begin{aligned}
h(\eta_1, \eta_2) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\eta_1^2 - 2\rho\eta_1\eta_2 + \eta_2^2\right]\right)\\
&= \frac{1}{\sqrt{1-\rho^2}}\phi(\eta_1)\,\phi\left(\frac{\eta_2 - \rho\eta_1}{\sqrt{1-\rho^2}}\right),
\end{aligned}$$

where $\phi(\cdot)$ is the standard normal density function. The jacobian of the transformation is

$$\begin{aligned}
J &= \frac{\partial\eta_1}{\partial\beta}\frac{\partial\eta_2}{\partial\gamma}\\
&= \frac{\partial\Phi^{-1}\left[G\left(\frac{\beta}{\tau_1}; \alpha_1\right)\right]}{\partial\beta}\frac{\partial\Phi^{-1}\left[G\left(\frac{\gamma}{\tau_2}; \alpha_2\right)\right]}{\partial\gamma}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\partial \Phi^{-1}(u_1)}{\partial u_1} \frac{\partial G(t;\alpha_1)}{\partial t_1} \frac{1}{\tau_1} \frac{\partial \Phi^{-1}(u_2)}{\partial u_2} \frac{\partial G(t;\alpha_2)}{\partial t_2} \frac{1}{\tau_2} \\
&= \frac{1}{\phi\left[\Phi^{-1}(u_1)\right]} \left(\frac{1}{\Gamma(\alpha_1)} t_1^{\alpha_1 - 1} e^{-t_1} \right) \frac{1}{\tau_1} \frac{1}{\phi\left[\Phi^{-1}(u_2)\right]} \left(\frac{1}{\Gamma(\alpha_2)} t_2^{\alpha_2 - 1} e^{-t_2} \right) \frac{1}{\tau_2} \\
&= \frac{1}{\phi\left\{ \Phi^{-1}\left[G\left(\frac{\beta}{\tau_1}; \alpha_1 \right) \right] \right\}} \frac{1}{\Gamma(\alpha_1)} \left(\frac{\beta}{\tau_1} \right)^{\alpha_1 - 1} \exp\left(- \frac{\beta}{\tau_1} \right) \frac{1}{\tau_1} \\
&\quad \times \frac{1}{\phi\left\{ \Phi^{-1}\left[G\left(\frac{\gamma}{\tau_2}; \alpha_2 \right) \right] \right\}} \frac{1}{\Gamma(\alpha_2)} \left(\frac{\gamma}{\tau_2} \right)^{\alpha_2 - 1} \exp\left(- \frac{\gamma}{\tau_2} \right) \frac{1}{\tau_2},
\end{aligned}
$$

where $t_1 \equiv \beta/\tau_1$, $t_2 \equiv \gamma/\tau_2$, $u_1 \equiv G(t_1;\alpha_1)$, and $u_2 \equiv G(t_2;\alpha_2)$ have been used in the intermediate equalities. Hence, the probability density function of (β, γ) is

$$
\begin{aligned}
f(\beta,\gamma) &\equiv |J|\, h[\eta_1(\beta), \eta_2(\gamma)] \\
&= \frac{1}{\tau_1 \tau_2 \Gamma(\alpha_1)\Gamma(\alpha_2)\sqrt{1-\rho^2}} \left(\frac{\beta}{\tau_1} \right)^{\alpha_1 - 1} \left(\frac{\gamma}{\tau_2} \right)^{\alpha_2 - 1} \exp\left(- \frac{\beta}{\tau_1} - \frac{\gamma}{\tau_2} \right) \\
&\quad \times \frac{\phi\left\{ \dfrac{\Phi^{-1}\left[G\left(\frac{\gamma}{\tau_2}; \alpha_2 \right) \right] - \rho \Phi^{-1}\left[G\left(\frac{\beta}{\tau_1}; \alpha_1 \right) \right]}{\sqrt{1-\rho^2}} \right\}}{\phi\left\{ \Phi^{-1}\left[G\left(\frac{\gamma}{\tau_2}; \alpha_2 \right) \right] \right\}}.
\end{aligned}
$$

Consequently, the probability density function of $r = \gamma/\beta$ is

$$
\begin{aligned}
g(r) &\equiv \int_0^{+\infty} \beta f(\beta, r\beta)\, \mathrm{d}\beta \\
&= \int_0^{+\infty} \beta \frac{1}{\tau_1 \tau_2 \Gamma(\alpha_1)\Gamma(\alpha_2)\sqrt{1-\rho^2}} \left(\frac{\beta}{\tau_1}\right)^{\alpha_1 - 1} \left(\frac{r\beta}{\tau_2}\right)^{\alpha_2 - 1} \exp\left(-\frac{\beta}{\tau_1} - \frac{r\beta}{\tau_2}\right) \\
&\quad \times \frac{\phi\left\{\dfrac{\Phi^{-1}\left[G\left(\dfrac{r\beta}{\tau_2}; \alpha_2\right)\right] - \rho\Phi^{-1}\left[G\left(\dfrac{\beta}{\tau_1}; \alpha_1\right)\right]}{\sqrt{1-\rho^2}}\right\}}{\phi\left\{\Phi^{-1}\left[G\left(\dfrac{r\beta}{\tau_2}; \alpha_2\right)\right]\right\}}\, \mathrm{d}\beta \\
&= \frac{\psi^{\alpha_2} r^{\alpha_2 - 1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\sqrt{1-\rho^2}} \int_0^{+\infty} w^{\alpha_1 + \alpha_2 - 1} \exp\left[-(1+\psi r)w\right] \\
&\quad \times \frac{\phi\left\{\dfrac{\Phi^{-1}\left[G(\psi r w; \alpha_2)\right] - \rho\Phi^{-1}\left[G(w; \alpha_1)\right]}{\sqrt{1-\rho^2}}\right\}}{\phi\left\{\Phi^{-1}\left[G(\psi r w; \alpha_2)\right]\right\}}\, \mathrm{d}w,
\end{aligned}
$$

where $\psi \equiv \tau_1/\tau_2$. In the special case $\rho = 0$, this reduces to

$$
\begin{aligned}
g(r) &= \frac{\psi^{\alpha_2} r^{\alpha_2 - 1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^{+\infty} w^{\alpha_1 + \alpha_2 - 1} \exp\left[-(1+\psi r)w\right]\, \mathrm{d}w \\
&= \frac{\psi^{\alpha_2} r^{\alpha_2 - 1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)(1+\psi r)^{\alpha_1+\alpha_2}} \int_0^{+\infty} v^{\alpha_1 + \alpha_2 - 1} e^{-v}\, \mathrm{d}v \\
&= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{\psi^{\alpha_2} r^{\alpha_2 - 1}}{(1+\psi r)^{\alpha_1+\alpha_2}}.
\end{aligned}
$$

Hogg and Klugman (1983) call this the *generalized Pareto* distribution. The well-known class of F-distributions is a subset of this class of distributions. Note that, because $g(r)$ must integrate to one, this implicitly proves

$$
\int_0^{+\infty} \frac{\psi^{\alpha_2} r^{\alpha_2 - 1}}{(1+\psi r)^{\alpha_1+\alpha_2}}\, \mathrm{d}r = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}.
$$

Hence, if $\rho = 0$, the moments of r are

$$\begin{aligned}
\mathrm{E}(r^k) &\equiv \int_0^{+\infty} r^k g(r)\,\mathrm{d}r \\
&= \int_0^{+\infty} r^k \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{\psi^{\alpha_2} r^{\alpha_2-1}}{(1+\psi r)^{\alpha_1+\alpha_2}}\,\mathrm{d}r \\
&= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\psi^k} \int_0^{+\infty} \frac{\psi^{k+\alpha_2} r^{k+\alpha_2-1}}{(1+\psi r)^{\alpha_1+\alpha_2}}\,\mathrm{d}r \\
&= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\psi^k} \frac{\Gamma(\alpha_1-k)\Gamma(\alpha_2+k)}{\Gamma(\alpha_1+\alpha_2)} \\
&= \frac{\Gamma(\alpha_1-k)\Gamma(\alpha_2+k)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\psi^k},
\end{aligned}$$

provided $k < \alpha_1$, the shape parameter of the distribution of β. If $k > \alpha_1$, this moment does not exist. In particular,

$$\begin{aligned}
\mathrm{E}(r) &= \frac{\Gamma(\alpha_1-1)\Gamma(\alpha_2+1)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\psi} \\
&= \frac{\alpha_2}{(\alpha_1-1)\psi}; \\
\mathrm{E}(r^2) &= \frac{\Gamma(\alpha_1-2)\Gamma(\alpha_2+2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\psi^2} \\
&= \frac{\alpha_2(\alpha_2+1)}{(\alpha_1-1)(\alpha_1-2)\psi^2},
\end{aligned}$$

and the variance of r is

$$\mathrm{Var}(r) = \frac{\alpha_2(\alpha_1+\alpha_2-1}{(\alpha_1-1)^2(\alpha_1-2)\psi^2}.$$

F The density of the ratio of a normal variate and a gamma variate

Let $\beta = \tau_1 G^{-1}[\Phi(\eta_1);\alpha_1]$ and $\gamma = \mu_2 + \sigma_2\eta_2$, where η_1 and η_2 are jointly normally distributed with means zero, variances one, and correlation ρ. Let $h(\cdot\,,\cdot)$ be the density of (η_1, η_2):

$$\begin{aligned}
h(\eta_1,\eta_2) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\eta_1^2 - 2\rho\eta_1\eta_2 + \eta_2^2\right]\right\} \\
&= \frac{1}{\sqrt{1-\rho^2}}\phi(\eta_1)\,\phi\left(\frac{\eta_2-\rho\eta_1}{\sqrt{1-\rho^2}}\right),
\end{aligned}$$

where $\phi(\cdot)$ is the standard normal density function. The jacobian of the transformation is

$$
\begin{aligned}
J &= \frac{\partial \eta_1}{\partial \beta}\frac{\partial \eta_2}{\partial \gamma} \\
&= \frac{\partial \Phi^{-1}\left[G\left(\frac{\beta}{\tau_1}; \alpha_1\right)\right]}{\partial \beta} \frac{\partial \frac{\gamma - \mu_2}{\sigma_2}}{\partial \gamma} \\
&= \frac{\partial \Phi^{-1}(u)}{\partial u} \frac{\partial G(t; \alpha_1)}{\partial t} \frac{1}{\tau_1}\frac{1}{\sigma_2} \\
&= \frac{1}{\phi\big[\Phi^{-1}(u)\big]} \left(\frac{1}{\Gamma(\alpha_1)} t^{\alpha_1 - 1} e^{-t}\right) \frac{1}{\tau_1}\frac{1}{\sigma_2} \\
&= \frac{1}{\phi\left\{\Phi^{-1}\left[G\left(\frac{\beta}{\tau_1}; \alpha_1\right)\right]\right\}} \frac{1}{\Gamma(\alpha_1)} \left(\frac{\beta}{\tau_1}\right)^{\alpha_1 - 1} \exp\left(-\frac{\beta}{\tau_1}\right) \frac{1}{\tau_1}\frac{1}{\sigma_2},
\end{aligned}
$$

where $t \equiv \beta/\tau_1$ and $u \equiv G(t; \alpha_1)$ have been used in the intermediate equalities. Hence, the probability density function of (β, γ) is

$$
\begin{aligned}
f(\beta, \gamma) &\equiv |J| h\left[\eta_1(\beta), \eta_2(\gamma)\right] \\
&= \frac{1}{\phi\left\{\Phi^{-1}\left[G\left(\frac{\beta}{\tau_1}; \alpha_1\right)\right]\right\}} \frac{1}{\Gamma(\alpha_1)} \left(\frac{\beta}{\tau_1}\right)^{\alpha_1 - 1} \exp\left(-\frac{\beta}{\tau_1}\right) \frac{1}{\tau_1}\frac{1}{\sigma_2}\frac{1}{\sqrt{1-\rho^2}} \\
&\quad \times \phi\left\{\Phi^{-1}\left[G\left(\frac{\beta}{\tau_1}; \alpha_1\right)\right]\right\} \phi\left\{\frac{\frac{\gamma - \mu_2}{\sigma_2} - \rho \Phi^{-1}\left[G\left(\frac{\beta}{\tau_1}; \alpha_1\right)\right]}{\sqrt{1-\rho^2}}\right\} \\
&= \frac{1}{\sigma_2 \tau_1 \Gamma(\alpha_1)\sqrt{1-\rho^2}} \left(\frac{\beta}{\tau_1}\right)^{\alpha_1 - 1} \exp\left(-\frac{\beta}{\tau_1}\right) \\
&\quad \times \phi\left\{\frac{\gamma - \mu_2 - \rho\sigma_2 \Phi^{-1}\left[G\left(\frac{\beta}{\tau_1}; \alpha_1\right)\right]}{\sigma_2\sqrt{1-\rho^2}}\right\}.
\end{aligned}
$$

Consequently, the probability density function of $r = \gamma/\beta$ is

$$
\begin{aligned}
g(r) &\equiv \int_0^{+\infty} \beta f(\beta, r\beta)\,\mathrm{d}\beta \\
&= \int_0^{+\infty} \beta \frac{1}{\sigma_2 \tau_1 \Gamma(\alpha_1)\sqrt{1-\rho^2}} \left(\frac{\beta}{\tau_1}\right)^{\alpha_1 - 1} \exp\left(-\frac{\beta}{\tau_1}\right) \\
&\quad \times \phi\left[\frac{r\beta - \mu_2 - \rho\sigma_2 \Phi^{-1}\left[G\left(\frac{\beta}{\tau_1}; \alpha_1\right)\right]}{\sigma_2\sqrt{1-\rho^2}}\right] \mathrm{d}\beta \\
&= \frac{\tau_1}{\sigma_2 \Gamma(\alpha_1)\sqrt{1-\rho^2}} \int_0^{+\infty} w^{\alpha_1} e^{-w}\, \phi\left\{\frac{\tau_1 r w - \mu_2 - \rho\sigma_2\Phi^{-1}[G(w;\alpha_1)]}{\sigma_2\sqrt{1-\rho^2}}\right\}\,\mathrm{d}w.
\end{aligned}
$$

In the special case $\rho = 0$, this reduces to

$$
g(r) = \frac{\tau_1}{\sigma_2 \Gamma(\alpha_1)} \int_0^{+\infty} w^{\alpha_1} e^{-w}\, \phi\left\{\frac{\tau_1 r w - \mu_2}{\sigma_2}\right\}\,\mathrm{d}w.
$$

Because the numerator and denominator are independent in this case, the moments are easily found from $\mathrm{E}(r^k) = \mathrm{E}(\gamma^k)\,\mathrm{E}(\beta^{-k})$, provided $k < \alpha_1$, the shape parameter of the distribution of β. If $k > \alpha_1$, this moment does not exist. In particular,

$$
\begin{aligned}
\mathrm{E}(r) &= \mathrm{E}(\gamma)\,\mathrm{E}(\beta^{-1}) \\
&= \mu_2 \frac{\Gamma(\alpha_1 - 1)}{\Gamma(\alpha_1)} \\
&= \frac{\mu_2}{\alpha_1 - 1}; \\
\mathrm{E}(r^2) &= \mathrm{E}(\gamma^2)\,\mathrm{E}(\beta^{-2}) \\
&= (\mu_2^2 + \sigma_2^2)\,\frac{\Gamma(\alpha_1 - 2)}{\Gamma(\alpha_1)} \\
&= \frac{\mu_2^2 + \sigma_2^2}{(\alpha_1 - 1)(\alpha_1 - 2)},
\end{aligned}
$$

and, consequently, the variance of r is

$$
\mathrm{Var}(r) = \frac{\mu_2^2 + (\alpha_1 - 1)\sigma_2^2}{(\alpha_1 - 1)^2(\alpha_1 - 2)}.
$$