



A spline function class suitable for demand models

Rich, Jeppe

Published in:
Econometrics and Statistics

Link to article, DOI:
[10.1016/j.ecosta.2018.02.002](https://doi.org/10.1016/j.ecosta.2018.02.002)

Publication date:
2020

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Rich, J. (2020). A spline function class suitable for demand models. *Econometrics and Statistics*, 14, 24-37.
<https://doi.org/10.1016/j.ecosta.2018.02.002>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Accepted Manuscript

A spline function class suitable for demand models

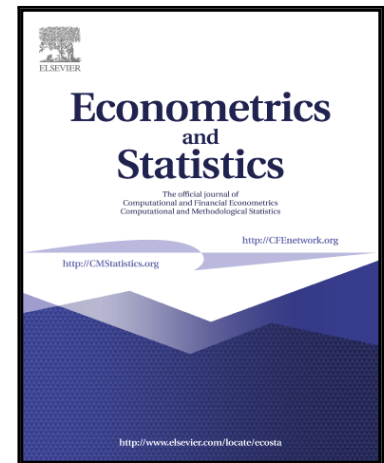
Jeppe Rich

PII: S2452-3062(18)30007-8
DOI: [10.1016/j.ecosta.2018.02.002](https://doi.org/10.1016/j.ecosta.2018.02.002)
Reference: ECOSTA 94

To appear in: *Econometrics and Statistics*

Received date: 17 March 2017
Revised date: 8 February 2018
Accepted date: 8 February 2018

Please cite this article as: Jeppe Rich , A spline function class suitable for demand models, *Econometrics and Statistics* (2018), doi: [10.1016/j.ecosta.2018.02.002](https://doi.org/10.1016/j.ecosta.2018.02.002)



This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A spline function class suitable for demand models

Jeppe Rich¹

Department of Management Engineering, Technical University of Denmark, Bygningstorvet 1, 2800 Lyngby, Denmark

Abstract

A function class suitable for estimating cost preferences in demand models is presented. The function class is applicable to any positive cost variable and is designed to be: i) monotonically decreasing, ii) to have decreasing marginal sensitivity with respect to cost, and iii) to be differentiable at every point. It is shown how suitable functions can be formed from sequences of tailored functions in a manner that ensures their continuity and differentiability at the knot points. The proposed functions are well suited for demand models where price elasticities exhibit a damped pattern as the values of their argument increase. The usual linear-in-parameter functions or non-linear functions, such as the Box-Cox function, do not have an equally flexible way of accounting for such a pattern. This can be relevant when estimating transport demand models where the sensitivity of demand with respect to transport costs is known to decline as the cost increases, i.e. the phenomenon of “cost-damping”. However, it may also be relevant as a means to capture the marginal return of investments or declining marginal utility of income. To provide an illustration, the functions are incorporated in a multinomial logit model that is estimated from synthetically generated data by maximum likelihood. A Monte Carlo simulation study shows that the estimator is able to recover the true parameters². The practical application of the function class is also considered within the new large-scale Danish National Transport Model.

Keywords: *discrete choice models; multinomial logit; functional form; cost-damping; spline functions*

¹ E-mail addresses: rich@dtu.dk, jeppe.rich@gmail.com

²The programs for generating the synthetic data and for estimating the models (in R and SAS software) are available as supplementary material to the electronic version of the paper.

1 Introduction

This paper proposes new parametric functions that could be used in estimating cost and time preferences in demand models. The proposed functions satisfy the requirement of Definition 1:

Definition 1: A cost-damping function class

Represents functions of positive cost that have the following characteristics.

- i) Utility decreases monotonically with respect to cost.*
- ii) Marginal sensitivity decreases with respect to cost (cost-damping).*
- iii) The functions are continuous and differentiable at every point.*

This class includes the well-known Box-Cox function (Box and Cox, 1964) as well as the logarithmic functions. The paper propose two new functions that comply with definition 1 and that offer some additional flexibility compared to existing functions. The first function is inspired by a Taylor expansion of the Box-Cox function and it exploits L'hopital's rule (Rich and Mabit, 2015). It inherits the desirable properties of the Box-Cox function, but it possesses a more flexible curvature. The second function is constructed directly from a sequence of Box-Cox functions. This function relaxes the assumption that the shape parameter is identical across the entire domain. The proposed functions are generally more flexible than traditional functions in the sense that the damping characteristics can be controlled more effectively. It is not discussed in detail why cost-damping is likely to be found in empirical data as this has been considered in other papers (Daly, 2010; Rich and Mabit, 2015; Daly et al., 2016). However, the reasons could include human preferences, unobservable attributes, heteroscedasticity, and selection bias.

The proposed functions are flexible, yet they are constrained to comply with typical properties of demand models, such as convexity in prices and cost-damping. More specifically, the functions are appropriate to circumstances where there is decreasing marginal sensitivity of demand with respect to prices. As examples of a decreasing sensitivity, it may be observed that doubling the money spent on a bottle of wine does not always double the pleasure that it will give, that the quality of life does not always increase in proportion to the available income (Layard et al., 2008), that there are often diminishing returns to scale in investments and that the disutility of increasing a driving distance from 10 to 20 KM is generally less than the disutility of increasing the distance from 200 to 210 KM.

The functions that are proposed in this paper are particularly relevant in analyzing the demand for transport as a function of its costs. Transport models are typically formulated within a random utility framework, where utility is represented as a function of transport costs and of travel time and estimated as discrete choice models. In the literature, there are several contributions to the issue of functional form specification in transport models. Piecewise linear approximations have been discussed in Ben-Akiva and Lerman (1985) in a logit modelling framework and, more recently, in Pinjari and Bhat (2006) in a mixed logit framework. Power series expansions are also discussed in Ben-Akiva and Lerman (1985) and they have been used in testing the adequacy of linear models. Other contributions include the early work on Box-Cox transformations in Gaudry and Wills (1978) and Hensher and Johnson (1981) and the more recent work in Gaudry (2010). Applications of the Box-Cox function are to be found in Gaudry et al. (1989), Ben-Akiva et al. (1987), Mandel et al. (1994) and Lapparent and de Palma (2002). Most recently, Rich and Mabit (2015) have

investigated several combinations of linear-in-parameter but non-linear-in-attribute forms, which, in several cases, have outperformed the Box-Cox model.

Compared to many of the existing functions in the literature, the two functions proposed in this paper can be seen as restricted splines (Wold, 1974) that satisfy the requirement of Definition 1. The segments of these functions are joined at knot points; and their parameters ensure connectivity and differentiability at these points. Several research papers have been concerned with spline-based approaches in the context of discrete choice models, including those of Kneib et al. (2007), Tutz and Scholz (2004), Huang and Nychka (2000) and Bilbao-Ubillos et al. (2015). However, these papers typically adapt a Bayesian approach; and they are generally concerned with the estimation of socioeconomic consumer preferences. They are not aimed at the estimation of price and demand relationships. The problem of using unconstrained spline functions (which are often piecewise polynomial functions) is that, in general, they do not fit into a context of random utility maximization. McFadden (1981) touched upon this in his seminal work, which later won him the Nobel price. In his discussion of indirect utility properties, he underlined the need for quasi-convexity in prices (page 207), which is also a requirement of the William- Daly- Zachary Theorem (Williams, 1977; Daly and Zachary, 1978). This is also mentioned in McFadden's work (page 211). Although it might be possible to estimate a piecewise polynomial function that is convex and complies with random utility theory, one cannot trust that this will be the case. Typically, such flexible functions conflict with the requirements of our definition in certain parts of the domain and for certain applications. This can lead to elasticities with the wrong sign and to biased policy experiments. For this reason, such unconstrained functions are rarely seen in the econometric literature in connection with discrete-choice demand modelling or in demand models more generally.

It is equally relevant to consider the contribution of these proposed functions when compared with piecewise linear functions. Three things can be highlighted in that respect. First, as the linear model is non-differentiable at the knot points, it is not possible to apply a traditional optimization algorithm such as the Newton-Raphson algorithm. Therefore, it is necessary either to set the knot points manually or to develop a heuristic approach to accommodate a search for the optimal placement of the knots. Secondly, a linear segment cannot represent the tail of the utility function adequately. Instead, it is desirable to adopt curves that flatten out in the tail in order to produce the kind of extreme damping that characterizes the log transformation. This case resembles a situation where the elasticities are scale-invariant with respect to the arguments of the function. Scale-invariance in the context of the linear model represents the situation where there is no effect of the attribute variable on the demand. This represents a situation that is typically not identified since it is absorbed in the constants of the model. The proposed functions use a log-transformed specification in the tail of the function and this leads to the desirable property of scale-invariance. Thirdly, from a pure behavioral perspective, it seems natural that preferences are indeed smooth across a population. Hence, applying functions that imply non-smoothness is a coarse approximation of preferences. In principle, this could have consequences for policy experiments as well.

In Section 2, we present the methodology by which the two utility functions that have the desired properties are derived. Section 3 considers model estimation based on a synthetically generated dataset. Section 4 considers the implications of the proposed functions for demand responses and their application within a large-scale transport model. Finally, section 5 contains the conclusions.

2 Methodology

In the following, for simplicity of notation, consider a linear-in-parameter function $\beta F(x_d)$ for a cost variable x_d where $d \in D$ represents destination alternatives from a universal choice-set D and where $\beta < 0$. To simplify the notation, d is suppressed in the following. To illustrate the connection between knots and intervals, consider Figure 1 below which illustrates a piecewise linear function connected by four knots.

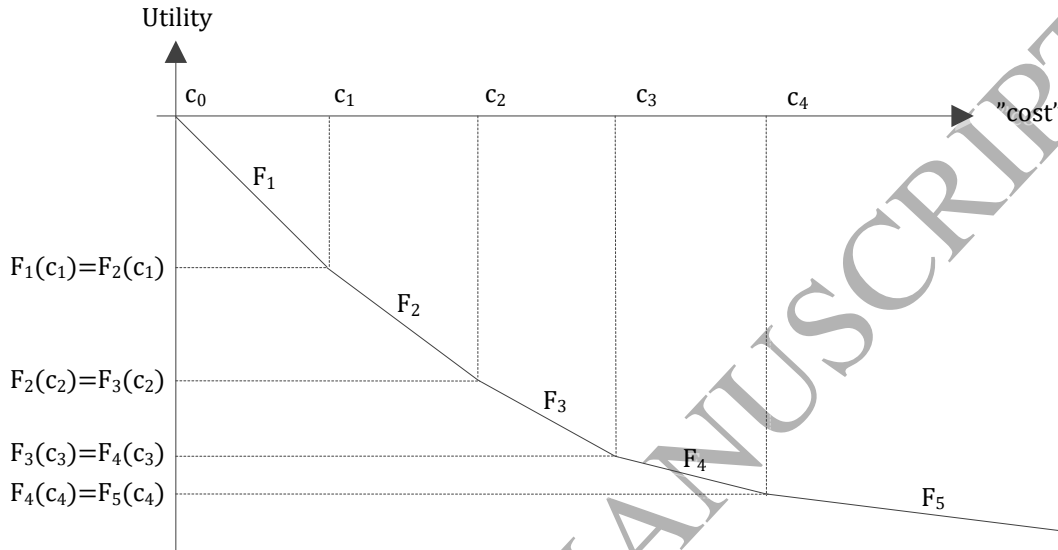


Figure 1: Piecewise linear functions connected by four knot points.

In order to make Definition 1 operational in a mathematical context, a general piecewise function is considered as presented in (1) below

$$F(x, c_1, \dots, c_{Q-1}) = \begin{cases} F_1(x, c_1), & c_0 \leq x < c_1 \\ F_2(x, c_1, c_2), & c_1 \leq x < c_2 \\ F_3(x, c_2, c_3), & c_2 \leq x < c_3 \\ \dots & \dots \\ F_q(x, c_{q-1}, c_q), & c_{q-1} \leq x < c_q \\ \dots & \dots \\ F_Q(x, c_{Q-1}), & x > c_{Q-1} \end{cases} \quad (1)$$

For functions to be consistent with Definition 1 it is required that

$$\frac{\partial F_q(x)}{\partial x} < 0, \forall x > 0, \forall q = 1, \dots, Q - 1, \quad (2)$$

$$\lim_{x \rightarrow c_q} (F_q(x) - F_{q+1}(x)) = 0, \forall x \geq 0, \forall q = 1, \dots, Q - 1 \quad (3)$$

and

$$\lim_{x \rightarrow c_q} \left(\frac{\partial F_q(x)}{\partial x} - \frac{\partial F_{q+1}(x)}{\partial x} \right) = 0, \forall x \geq 0, \forall q = 1, \dots, Q-1. \quad (4)$$

Hence, Equation (2)-(4) represents the formal mathematical recipe for the function class to be considered.

2.1 Function candidates for $F_1(x), \dots, F_Q(x)$

The simplest possible set of function candidates $\mathcal{F}(x) = \{F_1(x), \dots, F_Q(x)\}$ which are consistent with property i) and ii) of Definition 1 is the set of linear functions as illustrated in Figure 1. If $c_0 = 0$ and $c = \{c_1, \dots, c_{Q-1}\}$ represents the vector of knot points, the linear spline class can be written as in (5) below

$$\mathcal{F}(x) = \sum_{q=1}^Q 1_q(x) F_q(x) = \sum_{q=1}^Q 1_q(x) [\theta_q x + \alpha_q], \quad (5)$$

where the indicator function $1_q(x)$ is defined such that $1_q(x) = 1 \Leftrightarrow x \in [c_{q-1}, c_q]$ and zero elsewhere. If it is assumed that $F_1(c_0) = 0$ the slope parameters can be found from (6)

$$\{\theta_1, \dots, \theta_Q\} = \left\{ \frac{-F_1(c_1)}{c_1}, \frac{F_1(c_1) - F_2(c_2)}{c_2 - c_1}, \dots, \frac{F_{Q-1}(c_{Q-1}) - F_Q(c_Q)}{c_Q - c_{Q-1}} \right\}. \quad (6)$$

The first intercept parameter α_1 is set to 0 and the remaining intercepts can be found using (3) recursively. Hence,

$$\{\alpha_1, \dots, \alpha_Q\} = \{0, c_1(\theta_1 - \theta_2), \alpha_1 + c_2(\theta_1 - \theta_2), \dots, \alpha_{Q-1} + c_Q(\theta_{Q-1} - \theta_Q)\}.$$

The piecewise function, however, has undesirable properties as mentioned in the introduction.

First, as seen in Figure 1, the function is not consistent with Definition 1 as it violates condition iii), e.g. it has “kinks” in the first-order derivative at the knot points. This is not realistic as one often expects a smooth transition between cost attributes and perceived marginal utility. Hence, the linear spline formulation is, presumably, a piecewise linear approximation of a smooth process. In practice, it could have undesirable impacts with respect to policy analysis. If a policy intervention aimed at trips above a certain distance (i.e., related to the construction of a large bridge), was compared to a policy aimed at shorter trips, it could potentially bias the corresponding consumer surplus calculation. Secondly, the linear spline requires a knot to be defined at the endpoint to enable the calculation of the slope and the intercept of the final segment. This raises an identification issue as a scale-invariant specification is absorbed in the alternative specific constant.

A solution to these problems would be to introduce many knot points. However, this raises the problem of how to choose the knot points. As the function has kinky first-order derivatives the search for optimal knot points c_1, \dots, c_{Q-1} will have to be based on heuristic methods, where the likelihood function needs to be evaluated for each combination of points. In other words, a Newton-like procedure is not possible.

In the following, two function candidates are proposed that fulfill Definition 1 and solve the “tail-issue” of the piecewise linear function. This is accomplished by imposing a more flexible non-linear curvature.

2.2 Log-power spline

The first candidate arises from a power series expansion of the logarithmic function. Hence, consider the function shown in Equation (7) below

$$\mathcal{F}(x) = \sum_{q=1}^Q 1_q(x) [\theta_q \ln(x)^{Q-q+1} + \alpha_q]. \quad (7)$$

The function is connected in $Q - 1$ knot points c_1, \dots, c_{Q-1} and is defined in such a way that the $\theta_1 \ln(x)^Q + \alpha_1$ function operates in the first part of the curve where $c_0 \leq x < c_1$. For $c_1 \leq x < c_2$ the function $\theta_2 \ln(x)^{Q-1} + \alpha_2$ is active and this is continued successively, such that for the last part of the curve, the function is modelled using a purely logarithmic form. It should be acknowledged that $\mathcal{F}(x) < 0$ if $x < 1$. This is not necessarily a problem; however, should non-negativity be desirable the function can be shifted accordingly.

It is worth noting that an approximation of the Box-Cox function emerges as a special case of (7) with $\alpha_q = \left\{ \frac{1}{\gamma}, 0, \dots, 0 \right\}$ and $\theta_q = \frac{\gamma^q}{q!}$ for all $q = 1, \dots, Q$ where γ represents the Box-Cox parameter. This can be shown using L'hospital's rule and the precision of the approximation depends on the size of Q (Gaudry, 2010; Rich and Mabit, 2015). However, the approximation in a given point can be arbitrarily good and the difference can be expressed as the Taylor remainder term $o(Q)$. The proposed function shares all of the desirable properties of the Box-Cox function yet allows for a more flexible curvature through the introduction of knot points.

It now remains to find the form of the θ and the α parameters. First, note that of the Q spline parameters $\theta_1, \dots, \theta_Q$ and $\alpha_1, \dots, \alpha_Q$, only $Q - 1$ is identified. Hence, the first scale θ_1 is normalized to 1 and the first intercept α_1 is set to 0.

Proposition 1

Consider an arbitrary log-power spline function of degree Q as described in (7) which complies with Definition 1. Let $\theta_1 = 1$ and $\alpha_1 = 0$. Unique spline parameters $\theta_2, \dots, \theta_Q$ and $\alpha_2, \dots, \alpha_Q$ exist and can be found from

$$\theta_q = \frac{Q}{Q - q + 1} \prod_{r=2}^q \ln(c_{r-1}), \forall q = 2, \dots, Q, \quad (8)$$

and

$$\begin{aligned} \alpha_2 &= \frac{1}{Q-1} \ln(c_1)^Q, \\ \alpha_3 &= \alpha_2 + \frac{2!}{Q-1} \ln(c_1) \ln(c_2)^{Q-1}, \\ \alpha_4 &= \alpha_3 + \frac{3!}{Q-1} \ln(c_1) \ln(c_2) \ln(c_3)^{Q-2}, \end{aligned} \quad (9)$$

$$\alpha_5 = \alpha_4 + \frac{4!}{Q-1} \ln(c_1) \ln(c_2) \ln(c_3) \ln(c_4)^{Q-3},$$

and

$$\alpha_q = \alpha_{q-1} + \frac{(q-1)!}{Q-1} \ln(c_{q-1})^{Q-q+2} \prod_{r=1}^{q-2} \ln(c_r).$$

Proof

Equation (2)-(4) can be used in a recursive manner to determine the parameters $\theta_2, \dots, \theta_Q$ and $\alpha_2, \dots, \alpha_Q$. The easiest way to find the closed-form expressions in (8) and (9) is to find all scaling parameters θ and then subsequently work out the α intercepts recursively. Per definition, $\theta_1 = 1$ and $\alpha_1 = 0$, and the first scale parameter for $q = 2$ can be based on Equation (4) as shown in Equation (10) below. The θ_2 parameter is then applied in the calculation of the next scale parameter as shown in Equation (11). This process continues, and it can be shown, by algebraic manipulations that the q 'th scale parameter can be represented as in Equation (12) which is identical to (8).

$$\left. \frac{\partial F_1(x)}{\partial x} \right|_{x=c_1} = \left. \frac{\partial F_2(x)}{\partial x} \right|_{x=c_1} \Leftrightarrow \frac{Q \ln(c_1)^{Q-1}}{(Q-1) \ln(c_1)^{Q-2}} = \frac{Q \ln(c_1)}{(Q-1)} = \theta_2(c_1, 0, \dots), \quad (10)$$

$$\left. \frac{\partial F_2(x)}{\partial x} \right|_{x=c_2} = \left. \frac{\partial F_3(x)}{\partial x} \right|_{x=c_2} \Leftrightarrow \theta_3(c_1, c_2, 0, \dots) \left(\frac{Q-1}{Q-2} \right) \ln(c_2) = \frac{Q \ln(c_1)}{(Q-2)} \ln(c_2) = \theta_3(c_1, c_2, 0, \dots), \quad (11)$$

...

$$\left. \frac{\partial F_q(x)}{\partial x} \right|_{x=c_q} = \left. \frac{\partial F_{q+1}(x)}{\partial x} \right|_{x=c_{q+1}} \Leftrightarrow \frac{Q}{(Q-q+1)} \prod_{r=1}^q \ln(c_r) = \theta_q(c_1, \dots, c_q, 0, \dots). \quad (12)$$

The existence and uniqueness of $\theta_2, \dots, \theta_Q$ follow directly from the recursive solution procedure. For the first equation, there is only one possible solution for θ_2 and this in turn causes the calculation of θ_3 to be unique and so on. The next step is to find the interception parameters $\alpha_2, \dots, \alpha_Q$, to make sure that the utility functions are connected in all of the knots. These parameters can be found recursively from

$$\begin{aligned}
F_1(x, c_1)|_{x=c_1} &= F_2(x, c_1, c_2)|_{x=c_1}, \\
F_2(x, c_1, c_2)|_{x=c_2} &= F_3(x, c_2, c_3)|_{x=c_2}, \\
&\dots \\
F_q(x, c_{q-1}, c_q)|_{x=c_q} &= F_{q+1}(x, c_q, c_{q+1})|_{x=c_q}.
\end{aligned}$$

For the log-power spline in (7), this can be written as

$$\begin{aligned}
\ln(c_1)^Q &= \theta_2 \ln(c_1)^{Q-1} + \alpha_2, \\
\theta_2 \ln(c_2)^{Q-1} + \alpha_2 &= \theta_3 \ln(c_2)^{Q-2} + \alpha_3, \\
&\dots \\
\theta_{q-1} \ln(c_q)^{Q-q+1} + \alpha_{q-1} &= \theta_q \ln(c_2)^{Q-q} + \alpha_q.
\end{aligned} \tag{13}$$

By inserting the calculated scale parameters $\theta_2, \dots, \theta_Q$, as found from (10)-(12), the form of the intercepts can be found recursively. By trivial algebraic manipulations, the form in Equation (9) can be shown to be valid. Uniqueness follows directly from the recursive solution.

2.3 Power splines

The second function candidate arises from a sum of power functions for which the power is gradually decreasing and always less or equal to 1. Hence, the function to consider is given by

$$\mathcal{M}(x) = \sum_{q=1}^Q 1_q(x) [\theta_q x^{\pi_q} + \alpha_q], 1 \geq \pi_1 \geq \dots \geq \pi_Q \geq 0. \tag{14}$$

The curvatures of the power spline and of the log-power spline of (7) are similar. It is also possible to define a more direct Box-Cox spline that is formed as a sequence of connected Box-Cox functions, hence

$$B(x) = \sum_{q=1}^Q 1_q(x) \left[\theta_q \left(\frac{x^{\pi_q} - 1}{\pi_q} \right) + \alpha_q \right], 1 \geq \pi_1 \geq \dots \geq \pi_Q \geq 0. \tag{15}$$

The curvature of the above Box-Cox function is similar to that of the power spline. The only difference is the division by the Box-Cox parameters $\pi_q \forall q$. However, this scaling is absorbed in the θ_q parameters for all q .

For the power-spline function, using similar manipulations as in Proposition 1, it can be shown that the corresponding scaling and intercept parameters can be found from (16) and (17)

$$\theta_q = \prod_{r=1}^{q-1} \frac{\pi_r}{\pi_{r+1}} c_r^{\pi_r - \pi_{r+1}}, \forall q = 1, \dots, Q-1, \tag{16}$$

and

$$\alpha_2 = c_1^{\pi_1} \left(1 - \frac{\pi_1}{\pi_2} \right), \tag{17}$$

$$\alpha_3 = \alpha_2 + \theta_2 c_2^{\pi_2} - \theta_3 c_2^{\pi_3},$$

$$\alpha_4 = \alpha_3 + \theta_3 c_3^{\pi_3} - \theta_4 c_3^{\pi_4},$$

...

$$\alpha_q = \alpha_{q-1} + \theta_{q-1} c_{q-1}^{\pi_{q-1}} - \theta_q c_{q-1}^{\pi_q}.$$

The corresponding parameters for the Box-Cox spline function have been included in Appendix 2.

3 Estimating knot points and parameters

As a result of requirement iii) in Definition 1, and the way the functions are constructed, each of the proposed function candidates in Equations (7) and (14) are differentiable with respect to the spline parameters. Consequently, it is possible to calculate first- and second-order conditions for the log-likelihood function for all $x \geq 0$. This implies that the optimal knots can be estimated using maximum likelihood estimation simultaneously with the remaining parameters.

As seen from Equation (8) and (9), even low-dimensional spline functions are highly nonlinear and cannot be estimated using standard software. However, today most econometric software packages facilitate the estimation of nonlinear functions based on auto-generated derivatives. In the following, the joint estimation of all parameters using maximum likelihood estimation is explored. To test the procedure, a controlled experiment consisting of a synthetic data generating process that resembles a multinomial logit model is constructed. The synthetic model is formed from two spline functions each having two knot points and it is estimated using a SAS procedure called PROC NL MIXED (SAS Software, 2016). The setup can easily be generalized to more choices, other types of error terms (e.g., nested logit and mixed logit), and more spline variables, each having more knot points.

3.1 Synthetic MNL data

In the following, a generic data generating model is defined. It can be tailored to different specification tests, of which only a few are analyzed in this paper. Consider in (18) the utility functions $V_{i,n}$ for choices $i = 1, \dots, 4$ for a synthetic population $n = 1, \dots, N$

$$V_i = k_i + \beta_1 \mathcal{F}(x_i, cx_1, cx_2) + \beta_2 \mathcal{F}(z_i, cz_1, cz_2) + \varepsilon_i, \forall i = 1, \dots, 4. \quad (18)$$

The index of n is suppressed for ease of notation. It is assumed that $\mathcal{F}(x_i, cx_1, cx_2)$ and $\mathcal{F}(z_i, cz_1, cz_2)$ represent spline functions in the form of log-power splines. Hence, they are expressed in their knot points for which they each have two. That is,

$$\begin{aligned} \mathcal{F}(x_i, cx_1, cx_2) = & 1_{[0, cx_1]} \ln(x_i)^3 + 1_{[cx_1, cx_2]} [\theta_2(cx_1) \ln(x_i)^2 + \alpha_2(cx_1)] \\ & + 1_{[cx_2, cx_3]} [\theta_3(cx_1, cx_2) \ln(x_i) + \alpha_3(cx_1, cx_2)]. \end{aligned}$$

Here the spline parameters, according to (7), are given by $\theta_2(cx_1) = \frac{3}{2} \ln(cx_1)$, $\theta_3(cx_1, cx_2) = 3 \ln(cx_1) \ln(cx_2)$, $\alpha_2(cx_1) = -\frac{1}{2} \ln(cx_1)^3$ and $\alpha_3(cx_1, cx_2) = -\frac{1}{2} \ln(cx_1) [3 \ln(cx_2)^2 + \ln(c_1)^2]$. The notation for $\mathcal{F}(z_i, cz_1, cz_2)$ is similar and is not repeated.

Furthermore, assume that ε_i follows a standard Gumbel distribution. Hence, if $\omega \sim \text{Exp}(1)$, then $\varepsilon_i \sim -\ln(\omega)$.

The exogenous variables x_i and z_i are constructed from random distributions. In the following, these are assumed to be log-normally distributed, as this resembles a typical income or price distribution. To be able to define the mean for the log-normal distributions, these are constructed from the following distributions

$$x_i \sim \exp(\varphi(\mu_x, \sigma_x^2)) \text{ and } z_i \sim (1 - \rho) \exp(\varphi(\mu_z, \sigma_z^2)) + \rho x_i.$$

The mean and variance are defined by

$$\vartheta_x = \sqrt{m_x + m_x^2}, \mu_x = \ln(m_x^2/\vartheta_x) \text{ and } \sigma_x^2 = \tau_x \sqrt{\ln(\vartheta_x^2/m_x^2)}$$

and

$$\vartheta_z = \sqrt{m_z + m_z^2}, \mu_z = \ln(m_z^2/\vartheta_z) \text{ and } \sigma_z^2 = \tau_z \sqrt{\ln(\vartheta_z^2/m_z^2)}.$$

Here, m_x and m_z define the mean of the final log-normal distribution, and τ_x and τ_z are parameters used to control the scale of the variation. Assume that $\rho = 0$, in order to have uncorrelated variables, and use $\tau_x = \tau_z = 4$ to force some variation in the data. Also assume that the true knot points are selected in such a way that they are approximately equal to the corresponding 25% and 75% quartile of x_i and z_i , respectively, but represented as integers for ease of comparison.

After generating the utility function as described above, the utility maximization principle is imposed for the choice process. Refer to the choice variable as y_{ni} , where $y_{ni} = 1 \Leftrightarrow V_{ni^*} \geq V_{ni}, \forall i$ and $i \neq i^*$. The code for generating the synthetic data in R or SAS can be provided upon request.

3.2 MLE and inference

Based on the above data y_{ni} , and utility function V_{ni} , the log-likelihood function for the MNL model is defined as

$$\begin{aligned} L(\boldsymbol{\theta}|X) &= \prod_n \prod_i y_{ni} \ln P_{ni} \\ &= \prod_n \prod_i y_{ni} \ln \left(\frac{\exp(V_{ni}(\boldsymbol{\theta}|x_i, z_i))}{\sum_j \exp(V_{nj}(\boldsymbol{\theta}|x_i, z_i))} \right), \end{aligned}$$

where, in the most general form, $\boldsymbol{\theta} = \{k_1, \dots, k_3, \beta_x, \beta_z, cx_1, cx_2, cz_1, cz_2\}$. The models are estimated using PROC NL MIXED (SAS) and R and the code are provided as supplementary material.

To check that the model can approximately reproduce the true parameters for different data generating processes, the generation of the data and the estimation of the models have been embedded in a Monte-Carlo framework which allows us to estimate the model repeatedly but with different random seeds for the data generation. Hence, x_i and z_i are maintained across each Monte-Carlo iteration (for the same model),

but with different realizations of the error-term ε_i . This allows the model to be tested, and the parameters that have been estimated from different finite samples to be compared.

We start with the simplest possible model, namely a single-variable model referred to as M1 with utility given by

$$V_i = \beta_1 \mathcal{F}(x_i, cx_1, cx_2) + \varepsilon_i, \forall i = 1, \dots, 4.$$

The results for M1 are given below in Table 1.

Parameter	True	Nobs=1.000		Nobs=10.000		Nobs=100.000	
		Mean	MC std	mean	MC std	mean	MC std
β_x	-0.2	-0.232	0.108	-0.203	0.027	-0.200	0.005
(T-stat)		(-6.422)	4.398	(-13.090)	5.345	(-37.825)	4.117
cx_1	6	6.813	2.774	6.280	1.420	6.050	0.332
(T-stat)		(2.415)	1.376	(5.182)	1.256	(17.098)	1.232
cx_2	14	20.703	34.052	13.961	2.017	14.006	0.517
(T-stat)		(2.902)	1.461	(7.635)	1.815	(25.437)	1.272

Table 1: Maximum likelihood model parameters for different sample sizes for 200 Monte Carlo runs for model M1.

As can be seen from Table 1 the spline parameters, as well as the scale parameters β_x , approached the true values as the sample size increases. The “MC std” columns represent the standard deviation of the attribute across the Monte-Carlo sample.

It should be noted that the likelihood function sometimes tends to become relatively flat around the optimal knot points. This was experienced in this synthetic data setup, as well as for practical applications (refer to Section 4).

The next model that is considered here is referred to as M2. This model is a slight generalization of the previous model in that we now allow for a new linear term $\beta_2 z_i$, that is

$$V_i = \beta_1 \mathcal{F}(x_i, cx_1, cx_2) + \beta_2 z_i + \varepsilon_i, \forall i = 1, \dots, 4.$$

Parameter	True	Nobs=1.000		Nobs=10.000		Nobs=100.000	
		mean	MC std	mean	MC std	mean	MC std
β_x	-0.2	-0.215	0.064	-0.207	0.030	-0.200	0.006
(T-stat)		(-6.961)	9.974	(-11.559)	5.125	(-35.499)	4.446
β_z	-0.1	-0.100	0.007	-0.100	0.002	-0.001	0.001
(T-stat)		(-15.151)	8.545	(-47.360)	0.237	(-150.242)	0.268
cx_1	6	6.813	2.666	6.043	1.457	6.019	0.393
(T-stat)		(2.304)	0.890	(4.767)	1.245	(16.116)	1.188
cx_2	14	20.645	37.738	13.957	1.912	13.942	0.517
(T-stat)		(3.135)	3.605	(7.804)	1.823	(25.795)	1.458

Table 2: Maximum likelihood model parameters for different sample sizes for 200 Monte Carlo runs of model M2.

The final model to consider, the M3 model, is defined in (17), except that all alternative specific constants are assumed to be zero. This model allows for two different spline functions, with two spline knot points each, and a corresponding scale parameter. This result in a total of six parameters.

Parameter	True	Nobs=1.000		Nobs=10.000		Nobs=100.000	
		Mean	MC std	Mean	MC std	mean	MC std
β_x	-0.2	-0.221	0.080	-0.205	0.037	-0.200	0.006
(T-stat)		(-9.605)	34.298	(-13.625)	9.704	(-36.604)	4.251
β_z	-0.1	-0.128	0.146	-0.104	0.043	-0.103	0.015
(T-stat)		(-6.401)	21.749	(-4.640)	3.040	(-8.402)	2.057
cx_1	6	6.997	2.704	6.212	1.487	5.924	0.365
(T-stat)		(2.718)	2.058	(5.232)	1.290	(17.249)	1.231
cx_2	14	14.893	16.436	13.923	1.854	14.867	0.581
(T-stat)		(3.551)	3.728	(8.171)	2.687	(26.489)	1.267
cz_1	3	6.144	4.966	3.986	1.954	3.149	0.495
(T-stat)		(1.877)	1.527	(2.812)	0.979	(6.791)	0.941
cz_2	15	31.874	75.763	14.286	4.023	14.759	1.261
(T-stat)		(1.496)	0.486	(3.616)	0.821	(12.486)	0.863

Table 3: Maximum likelihood model parameters for different sample sizes for 200 Monte Carlo runs of model M3.

The main conclusion of the experiments is that if the underlying data generating process is described by one or more log-power spline functions, parameters can be estimated jointly using maximum likelihood estimation techniques.

As the models are highly non-linear, the likelihood functions are generally not convex across the entire parameter space. Therefore, some attention should be paid to starting values. However, in the above experiments, the results are not overly sensitive to starting values, and even with relatively poor starting values, the optimum can be attained rapid using a Newton-Raphson algorithm. Appropriate starting values may be obtained from a linear model combined with strictly positive and ordered knot points defined according to for example quartiles. More advanced search algorithms could be implemented, and are a topic for future research.

4 Impacts and practical application

Although the function classes were described mathematically in Section 2, and an estimation analysis was provided in Section 3, it is relevant to consider what the proposed functions imply in practice. We therefore start by considering the curvature and the implied elasticity curves in Section 4.1, then continue to discuss the mixing of functions in Section 4.2, and end by offering results and application experience from the Danish National Transport Model in Section 4.3.

Section 4.1 below consider a random utility model where spline functions are applied to express the indirect disutility of transport costs (and time).

4.1 On the curvature

To investigate the curvature in more detail, consider a simple destination choice model with (indirect) utility function U_d . The utility function includes a spline function $\mathcal{F}(x_d)$ defined on the basis of a generic cost variable x_d . A Gumbel distribution is attributed to the error term so that the model renders logit choice probabilities.

$$U_d = k + \beta \mathcal{F}(x_d) + \varepsilon_d, \forall d = 1, \dots, D$$

From this choice model, the model sensitivity can be represented by elasticity curves.

$$E\mathcal{F}_d = \beta \frac{\partial \mathcal{F}(x_d)}{\partial x_d} x_d (1 - P_d), \forall d = 1, \dots, D$$

To maintain simplicity, the analysis is limited to splines of the order of $Q = 3$, and only to the log-power spline as described in (7). For the log-power spline, the parametrization is straightforward as the powers are integers, e.g.

$$\mathcal{F}(x_d) = 1_{[0, c_1]} \ln(x_d)^3 + 1_{[c_1, c_2]} [\theta_2 \ln(x_d)^2 + \alpha_2] + 1_{[c_2, c_3]} [\theta_3 \ln(x_d) + \alpha_3], \quad (19)$$

where $1_{[0, c_1]}$ represents the indicator function for the interval $[0, c_1]$. The elasticity curves are given by

$$E\mathcal{F}_d = \beta (1_{[0, c_1]} 3 \ln(x_d)^2 + 1_{[c_1, c_2]} 2 \theta_2 \ln(x_d) + 1_{[c_2, c_3]} \theta_3) (1 - P_d),$$

and the spline parameters are directly available from Table 5 in Appendix 1. Below in Figure 2, the curvature of the utility function is illustrated for the simple linear case, for $\ln(x_d)^3$, and for the spline function as represented by Equation (19). The comparison between $\ln(x_d)^3$ and the spline is interesting as it reveals the gradual damping as cost increases. The function is constructed with $\beta = -0.05$ and knot points $\{c_2, c_3\} = \{50, 100\}$.

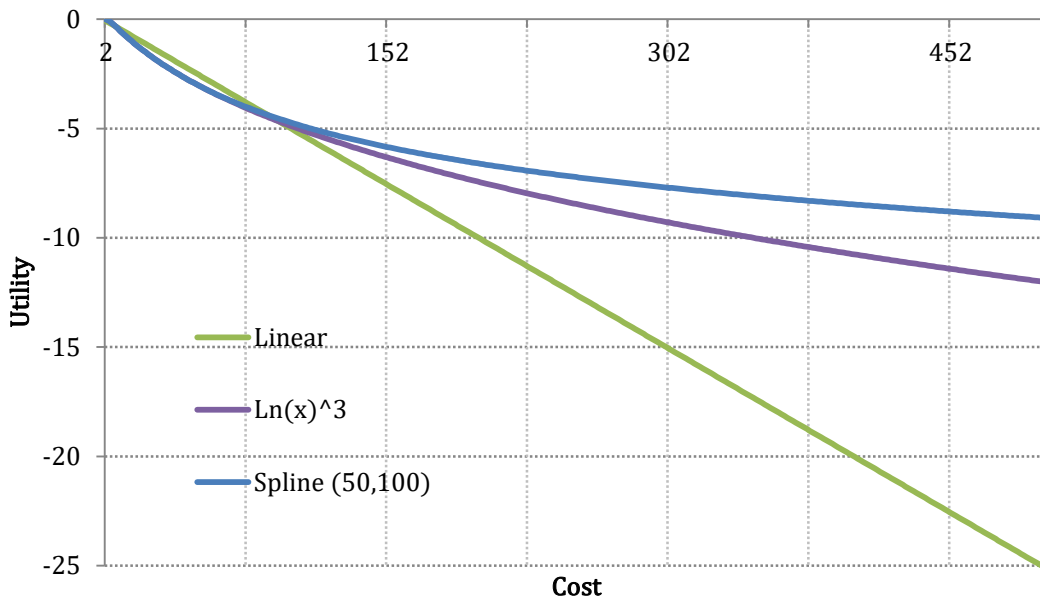


Figure 2: Curvature for utility functions as a function of cost.

The corresponding elasticity curves are shown below in Figure 3 from where it is clear how the tail of the spline function is damped. Contrary to this, the linear model tends to produce rather extreme and unrealistic elasticities due to the direct scaling with respect to cost. The $\ln(x_d)^3$ function provides some damping compared to the linear model; however for large cost values it tends to exaggerate elasticities as well. It is also worth noting that a combination of functions such as the logarithmic and linear functions (Daly, 2010) does not solve the damping issue entirely. When the cost variable becomes large the linear term dominates the logarithmic term in a logit type model.

In Figure 4, it is shown how different knot points produce different elasticity curves.

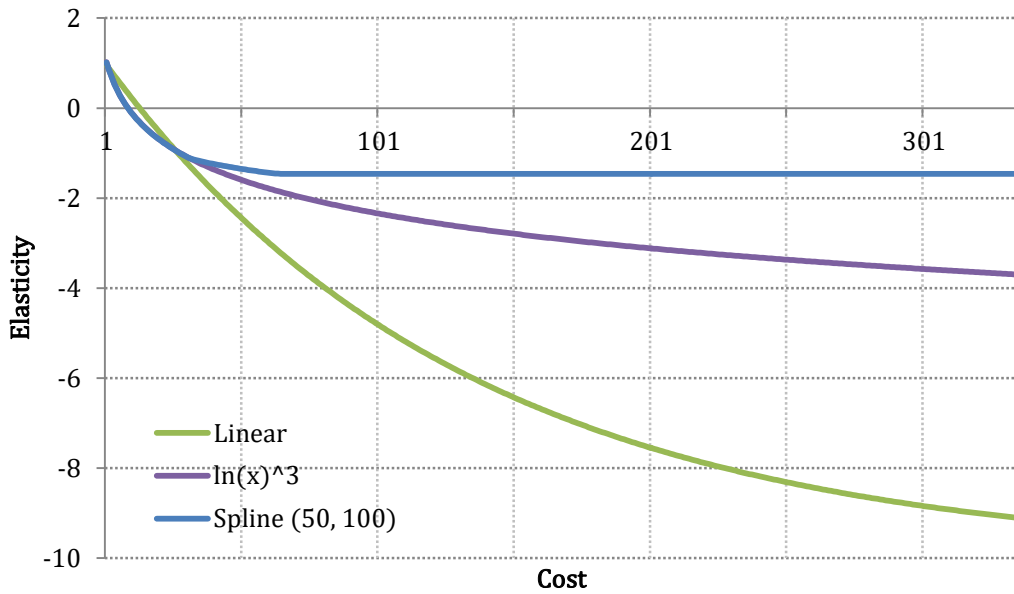


Figure 3: Curvature of the elasticity curves as a function of cost.

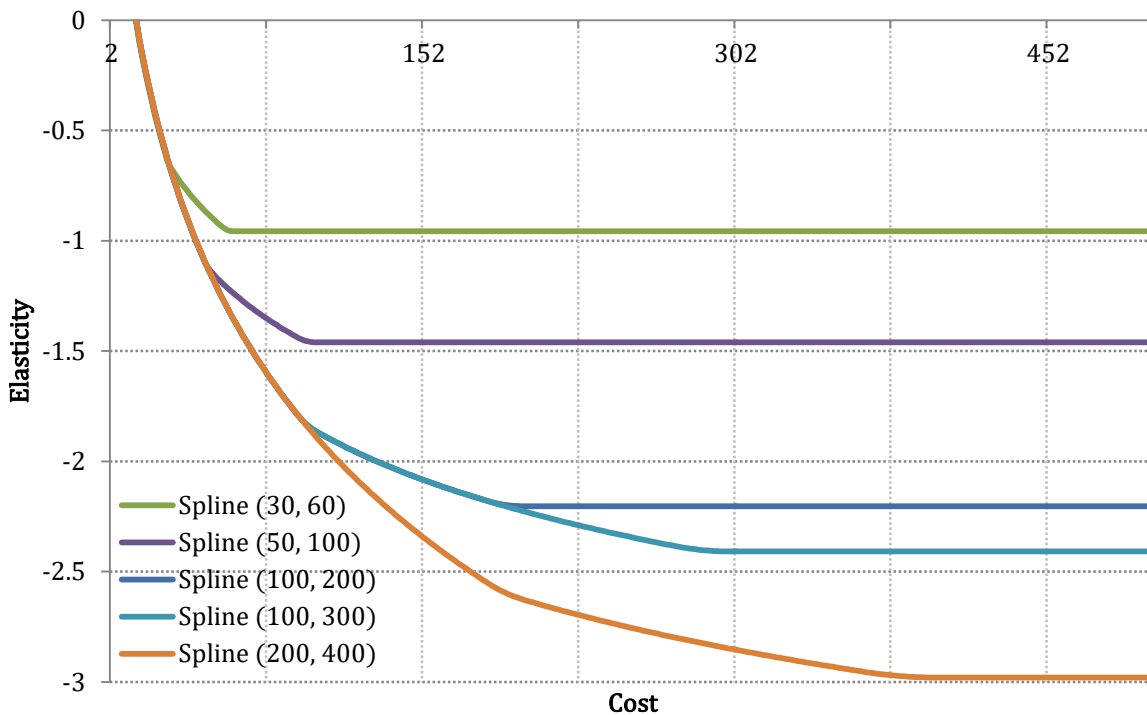


Figure 4: Comparison of elasticity curves for different knot points for a log-power spline with $Q=3$.

The real benefit of applying the spline function is that it is possible to control more effectively the behavior of the tail. This is not accommodated by any other simple functional forms such as the Box-Cox or alternative linear approximations of the Box-Cox as presented in Rich and Mabit (2015).

4.2 Mixing of functions

Additive mixing of different functions can be used to add even more flexibility to the curvature. A common way of mixing is the combined logarithmic and linear form $\beta_1 \ln(x_d) + \beta_2 x_d$. In this case, the linear curvature is mixed with the logarithmic parameter, and scale parameters are estimated separately for the two functions. In the case of the logarithmic and linear model, it is important to understand that $\beta_1 \ln(x_d)$ operates in the first part of the curve, whereas $\beta_2 x_d$ operates in the last part of the curve. As the cost increases, the linear term tends to dominate the logarithmic term and the exaggeration of elasticities caused by the linear term will prevail.

The spline function is also a good function candidate for mixing. However, it is particularly relevant to consider mixing candidates that are active in the first part of the curve. If spline functions are combined with functions that involve scale effects in the tail of the distribution, they will generally violate the damping properties. Hence, it is recommended that mixing candidates should have a similar or smaller order of power than the tail function of the spline. This suggests the pure logarithmic form as a perfect mixing candidate as it is active in the first part of the curve only. By allowing for a separate scaling of this function, we indirectly allow for a more flexible curvature. In any case, testing the significance of the added term(s) provides an easy way to assess whether the functions are sufficiently flexible.

4.3 Results from the Danish National Transport Model

Until now, it has been assumed that the underlying discrete choice model resembled a multinomial logit model and operated on a destination choice domain. However, it may be more relevant to consider a nested-logit model by adding a mode-choice dimension. It is trivial to see that if the logsum parameters do not change with the choice of destination or mode, all of the calculations carry over and the spline parameters are similar to those found above. Even if the logsum parameters varied with one or more choice dimensions, calculations would be relatively straightforward. In that case, the scaling parameters would remain essentially unchanged except for additional scaling of the logsum parameters.

In the Danish National Transport Model (Rich and Hansen, 2016), nested logit models with logsum parameters that do not change across choice of destination and mode are applied. The models are characterized by having a relatively wide destination domain. It has been difficult to find appropriate function candidates applicable to the entire destination domain. As a result, a log-power spline formulation has been applied and has been shown to work. In particular, the function has worked well for leisure travel and shopping travel segments as these tend to exhibit a strong damping pattern. However, this has not been the case for business and commuting travel to the same extent, which is less damped. The core function implemented in the Danish National Transport Model (only for weekday travel and only for primary trips) is shown in equation (20) below

$$F_n(GTT_{m,d|n}, \beta, c_1, c_2) = \beta \sum_{q=1}^3 1_q \left[\theta_q(c_1, c_2) \ln(GTT_{m,d|n})^{q+1-q} + \alpha_q(c_1, c_2) \right], \quad (20)$$

where $GTT_{m,d|n} = \frac{cost_{m,d}}{\mu_n} + time_{m,d}$, and μ_n is the value-of-time for individual n . The value-of-time is income-dependent, but is fixed prior to the estimation. Clearly, this implies that elasticity curves, as shown below for time and cost attributes, are essentially a weighted representation where the balance between

cost and time is controlled by the value-of-time. However, to illustrate the damping, this simplification is not strictly critical. The function in (20) is then further mixed with simple log-functions to introduce a more flexible description for shorter trips. Below in Figure 5, the log-likelihood profile of using the log-power spline class compared to the previous model for which the base model is actually identical to the parametrization with $c_1 = \infty$ and $c_2 = \infty$ is shown.

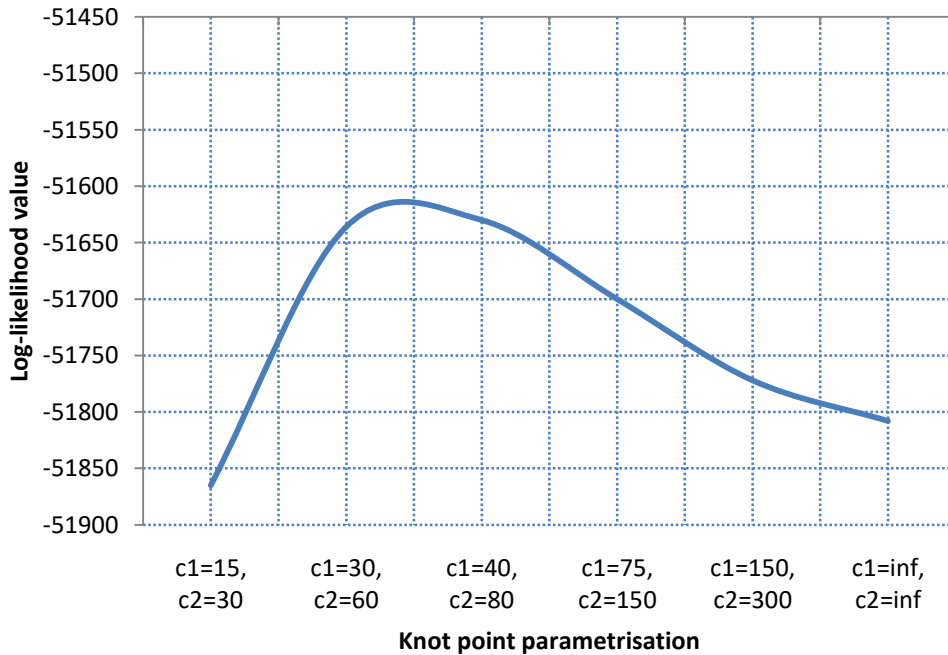


Figure 5: Log-likelihood performance of leisure trip segment as a function of different knot point parameters.

As can be seen, there is a significant improvement in log-likelihood of approximately 200 log-likelihood points. The estimation of the knot points was carried out by simulating likelihood values for the nested logit model over a grid of c_1 and c_2 parameters. Hence, it has not been possible to provide t-stats for the parameters and the parameters can only be considered as “near-optimal”. However, as also previously discussed, the curvature is relatively flat around the knot points. It makes no practical difference as long as all other parameters, including scale parameters, conditional on the near-optimal values of c_1 and c_2 , are estimated.

The implication for the model sensitivity of choosing a “near-optimal” combination of c_1 and c_2 , in contrast to choosing $c_1 = \infty$ and $c_2 = \infty$, is significant. The mean of the elasticities is slightly affected, as only a small share of the trips exceeds c_1 and c_2 . However, for longer trips the spline function prevents the elasticities from “exploding”, as it effectively introduces a maximum elasticity.

This has several implications. First, for large-scale infrastructure projects for which the share of long-distance trips is high, using a proper spline form significantly reduces the elasticities for these trips. The elasticity also tends to change from being sensitive to distance to being sensitive to the choice of mode. Figure 6 and Figure 7 are examples of the impact on the elasticity curves.

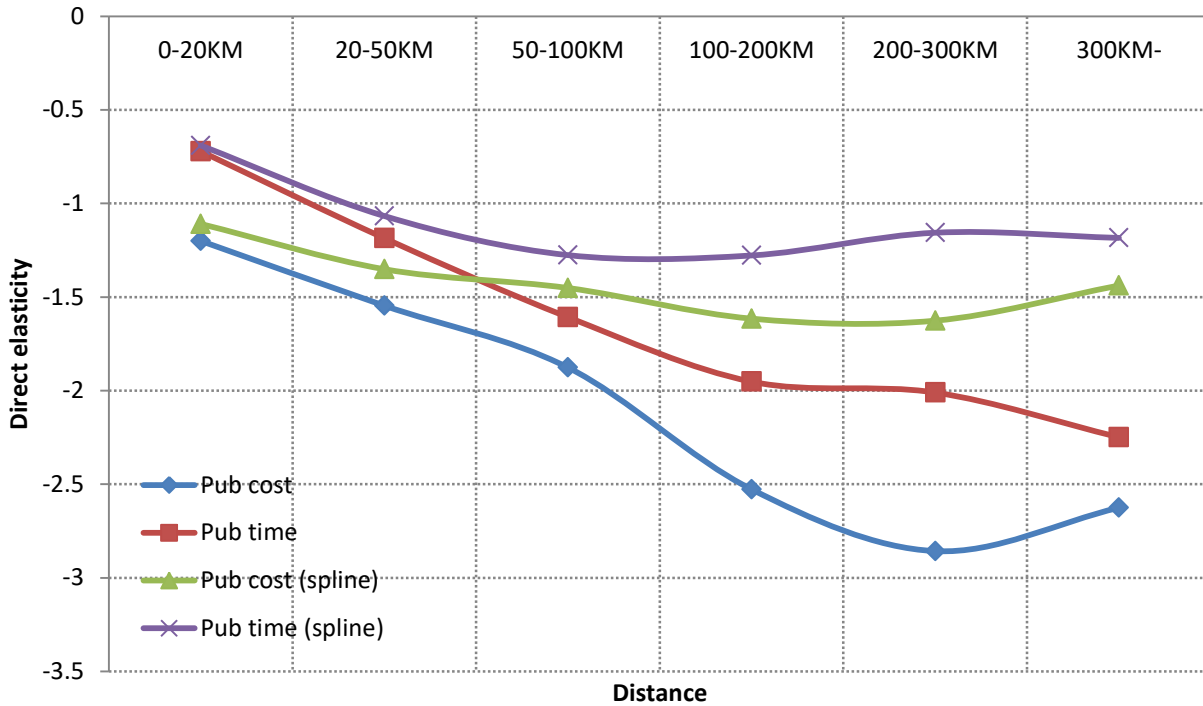


Figure 6: Elasticity curve for public transport cost and time for leisure travel for spline model versus normal model with $c_1 = \infty$ and $c_2 = \infty$.

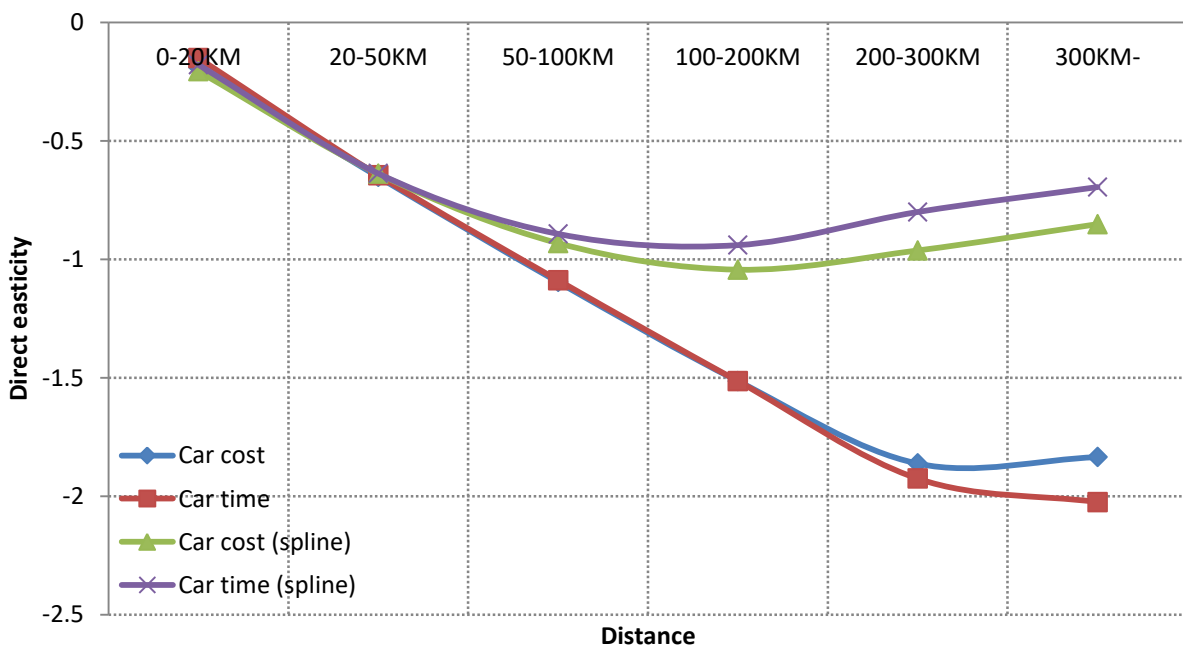


Figure 7: Elasticity curve for car cost and time for leisure travel for spline model versus normal model with $c_1 = \infty$ and $c_2 = \infty$.

As can be seen, there is a significant difference, not only in the tail of the distribution but also for semi-long distances. Even the mean is relatively different. For the more distant destinations above 300 KM, the

elasticity curve is actually declining. This can be attributed to mode shifts, but can also represent shifts in the destination choice. It should be remembered that there are very few observed choices in the tail of the distribution, and single observations may cause the curvature to fluctuate somewhat. The above simulations are based on a cost increase of 10% and this generally causes the average distance to decline. Hence, this will cause a move from right to left. If there are very few trips for some of the longest distance intervals, the share of these tends to drop at the expense of shorter trips, and this movement will result in a backward bending curve.

The performance of the function class has also been tested during two consecutive back casting exercises using the Danish National Transport Model. First, the model was back cast to year 2002. Then, in a more recent exercise the model was forecasted to 2015 and evaluated against observed transport measures for 2015. Overall, these results have been very encouraging, and it was specifically tested whether the model could reproduce the transport growth across the Great Belt Corridor in Denmark for which a price reduction of 20% was introduced in 2005. As this is a long-distance corridor with trips across the entire distance domain, this test was particularly interesting from the perspective of evaluating the spline function. In both of these tests the model was able to reproduce the observed sensitivity very well. More details with respect to the Danish National Transport Model can be found in Rich and Hansen (2016).

5 Summary and conclusions

The paper presents a new function class that may be suited for estimating demand models where individual demand responses are estimated as a function of prices. The proposed function class embraces a set of flexible yet constrained functions in order to comply with typical demand model properties such as convexity in prices and marginally decreasing sensitivity to cost. As such properties are often observed in empirical data, the function class provides a tool that may prove relevant in a number of econometric modelling situations.

From the general function class, two specific functions are proposed which are formed from a sequence of connected power-functions. The proposed functions have natural linkages to other spline functions, such as polynomial spline functions and linear splines, e.g. piecewise linear functions. However, the proposed functions differ in important ways. The proposed functions offer consistency with the random utility framework. Moreover, the functions offer a smooth and elegant way of introducing decreasing sensitivity to cost, something that is not accommodated by polynomial spline functions. Compared to linear splines, the proposed function has several advantages. As the functions are differentiable (by construction) the estimation of knot points can be based on standard maximum likelihood estimation techniques and estimated jointly with other model parameters. In addition, the functions provides an elegant way of modelling the tail of the distribution and obtain scale-invariance. This is not possible in a linear specification. Finally, the functions relaxes the problematic assumption of non-smooth preferences, which underlines the piecewise linear model.

To support those who might be interested in further investigating these functions, this paper offers examples of how to estimate such functions for a controlled data setup that resembles a multinomial logit model. The functions are also discussed from a more practical applied perspective with reference to a large-scale national transport model for Denmark.

Future research related to the proposed functions could include: i) two-stage estimation strategies for finding appropriate starting values, ii) search for other function candidates that may complement those proposed in the paper, and iii) more examples of applications.

Acknowledgments

I would like to thank Andrew Daly for his valuable comments during the preparation of an earlier version of the paper. Andrew Daly, who has deep knowledge of transport model econometrics, functional form analysis and the nature of cost-damping, contributed with specific as well as general comments, all of which have improved the paper.

I would also like to thank Joar Lind from the Swedish Road Directorate for developing the R-code for generating the synthetic data and estimating the models. The code is available as supplementary material.

The work in this paper has been carried out in connection with the development of the Danish National Transport Model funded by the Danish Ministry of Transport.

ACCEPTED MANUSCRIPT

References

- Ben-Akiva, M., Lerman, S., 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge.
- Ben-Akiva, M., Daly, A., Gunn, H.F., 1987. Destination Choice Models: Design and Appraisal. In: *Proceedings of the 10th PTRC Summer Annual Meeting, University of Warwick, 1987*.
- Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. *Journal of the Royal Statistical Society B* 26 (2), 211–252.
- Daly, A., 2010. Cost damping in travel demand models. Technical report, Rand Europe.
URL: http://www.rand.org/pubs/technical_reports/TR717.html
- Daly, A., Carrasco, J. 2009. The influence of trip length on marginal time and money values. In: Kitamura, R. et al. (eds.) *The Expanding Sphere of Travel Behaviour Research: Selected papers from the proceedings of the 11th Conference on Travel Behaviour Research*. Emerald Books.
- Daly, A., Sanko, N., Wardman, M., 2016. Cost and time damping: evidence from aggregate rail direct demand models. *Transportation*. <https://doi.org/10.1007/s11116-016-9711-9>
- Daly, A. and S. Zachary, 1978. Improved multiple choice models, in D. Henscher and Q. Dalvi (eds.) *Identifying and Measuring the Determinants of Mode Choice*, Teakfields, London.
- Bilbao-Ubillos, J., Fernandez-Sainz, A., Heidenreich, N., Sperlich, S., 2015. Flexible estimation of transport demand functions: recommendations for public policy makers. *Transportation Letters* 7:5, 241-251
- Gaudry, M., 2010. Quebec-Windsor Corridor High Speed Rail Market Forecast Profiles in Context: Level-of-Service Response Curvature Sensitivity and Attitude to Risk or to Distance in Forty Logit Core Model Applications of the Law of Demand, *Agora Jules Dupuit, AJD-127*, www.e-ajd.org.
- Gaudry, M., Jara-Diaz, S., Ortuzar, J.D., 1989. Value of time sensitivity to model specification. *Transportation Research B* 23 (2), 151-158.
- Gaudry, M., Wills, M., 1978. Estimating the Functional Form of Travel Demand Models. *Trans. Research* 12, 257-289.
- Hensher, D., Johnson, L., 1981. *Applied Discrete Choice Modelling*. Croom Helm, London.
- Huang, J. C., Nychka, D., 2000. A nonparametric multiple choice model within the random utility framework. *Journal of Econometrics*, 97(2), 207–225.
- Kneib, T., Baumgartner, B. and Steiner, W. J. , 2007. Semiparametric multinomial logit models for analysing consumer choice behaviour. *Advances in Statistical Analysis*, 91(3), 225–44.
- Layard, R., Nickell, S., Mayraz, G., 2008. The marginal utility of income. *Journal of Public Economics*, 92 1846-1857.

- Lapparent, M., de Palma, A., 2002. Nonlinearities in the Valuations of Time Estimates. In: Proceedings of the 30th PTRC Annual Meeting, PTRC, London (2002).
- Mandel, B., Gaudry, M., Rothengatter, W., 1994. Linear or Nonlinear Utility Function in Logit Models? The Impact on German High-Speed Rail Demand Forecast. *Transportation Research B* 28 (2), 91-101.
- McFadden, D., 1981. Econometric Models of Probabilistic Choice. In: Manski, C. and McFadden, D., Eds., *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, 198-272.
- Pinjari, A.R., Bhat, C.R., 2006. Nonlinearity of Response to Level of Service Variables in Travel Mode Choice Models. *Transportation Research Record*, Vol. 1977, 67-74.
- Rich, J., Mabit, S.L., 2015. Cost damping and functional form in transport models. *Transportation*. 43:889:912.
- Rich, J. and Hansen, C.O., 2016. The Danish National Passenger Model – Model specification and results. *European Journal of Transport and Infrastructure Research*. 16(4): 573-599
- SAS Software, 2016. The NL MIXED Procedure. SAS/STAT(R) 9.2 User's Guide, Second Edition.
- url:
https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#nlmixed_toc.htm
- Tutz, G. and Scholz, T., 2004. Semiparametric modelling of multicategorical data. *Journal of Statistical Computation and Simulation* 74(3), 183–200.
- Williams, H.C.W.L., 1977. On the Formulation of Travel Demand Models and Economic Measures of User Benefit, *Environment & Planning A* 9 (3), 267-285.
- Wold, S., 1974. Spline Functions in Data Analysis. *Technometrics* Vol. 16 (1), 1-11.
- Url: <http://www.stat.ucla.edu/~cocteau/stat204/readings/wold.pdf>

Appendix 1: Pre-calculated spline parameters

Usually, not more than four knot points are needed. Below we provide pre-calculated spline parameters for the standard cases of the log-power function with integer powers.

Log-power spline parameters

Table 4-Table 6 provide spline parameters deducted from Proposition 1. In this case we have already scaled with the β -spline parameter, which is assumed identical across spline intervals. Note that this does not affect the θ_q -parameters as these are invariant to uniform scaling.

q	θ_q	α_q
2	$\frac{1}{2}\ln(c_1)$	$-\ln(c_1)^2$

Table 4: Spline parametrization for the log-power spline class for $Q = 2$.

q	θ_q	α_q
2	$\frac{3}{2}\ln(c_1)$	$-\frac{1}{2}\ln(c_1)^3$
3	$3\ln(c_1)\ln(c_2)$	$-\frac{1}{2}\ln(c_1)[3\ln(c_2)^2 + \ln(c_1)^2]$

Table 5: Spline parametrization for the log-power spline class for $Q = 3$.

q	θ_q	α_q
2	$\frac{4}{3}\ln(c_1)$	$-\frac{1}{3}\ln(c_1)^4$
3	$2\ln(c_1)\ln(c_2)$	$-\frac{1}{3}[\ln(c_2)^4 + 2\ln(c_1)\ln(c_2)^3]$
4	$4(c_1)\ln(c_2)\ln(c_3)$	$-\frac{1}{3}[\ln(c_2)^4 + 2\ln(c_1)\ln(c_2)^3 + 6\ln(c_1)\ln(c_2)\ln(c_3)^2]$

Table 6: Spline parametrization for the log-power spline class for $Q = 4$.

Power spline parameters

For the power spline function it is slightly easier to calculate the scale parameters as these are invariant with respect to Q . For $\pi_1 = 1$ we have;

q	θ_q	α_q
2	$\frac{c_1^{1-\pi_2}}{\pi_2}$	$c_1 \left(\frac{\pi_2 - 1}{\pi_2} \right)$
3	$\frac{c_1^{1-\pi_2} c_2^{\pi_2-\pi_3}}{\pi_3}$	$c_1 \left(\frac{\pi_2 - 1}{\pi_2} + \frac{c_2^{\pi_2}}{c_1^{\pi_2}} \left(\frac{1}{\pi_2} - \frac{1}{\pi_3} \right) \right)$
4	$\frac{c_1^{1-\pi_2} c_2^{\pi_2-\pi_3} c_3^{\pi_3-\pi_4}}{\pi_4}$	$c_1 \left(\frac{\pi_2 - 1}{\pi_2} + \frac{c_2^{\pi_2}}{c_1^{\pi_2}} \left(\frac{1}{\pi_2} - \frac{1}{\pi_3} + \frac{c_3^{\pi_3}}{c_2^{\pi_3}} \left(\frac{1}{\pi_3} - \frac{1}{\pi_4} \right) \right) \right)$

Table 7: Spline parametrization for the power spline class for $Q = 4$.

It is easy to generalize the above expression to the general case.

Appendix 2: Spline parameters for Box-Cox spline

$$\theta_q = \prod_{r=1}^{q-1} c_r^{\pi_r - \pi_{r+1}}, \forall q = 1, \dots, Q - 1,$$

$$\alpha_2 = \frac{c_1^{\pi_1} - 1}{\pi_1} - \theta_2 \frac{c_1^{\pi_2} - 1}{\pi_2},$$

$$\alpha_3 = \alpha_2 + \theta_2 \frac{c_2^{\pi_2} - 1}{\pi_2} - \theta_3 \frac{c_2^{\pi_3} - 1}{\pi_3},$$

$$\alpha_4 = \alpha_3 + \theta_3 \frac{c_3^{\pi_3} - 1}{\pi_3} - \theta_4 \frac{c_3^{\pi_4} - 1}{\pi_4}$$

and

$$\alpha_q = \alpha_{q-1} + \theta_{q-1} \frac{c_{q-1}^{\pi_{q-1}} - 1}{\pi_{q-1}} - \theta_q \frac{c_{q-1}^{\pi_q} - 1}{\pi_q}.$$

ACCEPTED MANUSCRIPT