LAP
LIMA AIRPORT PARTNERS

AEROPUERTO INTERNACIONAL JORGE CHÁVEZ
Avenida Elmer Faucett s/n – Callao
Edificio Central, Piso 7
T (511) 517 3100
www.lima-airport.com

C-LAP-GPF-2024-0179

Callao, 19 de Noviembre de 2024

Señores
ORGANISMO SUPERVISOR DE LA INVERSIÓN EN INFRAESTRUCTURA DE TRANSPORTE DE USO PÚBLICO – OSITRAN
Presente. –

Atención : Dra. Veronica Zambrano
Presidenta Ejecutiva

Sr. Ricardo Quesada Oré
Gerente de Regulación y Estudios Económicos

Asunto : Propuesta: Revisión tarifaria de oficio del Factor de Productividad en el Terminal Portuario General de San Martín – Pisco para el periodo 2025-2030

Referencia : a) Resolución de Presidencia N° 0065-2024-PD-OSITRAN
b) Informe “Propuesta: Revisión tarifaria de oficio del Factor de Productividad en el Terminal Portuario General San Martín – Pisco para el periodo 2025-2030”

De nuestra consideración,

Nos dirigimos a ustedes con relación a la resolución a) de la referencia, mediante la cual se aprueba el informe de la referencia b) y se otorga un plazo para que los interesados remitan por escrito sus comentarios o sugerencias al OSITRAN. Al respecto, Lima Airport Partners S.R.L. (LAP) tiene los siguientes comentarios.

De acuerdo con el informe de la referencia b), en el “capítulo V. Cálculo del factor de productividad elaborado por estas Gerencias”, subcapítulo “V.1.2. Cálculo del índice de cantidades de insumos”, sección “V.1.2.3. Capital” se indica lo siguiente:

> *“171. Para el cálculo de las unidades de capital, se procede a dividir el valor de stock de capital entre un precio representativo para cada categoría de activo, toda vez que no se dispone de la información de precios específicos de cada categoría, es decir, se hace una construcción indirecta de la serie de unidades de capital.*
>
> *172. De este modo, se emplea como variable proxy del precio representativo de las categorías de activos al IPME y al IPMC, de acuerdo con la naturaleza del activo analizado. Se considera como precio representativo al valor promedio del índice elegido mensual de enero a diciembre de cada año; sin embargo, dado que el valor del stock de capital se encuentra expresado en dólares, resulta necesario realizar un ajuste por tipo de cambio. Las series antes señaladas, así como los respectivos ajustes por tipo de cambio, son presentados en el cuadro siguiente […]*
>
> *[…] 173. En tal sentido, se ha procedido a identificar los principales componentes de cada uno de los activos considerados en el cálculo del factor de productividad del TPGSM, ello con el objetivo de seleccionar el índice de precios (IPME o IPMC) que debe ser considerado como una variable proxy de su precio.*

174. **Cabe indicar que los índices de precios (IPME e IPMC) son utilizados también para estimar el precio del capital. Sobre ello, al estimar el precio del capital de la categoría de activos fijos “Instalaciones” y de algunas de las categorías de activos intangibles, se observa que el precio de capital, en el año 2022, resulta ser negativo, lo cual, en el presente caso no tiene interpretación económica. Con la finalidad de corregir ello, para dichas categorías se utilizará el IPME ajustado por tipo de cambio como índice de precios.**[1] En el siguiente cuadro se detalla el índice de precios que se utiliza para deflactar cada categoría de activo de la serie de stock de capital del TPGSM.”

Como se puede observar, OSITRAN utiliza el Índice de Precio de Maquinaria y Equipo (IPME) y el Índice de Precios de Materiales de Construcción (IPMC) para calcular las cantidades de Capital con el objetivo de calcular, posteriormente, el precio de alquiler de capital a partir de la fórmula planteada por Christensen y Jorgenson (1969), la cual se detalla a continuación:

$$\widehat{W}_{m,t}^{K} = \frac{r_t * \rho_{m,t-1} + \bar{\delta}_m * \rho_{m,t} - (\rho_{m,t} - \rho_{m,t-1})}{1 - u_t}$$

Donde:

$\widehat{W}_{m,t}^{K}$	: Precio del capital correspondiente al activo m en el año t.
r_t	: Costo del capital de la empresa en el año t.
$\bar{\delta}_m$	: Tasa de depreciación correspondiente al activo m.
$\rho_{m,t}$	: Precio representativo del activo m en el año t.
$\rho_{m,t-1}$	: Precio representativo del activo m en el año $t-1$.
u_t	: Tasa impositiva de la empresa en el año t.

Al respecto, el regulador manifiesta que para la categoría de activos fijos “Instalaciones” y para algunas de las categorías de activos intangibles no es posible utilizar el IPMC debido a que el precio de alquiler de capital resulta negativo, lo cual, no tendría interpretación económica. Por lo tanto, en su lugar, el IPMC es reemplazado por el IPME. De esta manera, no se registra un precio negativo. Sin embargo, no resulta ser tan consistente utilizar el Índice de Precios de Maquinaria y Equipo (IPME) sobre aquellos activos que se encuentran relacionados a Infraestructura. Lo más adecuado sería utilizar el Índice de Precios de Materiales de Construcción (IPMC).

Para analizar la dificultad planteada, es importante centrarse en el componente **“($\rho m,t - \rho m,t-1$)”** de la fórmula antes mencionada, ya que dada la naturaleza de la fórmula, existe la posibilidad de que una alta variación entre los índices de precios resulte en un precio proxy de capital negativo. Aunque el cálculo matemático pueda arrojar un número negativo, en la práctica económica, el precio de capital no puede ser negativo. Por tanto, un resultado cercano menor a cero indica una situación límite en la que el valor del capital es mínimo. Ante este escenario, podría rescatarse el resultado de la fórmula original mediante metodologías desarrolladas previamente en situaciones similares. Para ello, el regulador podría utilizar las siguientes metodologías para determinar el precio proxy del alquiler de capital:

- **Transformación seno hiperbólico de área**

  $Pproxy = sinh^{-1}(p) = ln(p + \sqrt{p2 + 1}) \approx sign(p) * ln(2|p|)$

[1] En el caso de las categorías “IC: Reubicación de las torres de enfilamiento del Terminal” y “IC: Cerco perimétrico norte” no se aprecian precios negativos del capital, por lo que se utilizará el IPMC ajustado por el tipo de cambio.

- **Transformación seno-arcoseno hiperbólico**
  $Pproxy = sinh\,(\varepsilon + \sigma * arcsinh(p))$

- **Transformación logarítmica**
  $Pproxy = \ln\,(p - \varepsilon)$

Enviamos en adjunto, literatura que ha analizado este tipo de situaciones relacionadas a precios negativos como Sewalt y De Jong (2003), Jones y Pewsey (2009) y Crastes (2021), la cual debería ser utilizada por el regulador para evitar asignar índices de precios que no se encuentran asociados a la característica del activo.

Agradeciendo de antemano la atención a la presente, quedamos de ustedes.

Atentamente,

LIMA AIRPORT PARTNERS S.R.L.

MARIA ELENA REAÑO
Apoderada

Adj.
Sewalt y De Jong (2003)
Jones y Pewsey (2009)
Crastes (2021)

Negative Prices in Electricity Markets

In this paper we describe how liberalisation has lead to the segmentation of trading opportunities for electricity with different periods to delivery. We clarify the price characteristics in each segment, including the extreme volatility in short-term prices and the phenomenon that electricity prices can become negative close to the time of delivery. With the Dutch market as an example, we show the implications for risk management and the valuation of derivatives. We argue that a distinct price model is required for risk management and derivative valuation in each market segment. Derivative valuation goes beyond the financial contract itself and can be very useful for taking strategic decisions on flexible generation assets.

By **MICHAEL SEWALT & CYRIEL DE JONG**[1]

LIBERALISATION AND DEREGULATION in electricity markets have resulted in active trading between generators, suppliers, distributors, large end users and several intermediaries for hedging and speculation purposes. In sharp contrast to conventional markets trading has been clearly segmented, both geographically as well as in terms of delivery period. Geographical segmentation is the result of limited cross border transport opportunities and different regulations per country. Most noteworthy however is the segmentation that is due to the non-storability of the commodity: separate trading mechanisms and markets exist for electricity with different periods to delivery, ranging from long-term forward markets to (very short-term) imbalance markets. Each market segment is characterised by distinct price characteristics that provide a challenge for risk management, derivative valuation and

> Each market segment is characterised by distinct price characteristics <

asset optimisation. In this paper we clarify the properties of the market segments and focus on the extraordinary characteristics of very short-term prices. We highlight the implications for option valuation, which provides the means for valuing flexible generation assets.

The interest in option valuation stems from the limited liquidity and large pricing differences between electricity options. Therefore, market prices do not provide the desired benchmark on which to base strategic decisions. For example, if a generation plant can be treated as an option on spot electricity prices, then we would ideally value this plant based on tradable options. The illiquid market and the lack of valuation benchmarks is partially due to the inapplicability of standard pricing models to price electricity options with short periods to delivery, especially when the short-term electricity prices become negative. This motivates the growing attention for option pricing models for electricity in general and our focus in this paper on the phenomenon of negative prices. Option premiums on short-term prices are much higher than can be expected from standard pricing models as Black-Scholes (1973) or Black (1976)[2]. We therefore analyse the requirements that a pricing model should fulfil for the shortest-term (quarter-hourly imbalance) prices. An appropriate price model would also improve strategic decisions on flexible real assets.

The Forward Market

The ongoing liberalisation of electricity markets has resulted in a relatively liquid trade of longer-term contracts between several market participants. Popular forward contracts are week-ahead, coming months, quarters and years. The major part of the trading is settled OTC where the parties come together, sometimes facilitated by brokers. Even though some exchanges are quite successful (e.g. Nord Pool, EEX), in those markets OTC trades still form a major portion of the total trading in forwards. OTC trading is facilitated with the adoption of master agreements, which increasingly follow the standards of the European Federation of Energy Traders (EFET). In addition, information providers such as Platt's and Heren provide some transparency by publishing forward prices.

Market participants mainly organise electricity trading on a country-by-country basis in the form of country desks, because national grids still have their own procedures and limited exchange capacities between them. Prices in the forward market are quite well comparable to those in other commodity markets. Volatility is limited and forward returns conform to the normality assumptions pretty well, as shown in Table 1: skewness and kurtosis do not deviate significantly from 0, so prices exhibit few outliers. Because of this price behaviour and because hedging with forwards is possible to some extent, the standard Black (1976) formula may be applied to value European-style call and put options on forwards.[3] Those instruments are traded on the Nord Pool exchange and OTC, and provide a means to manage longer-term risks. Apart from limited excess kurtosis and skewness, Table 1 also highlights a first indication of term-structure effects in forward prices: shorter-term forwards experience higher volatility than longer-term forwards. This effect is much stronger in electricity markets than most other markets and is due to the non-storability of the commodity, which prevents arbitrage between periods.

The Spot Market

Forward trading is mostly organised without the need to trade on an exchange. On the other hand, spot trading for day-ahead delivery is largely conducted on organised spot markets such as (in Europe) those from Nord Pool, EEX, APX, UKPX, COMEL and Powernext. The advantage of centralised markets is not only an increase of price transparency, but also a reduction in credit and counterparty risk. In the Netherlands for example, the APX takes full responsibility for counterparty risk, like a general clearing institute, and facilitates the exchange of power. On a daily basis potential buyers and sellers can hand in bids and offers for power on a specific hour for the day ahead. Based on the resulting supply and demand curves a market clearing price and a market clearing volume are determined for every single hour the next day. A transaction will be settled by the APX when a bid or ask (buy or sell) is hit. The most important function of the day-ahead market is giving market participants the opportunity to balance their own delivery or procurement on a short-term basis. Both before and after settlement on the exchange spot trading also takes place on OTC markets. However, the advantage of the exchange is that it looks after the financial settlement and guarantees the physical delivery. Therefore, the counterparty risk is fully reduced in contrast with bilateral agreements.

Table 1. Return Properties of Electricity Contracts

	Volatility	Skewness	Excess Kurtosis
M1	32.9%	-0.36	-2.14
M2	20.3%	-0.83	1.00
M3	14.3%	-0.01	-0.12
Q1	15.7%	0.04	-0.02
Q2	8.6%	-0.18	2.64
Q3	8.7%	-0.25	1.05
Y1	7.4%	0.13	2.11

Forward statistics are based on weekly (5-daily) returns of German baseload contracts (volatility is annualised). Period: Jan 2002 – Mar 2003.
Source: Platts/Moneyline.

Being much closer to delivery than forward contracts, day-ahead spot price dynamics are inherently different from forward price dynamics. Since spot price changes are not normally distributed, the standard Black (1976) model is inappropriate for valuing (daily exercisable) options, caps, floors or collars. Distinguishing features of the prices are a strong level of mean-reversion, seasonality (across seasons and weekdays), extreme and possibly time-varying volatility (reaching daily levels of 1,000%), and occasional spikes. These characteristics have extensively been analysed by academics and practitioners alike and different modelling approaches have been proposed. A common approach is a mean-reverting model with stochastic jumps to account for occasional spikes. Since spikes are often very short-lived, the stochastic jump process (which assumes a long-lasting impact of spikes) does not work well for some electricity spot markets. A recent development is the application of regime-switches in which spikes are modelled as a separate price process. The advantage of a separate spike regime is that it better reflects the temporary nature of spikes. An additional advantage of this approach is that it allows (under certain conditions) for the derivation of closed-form formulas for European-style options on spot prices. Those formulas or simulation-based methods may be applied to value flexible end-user contracts with caps or floors, tradable daily exercisable options and generation assets with the flexibility to manage output on a day-to-day level. Outcomes of those approaches will generally be quite distinctive from standard option-formulas and yield much higher values especially for out-of-the-money call options. This may lead to price caps being sold too cheaply and flexibility in energy generation plants being valued too low.

> Prices in the forward market are quite well comparable to those in other commodity markets <

Simply modelling baseload or peakload prices is not sufficient for some options and the valuation of flexible generation: hourly prices should instead be modelled. Since spikes normally last for several hours in a row and revert back to normal levels more gradually, it is not so convenient to transfer spikes on an hourly level to a spike regime. Approaches that have been applied instead are state-of-the-art time-series models (Guthrie and Videbeck (2002), Cuaresma et al (2002)). The main challenge here is to accurately capture the interdependencies between prices on the same day and between similar hours on different days. In Table 2 we observe for example that the correlation between hours that are 7 hours apart is lower (0.21) than between a single hour on consecutive days (0.61). A similar complex interdependency exists in volatilities of prices (final column). Finally, some sort of jump behaviour (positive for peak-hours, negative for off-peak hours) may be required to capture the outliers. Monte Carlo simulations of hourly price models are used in risk management applications and form the basis for the val-

> The Black (1976) model is inappropriate for valuing daily exercisable options <

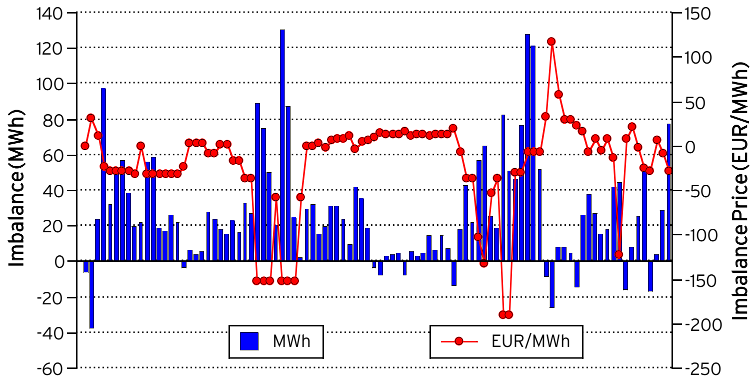
Table 2. Correlations Between Hourly Spot Prices

Time Lag	Correlation: Price	Correlation: Squared Price
1	0.78	0.63
2	0.63	0.48
3	0.51	0.37
4	0.44	0.34
5	0.34	0.23
6	0.27	0.20
7	0.21	0.16
24	0.61	0.60
48	0.36	0.31

Correlations between hourly spot prices and squared spot prices on the Amsterdam Power Exchange.

Figure 1. Imbalance Results

Quarter-hourly imbalance and imbalance prices (average of supply and extraction price) on 9 March 2003 for the Dutch power market.



Source: Tennet

uation and optimal management of assets that may be operated on an hourly level.

The Imbalance Market

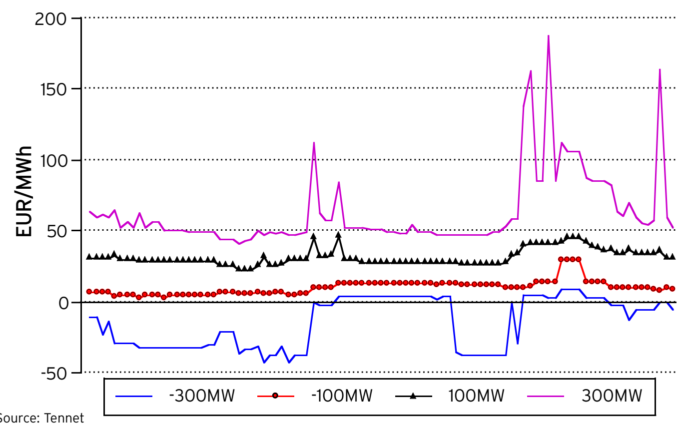
In order to keep an electricity network functioning, the balance of power (supply equals demand) must be maintained

> Price modelling & options valuation on imbalance markets is still largely unexplored <

at all times. In the Netherlands for example 34 market participants have the so-called 'programme responsibility', which includes the requirement to supply a daily schedule of expected supply and demand on a quarter-hourly basis. Just before the electricity is generated and consumed, the network operator TenneT settles any discrepancies between forecasted and actual supply and demand (see Figure 1 for the discrepancies on 9th March 2003). Apart from its own emergency capacity that it may use, the network operator organises an imbalance market on a daily basis to 'smooth' the discrepancies. On this market variable capacity may be offered to TenneT: participants can bid on both increasing or decreasing their supply or demand (Figure 2). A quarter of an hour before delivery, TenneT determines the required capacity. This results in 96 imbalance prices for each of the 96 daily quarters of an hour. Market participants with a negative imbalance pay according to the imbalance market results and participants with a positive imbalance earn according to the imbalance market results. Given the technical problems of an immediate shutdown or start-up of a facility the imbalance market is much more volatile than the spot market, which on its turn is much more volatile than the forward market. Price modelling and option valuation on imbalance markets is still largely unexplored; that's why we explore this topic in more detail.

Figure 2. Imbalance Bid Curves

Quarter-hourly bid curves on 9 March 2003 for the Dutch power market. Each line (-300, -100, 100, 300 MW) represents the price for which power generators are willing to supply power up to the specific volume.



Source: Tennet

Negative Imbalance Prices

In this paragraph we explore a unique phenomenon in electricity markets: negative prices. Negative prices mean that the destruction of the commodity has more value than its creation: electricity is a waste product and is dumped on the market. How does this situation arise and why can electricity be seen as a waste product? As discussed before, there must always be a balance between supply and demand on a power network. Primarily during the night power supply can be higher than demand. This nightly imbalance is caused, for instance, by the installation of combined-cycle facilities and the so-called must-run character of non-flexible generators. Combined-cycle installations are basically installed for the generation of heat (steam) whereby electricity is a co-product. Reducing the must-run output is hardly possible from a technical perspective or it involves high shutdown costs.

Negative prices are acceptable to power suppliers because the opportunity costs of a shutdown period are sometimes much higher. Generally, prices will be negative in only a short period of time and mainly during the night. However, Figure 1 shows that negative prices can sometimes last for long periods of time and can attain extreme levels. The graph contains the imbalance market results in the Netherlands on March 9th 2003; it shows that the market is very volatile and prices can jump from -190 €/MWh up to +120 €/MWh within two hours. When negative prices last for a longer period (corresponding to a positive imbalance), shutting down generation capacity will pay off and imbalance prices will automatically increase.

Negative prices cause sizeable operational problems, for example in energy risk management systems. Not all systems can handle a negative deal in their VaR-calculations, cash-flow projections or invoice procedures. An even larger challenge is the appropriate modelling of negative prices for optimisation and realistic valuation of the most flexible generation assets.

Modelling Negative Prices

Since options on imbalance prices are barely traded, modelling imbalance prices is mainly for risk management, and for optimal management and valuation of the most flexible generation units. A flexible unit will be generating power when the facility is 'in-the-money', meaning that the spark spread (equal to commodity price minus variable cost of fuel) is positive. Many energy risk management systems use the so-called delta hedging strategy [4] in order to forecast the optional production capacity in advance. This strategy is based on the same assumptions as other standard option formulas. These models assume that electricity and fuel prices evolve according to a gradual process ('Brownian Motion') with no extreme changes, mean reversion or negative prices: it assumes prices are lognormally distributed.

To understand how negative prices can be dealt with, it is important to understand why negative prices lead to modelling problems. The problems all stem from the fact that standard price models are based on price returns (or in fact logreturns). A return becomes in fact infinite when prices approach zero and is not defined at all for negative prices.[5] Incorporating negative prices can basically be achieved with two approaches: an indirect (structural) approach and a direct approach. The structural approach does not model prices directly, but models them as the outcome of a price formation process. This process may include for example the imbalance (Figure 1) and the imbalance bid curves (Figure 2), from which imbalance prices result. A structural approach offers valuable insights in the formation of prices and is appealing to industry professionals, who 'recognise' in it the functioning of the market. However, for risk management systems they easily become too complex, because they need to contain several stochastic variables (such as imbalance and imbalance bid curves), which provide challenging modelling and implementation tasks by themselves.

A direct price modelling approach is not straightforward either, but at least reduces the problem to one variable: the price. We propose to allow for negative prices by setting a lower bound on the actual price and re-scaling prices with respect to this lower bound. An important advantage of this approach is that the lower bound can be based on economic rationales and market experience. It also allows for the extreme positive outliers, while limiting the negative outliers in prices. Moreover, this approach permits the usage of relatively standard time-series models on the re-scaled price returns. Imbalance prices exhibit sudden jumps and similar complex interrelations as those in hourly spot prices: within days and across days in both price levels and price volatility. For a realistic model it is necessary to include both types of interrelations. We suggest that a combination of the periodic autoregressive model in Guthrie and Videbeck (2002) and specifications that model each time period separately (Cuaresma et al., 2002) can achieve this goal.

Conclusions

Options are helpful products for managing unexpected price and volume fluctuations. Given the high volatility of short-term electricity prices it could be expected that electricity options are very popular. However, currently bilateral options are traded in the OTC market only on a small scale, and the exchange trade of options is even lower. The few options that are traded are mainly on forward contracts, such as for example an option on the forward 2004. But also options on day-ahead spot prices (daily exercisable options) are traded occasionally. A major explanation for the low trading volume is the difficulty to value those contracts. When current methods of option valuation are improved, the trading of options could be stimulated. Before trading in short-term options really takes off, option valuation techniques are already required: to manage and value flexible generation assets. In order to determine whether an investment is worthwhile or not a power generator can be considered as an option on power production. This method of real option valuation becomes more and more familiar. However for a proper valuation of the most flexible generation assets, it is important that standard option pricing models will be adjusted to price spikes as well as negative prices.

> Imbalance prices exhibit similar complex interrelations as those in hourly spot prices <

The study of negative commodity prices has made clear that negative prices have a special impact on option pricing models. Standard models as Black-Scholes (1973) and Black (1976) are not applicable to options with negative underlying value. New models can be very helpful for financial options as well as for strategic decisions (real options) like the management of flexible generation assets ■

Footnotes

1. The authors thank Kasper Walet (Maycroft Consultancy) and Gerard van Baar (Deloitte & Touche) for helpful comments.

2. Black's (1976) model is similar to the famous Black-Scholes (1973) model, but applicable to options on forwards and futures.

3. If returns are skewed or exhibit clear kurtosis the extended Black formula with separate terms for skewness and kurtosis (Jarrow and Rudd, 1982) yields more reliable results.

4. Without the possession of the option in reality the increase or decrease of an option value can be optimally simulated by holding an amount of the underlying asset equal to the option delta.

5. For example, what is the return if prices increase from -10 to +10?

MICHAEL SEWALT is consultant at Deloitte & Touche's Energy & Commodity Risk Management Centre of Excellence. msewalt@deloitte.nl

CYRIEL DE JONG is professor at Erasmus University in Rotterdam and consultant in Energy Risk Management and Derivative Valuation.
cjong@fbk.eur.nl

Literature

De Jong, C. and R. Huisman, 2003, "Option formulas for power prices with spikes", Energy & Power Risk Management, February, Germany special issue

Jarrow, R. and A. Rudd, 1982, *"Approximate option valuation for arbitrary stochastic processes"*, Journal of Financial Economics, 10, 347-369.

Guthrie, G. and S. Videbeck, 2002, *"High-frequency electricity spot price dynamics: an intraday markets approach"*, research paper, New Zealand Institute for the study of competition and regulation.

Cuaresma, J.C., J. Hlouskova, S. Kossmeier and M. Obersteiner, 2002, *"Forecasting electricity spot prices using linear univariate time series models"*, working paper, University of Vienna.

A new shifted log-normal distribution for mitigating 'exploding' implicit prices in mixed multinomial logit models

Romain Crastes dit Sourd*

June 28, 2021

Abstract

This paper introduces a new shifted negative log-normal distribution for the price parameter in Mixed Multinomial logit models. The new distribution, labelled as the μ-shifted negative log-normal distribution, has desirable properties for welfare analysis and in particular a point-mass which is further away from zero than the negative log-normal distribution. This contributes to mitigating the 'exploding' implicit prices issue commonly found when the price parameter is specified as negative log-normal and the model is in preference space. The new distribution is tested on 10 stated preference datasets. Comparisons are made with standard alternative approaches such as the willingness-to-pay space approach. It is found that the new μ-shifted distribution yields much lower mean marginal WTP estimates compared to the negative log-normal specification (up to 99% lower) and similar to the values derived from a multinomial logit while at the same time fitting the data as well as the negative log-normal specification and much better than the willingness-to-pay space approach.

JEL Codes: Q51, C35.

*Centre for Decision Research, Management Division, Leeds University Business School and Choice Modelling Centre, University of Leeds.

1 Introduction

Choice models fitted to stated and revealed preference data for non-market valuation commonly employ specifications aiming at representing random heterogeneity in taste across decision makers (Hoyos, 2010). The use of mixed multinomial logit models is widespread not only in Environment but also Transport and Health for modelling choice data (and more commonly data from stated preference surveys). This category of models includes all the variants of the mixed logit family such as the generalized multinomial logit (Fiebig et al., 2010) and the logit mixed logit model (Scarpa et al., 2020; Train, 2016), all of which provide ways to derive marginal willingness-to-pay (WTP) distributions rather than point estimates. Both model fit and WTP estimates can vary greatly depending on model specification and the adequate modelling strategy is situation specific.

The literature often describes the model specification search as a trade-off between deriving reasonable WTP estimates and model fit to the data, where reasonable refers to an estimate for the distribution of WTP that is not extremely skewed and in line with the values found in other papers or derived from other revealed or stated preference methods (Sonnier et al., 2007; Train and Weeks, 2005). Perhaps the most emblematic illustration of this trade-off is the distinction between models estimated in preference space and models estimated in WTP space. Each specification is simply a re-parametrisation of the other and the main difference is the range of distributions allowed by each specification. A WTP space model for which the price is not specified as randomly distributed is identical to its counterpart in preference space. The WTP space approach allows to directly specify the WTP distribution for each of the (non-monetary) attributes. In preference space, the distribution of the utility derived from each attribute is specified by the analyst and the WTP distributions are derived from the ratio of the distribution of a given non-monetary attribute and the distribution of the price attribute (which requires the use of simulations). Performing such a ratio might give rise to the "exploding" implicit price problem depending on the distributional assumptions made by the analyst.

In most cases, assuming no random heterogeneity in preferences across decision makers for the price attribute yields reasonable WTP values but is behaviourally implausible and might be rejected against competing hypothesis, of which one of the most common is to assume that the price attribute is negative log-normal. The negative log-normal distribution has a point mass near zero, which can potentially introduce very small values in the denominator of the ratio of simulated distributions from which WTP measures are derived in preference space. This is an issue which has also been found (although to a lesser extend) when using a log-uniform distribution for the price attribute as suggested by Hess et al. (2017). The presence of a few large values in the simulated WTP distribution for a given non-monetary attribute can have an extreme influence on the mean of the distribution (and causing it to "explode", that is to reach implausibly large values). At the same time, models in preference space have often been found to fit the data better than models in WTP space, because the distributional assumptions allowed by working in preference space are simply more likely to better accommodate extreme preferences for a given attribute (or lower price sensitivity for some decision makers). This leads Train and Weeks (2005) to suggest that "*research is needed to identify distributions that fit the data better when applied in WTP space and/or provide more reasonable distributions of WTP when applied in preference space*".

Several authors have engaged with this proposition in the past and most of their work has consisted in finding distributions that yield more reasonable WTP estimates in preference space. As previously mentioned, Hess et al. (2017) proposed to used a negative log-uniform distribution for the price attribute in preference space. However, such a distribution might not fit the data as well as the negative log-normal distribution and still produces a few outliers which can have a strong impact on mean WTP estimates.

Other authors have suggested to use a normal distribution for the monetary attribute (Svenningsen and Jacobsen, 2018). However, Daly et al. (2012) have demonstrated that such an approach can prevent the existence of moments for the WTP distribution of the non-monetary attributes. A different approach has been suggested by Giergiczny et al. (2012), who proposed to introduce a cost-income variable in their model in an attempt to shift the (negative log-normal) distribution of the sensitivity for the price attribute away from zero. It has not become a standard approach in the literature, which is likely because of the additional data requirements (the income of each decision maker needs to be know) as well as because of the unorthodox way the cost-income ratio is introduced in the model (it is an interaction variable which is introduced by dividing the price attribute instead of multiplying it).

In this paper, we propose a new approach for deriving more reasonable WTP distributions from models estimated in preference space based around the shifted (negative) log-normal distribution. This approach is inspired by the work of Giergiczny et al. (2012) in the sense that it consists in shifting the distribution of the price sensitivity away from zero but, similarly to Hess et al. (2017) for example, it relies solely on a new specification for the price attribute and does not require to collect additional data to be implemented. The shifted log-normal distribution (also known as the three-parameters log-normal distribution) has been introduced by Sangal and Biswas (1970) and has been used in hydrology, among other fields. The shifted log-normal distribution is simply obtained by adding a shift parameter to a classic log-normal distribution (as defined in the remainder of this paper). In what follows, we demonstrate that the three parameters log-normal distribution is unable to provide moments for the WTP distributions when the shift parameter is found to be positive at model convergence. In addition, a recent contribution from McFadden and Robles (2019) shows that the three-parameters log-normal also suffers from identification issues which are further commented on in the remainder of this article. We propose two new specifications for tackling down these issues, namely the κ-shifted and μ-shifted log-normal distributions. The μ-shifted is of particular interest as it is found to always ensure the existence of moments for the WTP distributions and provide much more reasonable WTP estimates in comparison to the log-normal and the log-uniform distribution (and not significantly different from the multinomial logit model).

We test the new approaches derived from the shifted log-normal distribution on 10 datasets from the non-market valuation and transportation literature. Using a large number of datasets instead of a single one allows to show how well the proposed distribution performs in a large range of empirical contexts. For each dataset, we estimate a series of models and compare new and existing common parametrisations (multinomial logit, mixed logit with a non-random price attribute, log-normal price, log-uniform price, WTP-space). We estimate 90 models and derive WTP distributions in total. This large amount of data allows to use meta-analysis techniques to compare the new-shifted distributions to other existing approaches. We find in particular that the new μ-shifted parametrisation consistently leads to goodness-of-fit measures which are nearly identical or equal to the best fitting model for a corresponding dataset, and outperforms the corresponding WTP-space model in all cases. At the same time, the proposed approach yields WTP estimates which are not significantly different from those derived from a multinomial logit model, and far more reasonable than the values derived from models featuring a log-normally distributed price parameter (between 9.89% and 99.99% lower).

This paper is hence organised as follows. The next section presents the new shifted distributions. Section 3 describes the framework for empirical testing used to compare the new shifted distributions to existing alternatives in the literature. Section 4 introduces the data used in this analysis. Section 5 presents and discusses results. Section 6 concludes.

2 Modelling work

2.1 The Mixed Multinomial Logit framework

We start by describing the well-known Mixed Multinomial Logit (MMNL) specification in preference space. Let U_{int} be the utility that respondent n derives from alternative i in choice situation t. The utility includes a modelled component V_{nit} and a random component ε_{int} which follows a type 1 extreme value distribution. We have:

$$U_{int} = V_{int} + \varepsilon_{int} \tag{1}$$

$$V_{int} = ASC_i + \beta_n' x_{int} + \rho_n cost_{int} \tag{2}$$

where β_n is a vector of taste coefficients (excluding the sensitivity for the cost), x_{int} a vector of attributes for alternative i, ρ_n corresponds to the sensitivity for the cost attribute for respondent n and $cost_{int}$ corresponds to the value for the cost attribute for the alternative i faced by respondent n in choice situation t. We include alternative specific constants (ASCs) for all but one of the alternatives. The probability that respondent n chooses a given alternative i conditional on β_n, ρ_n and the ASCs in choice situation t corresponds to the well know Multinomial logit (MNL) probability. The elements in β_n and ρ_n can be allowed to vary randomly across respondents. A common assumption in non-market valuation is to assume that the elements in β_n are normally distributed.

$$\beta_{kn} = \mu_k + \sigma_k \zeta_{kn} \tag{3}$$

where μ_k corresponds to the mean and σ_k the standard deviation of the random parameter. In this example, ζ_{kn} is a random disturbance distributed $N(0,1)$ which is very common in non-market valuation, but other distributions (Uniform, Triangular, etc) can be used too. Recent contributions from Fosgerau and Mabit (2013) and Train (2016) (see also Scarpa et al. (2021)) provide different ways of introducing more flexible distributions in the mixed logit framework. For example, following Fosgerau and Mabit (2013), it is possible to specify β_n as a second order polynomial of a standard normal random variable as follows:

$$\beta_{kn} = \mu_k + \sigma 1_k \zeta_{kn} + \sigma 2_k \zeta_{kn}^2 \tag{4}$$

In some cases, the distribution of a given attribute needs to be constrained for behavioural reasons or for ensuring the existence of moments for the distribution of marginal WTP estimates. This is for example the case for the price attribute, which is generally assumed to be log-normal.

$$\rho_n = -e^{(\mu_{price} + \sigma_{price} \zeta_{price,n})} \tag{5}$$

As the actual value of β_n and ρ_n for a given respondent is not observed by the analyst, the choice probabilities are given by a multi-dimensional integral of the MNL probabilities. It is worth noting that constraining the standard deviation σ of all the random parameters to be zero leads to an MNL model.

Assuming that the sensitivity for the cost attribute follows a negative log-normal distribution has undesirable features when it comes to deriving marginal Willingness-To-Pay (mWTP) distributions. Indeed, mWTP distributions are derived for each non-monetary attribute by calculating the following ratio using a large number of draws (often several millions):

$$mWTP_{kn} = -\frac{\beta_k n}{\rho_n} \tag{6}$$

The log-normal distribution has a point-mass near zero, which means that the denominator of Equation 6 is likely to reach extremely small values, leading to very large mWTP estimates. These large values have been found to have a considerable impact on mean mWTP estimates. This problem is known as the 'exploding implicit price' problem in the literature (see Giergiczny et al. (2012) for example) and the mean mWTP values derived from such models have been branded as unreasonable or counter-intuitive based on expert knowledge (Scarpa et al., 2008). A series of alternative parametrisations have been suggested in the literature to circumvent this issue. However, each proposition has its drawbacks and a universal solution does not exist. The right modelling approach is always case specific and depends on the data at hand.

2.2 Alternative parametrisations

A popular alternative to the parametrisation presented above is the WTP space approach. The WTP space approach was first suggested by Train and Weeks (2005), although the concept was first put forward in Cameron (1988) in the context of referendum contingent valuation data. A model parametrised in WTP space is obtained by reformulating Equation 2 so that:

$$V_{int} = ASC_i + \rho_n \cdot (\beta_n' x_{int} - cost_{int}) \tag{7}$$

where the elements of β'_n are now directly interpretable as WTP estimates. It is worth noting that not all the elements of β'_n have to be specified in WTP space.

The main difference between the preference space and the WTP space approach is the assumptions made about the distribution of mWTP. Assuming normally distributed preferences for the non-monetary attributes, a model specified in preference space implies that the mWTP distribution for each attribute corresponds to the ratio of a normal distribution and a log-normal distribution, which leads in most cases to a heavy-tailed (on the right side) distribution. On the other hand, a model specified in WTP space implies that mWTP distributions are normally distributed. The fact that the preference space approach has often been found to fit the data better is likely to be driven by the fact that the implied distribution of mWTP has more potential for accommodating extremely high preferences for a given set of non-monetary attributes.

Other solutions often found in the literature consist in constraining σ_{price} to be equal to zero or using other distributions for ρ in preference space such as the negative log-uniform distribution, where $\zeta_{price,n}$ in Equation 5 is then assumed to be distributed $U(0, 1)$. In this paper, we compare the aforementioned parametrisations to a set of new distributions for the price attribute in preference space based on the shifted log-normal distribution as previously introduced.

2.3 Shifted negative log-normal distributions

2.3.1 The three-parameters log-normal distribution

The original shifted negative log-normal distribution introduced by Sangal and Biswas (1970) is simply obtained by adding one parameter to the two parameters negative log-normal distribution described in Equation 5:

$$\rho_n = \kappa - e^{(\mu_{price} + \sigma_{price} \zeta_{price,n})} \tag{8}$$

where κ is a shift parameter to be estimated. In the remainder of this paper, we refer to this specification as the three-parameters log-normal distribution. Introducing κ is expected to contribute to shift the denominator in Equation 6 away from zero and hence mitigate the "exploding ratio" issue. However, this can only be the case if κ is found to be negative. In the event where a positive κ is found, the model can not be used for welfare analysis as the distribution of ρ can span on both sides of zero. This is a problem because this would suggest that no finite mWTP moments exist (this issue is thoroughly documented in Daly et al. (2012)). Constraining κ to be negative *via* an exponentiation is also undesirable. More precisely, if κ is unconstrained and found to be positive, its constrained counterpart is likely to lead to a shift which is extremely close to zero, which in turn can lead to issues during model estimation, in addition to having no or very little impact when it comes to moving the numerator of Equation 6 away from zero.

Another concern regarding the use of a three-parameters log-normal distribution for the price attribute which has been recently pointed-out by McFadden and Robles (2019) is that the parameters μ and κ described in Equation 8 are collinear and were found to be "poorly estimated" in a series of Monte-Carlo simulations. The authors state that "*this is unsurprising since both act to determine the location of the [...] distribution, and have effects that are separately identified only through the behaviour of consumers facing the most extreme high prices*". In this paper, we propose two new specifications of the shifted log-normal distribution which seek to mitigate these issues.

2.3.2 The κ-shifted log-normal distribution

Our first proposition is to constrain μ_{price} in Equation 8 to be equal to zero in order to only let κ act as a location parameter and improve the chances to find a model where the value of κ at convergence is negative. It is worth noting that this does not imply that we assume that the mean of the resulting shifted negative log-normal distribution is zero. Indeed, the mean of a three parameters log-normal distribution is jointly influenced by κ, μ_{price} and σ_{price} as discussed by Sangal and Biswas (1970). Constraining μ_{price} to be zero simply transfers the heterogeneity meant to be captured by the parameter to κ and σ_{price}. We label this specification the κ-shifted negative log-normal distribution (or more simply, in the context of this paper, the κ-shifted distribution). it is formally described as such:

$$\rho_n = \kappa - e^{(\sigma_{price}\zeta_{price,n})} \tag{9}$$

It is worth noting that this solution improves the issue related to the location parameters reported by McFadden and Robles (2019) but still allows κ to be positive, thus leading to potential issues with the existence of mWTP moments.

2.3.3 The μ-shifted log-normal distribution

The second alternative specification for the three parameters log-normal distribution which we propose in this paper consists in replacing the shift parameter κ by the exponential of the mean of the underlying normal distribution, that is $-e^{\mu_{price}}$. Such a specification has the desirable property of ensuring that the shift does not become positive or very close to zero. We label the resulting distribution the μ-shifted log-normal distribution. Formally, we propose:

$$\rho_n = -e^{(\mu_{price})} - e^{(\mu_{price} + \sigma_{price}\zeta_{price,n})} \tag{10}$$

This specification improves both the issue related to the location parameters as well as the need to ensure moments for mWTP estimates. It also ensures that the point-mass is shifted further away from zero with respect to the 2-parameters negative log-normal distribution. Its usefulness is proven *via* an extensive amount of empirical tests which are described in the next section.

3 Framework for empirical testing

We test the usefulness of the new shifted distributions introduced in this paper by comparing their performance with other existing model specifications using 10 datasets introduced in the next section. The different model specifications tested are reported in Table 1 below:

Table 1: Model specifications

Model	**monetary attribute parameterization**	**Additional details**
MNL	$\rho_{price} = \mu_{price}$	No random heterogeneity
Fixed monetary attribute	$\rho_{price} = \mu_{price}$	
LN	$\rho_{price} = -e^{(\mu_{price}+\sigma_{price}\zeta_{price})}$	$\zeta \sim N(0,1)$
LU	$\rho_{price} = -e^{(\mu_{price}+\sigma_{price}\zeta_{price})}$	$\zeta \sim U(0,1)$
WTPS	$\rho_{price} = -e^{(\mu_{price}+\sigma_{price}\zeta_{price})}$	$\zeta \sim N(0,1)$, see Equation 7
WTPS P2	$\rho_{price} = -e^{(\mu_{price}+\sigma_{price}\zeta_{price})}$	$\zeta \sim N(0,1)$, $\beta_{kn} = \mu_k + \sigma 1_k \zeta_{kn} + \sigma 2_k \zeta_{kn}^2$
3-Param LN	$\rho_{price} = \kappa - e^{(\mu_{price}+\sigma_{price}\zeta_{price})}$	$\zeta \sim N(0,1)$
κ-shifted	$\rho_{price} = \kappa - e^{(\sigma_{price}\zeta_{price})}$	$\zeta \sim N(0,1)$
μ-shifted	$\rho_{price} = -e^{(\mu_{price})} - e^{(\mu_{price}+\sigma_{price}\zeta_{price})}$	$\zeta \sim N(0,1)$

All models are estimated with 2,000 mlhs draws (Hess et al., 2006)

The MNL specification is introduced as a baseline benchmark and also because this is still used for welfare analysis in some contexts. The *fixed monetary attribute* specification refers to a model where all the parameters are randomly distributed apart from the monetary attribute, for which no random heterogeneity is assumed. The LN model refers to a model where the monetary attribute is assumed to be distributed negative log-normal and will be one of the main comparison points together with the WTPS model. A negative log-uniform (LU) specification is also introduced and tested. The WTPS P2 model refers to a WTP-space model where each non-monetary attribute is specified as a second order polynomial of a standard normal random variable (or log-normal in a very few cases). The three remaining specifications refer to the three shifted log-normal distribution discussed in this paper (3-PARAM LN, κ-shifted and μ-shifted). The specification of the non-monetary attributes are the same across all the models[1] (apart from the MNL model and the WTPS P2 model). All models are estimated using the *Apollo* package for R (Hess and Palma, 2019).

3.1 Models comparison

A total of 90 models are estimated (9 for each set of data). The models are then compared as follows:

3.1.1 Model fit

We first compare the impact of distributional assumptions for the price attribute on model fit by comparing the log-likelihood as well as the BIC for each model. We do not use likelihood ratio tests because the large majority of models feature the same number of parameters (for a given dataset). Moreover, the only cases where it would be adequate to run a likelihood-ratio tests are for comparing the MNL model and the fixed monetary attribute model to the other models. However, it is well-known that such models result in

[1]For all the datasets, the large majority of the non-monetary attributes have been specified as normally distributed apart from a few which have been specified as log-normally distributed when constraining the sign was necessary.

a significantly poorer fit in a large majority of cases, and testing for this with likelihood ratio tests would only state the obvious.

3.1.2 Welfare estimates

Secondly, we compare the difference between the moments of the resulting mWTP for each model specification and each dataset. In particular, we compare the difference between the 3-parameters LN specification, the κ-shifted specification and the μ-shifted specification with the two parameters LN specification and the WTPS model.

3.2 Meta-analysis on welfare estimates

Meta-analysis techniques have been extensively used in non-market valuation for comparing studies and identify the factors that drive their results. Meta-analysis are also essential in the literature on benefit transfers, where the results from different studies are aggregated *via* a meta-regression in order, for example, to forecast the economic value of the ecosystem services provided by natural sites on which a stated preference survey has not been performed. In the context of this paper, we propose the following process:

i Compute the mean mWTP estimate for each attribute and for each model specification in all the model specifications

ii Each resulting mean mWTP corresponds to an observation

iii The dependent variable of the meta-regression is specified to be the natural logarithm of each mean mWTP, giving rise to a log-linear regression

iv The independent variables of the meta-regression are dummy variables corresponding to each dataset and dummy variables corresponding to each model (minus a base in each case)

Such an approach will allow to isolate the effect of the model specification on mean mWTP estimates and provide confidence intervals for the difference in terms of welfare estimates across model specification. The log-linear approach is favoured over a linear approach because the welfare estimates derived from each datasets might be expressed in different currencies and feature different orders of magnitude. Using a log-linear framework also allows to interpret results as the variations in percentage of the mean welfare estimates driven by the different model specification tested all else being equal.

3.2.1 Differences in distributions

This test consists in plotting the kernel density estimate of the mWTP distributions derived from each model (this excludes the MNL model, from which welfare distributions can not be derived) and compare the common area across model specification. Kernel density estimation (KDE) is simply a non-parametric technique for estimating the probability density function of a random variable. We then use the k-density test for comparing the common area of KDE proposed by Martínez-Camblor et al. (2008). This test allows to assess how similar or different two distributions are. More precisely, the k-density test gives a simple measure of the proximity of two kernel density estimates. This measure, known as the AC statistic, varies between 0 and 1. A value of 0 corresponds to an absolute discordance while a value of 1 corresponds to an absolute match of the distributions. This test will mainly allow to measure how (dis-)similar mWTP distributions are across model specifications for a given dataset.

3.3 Out-of-sample fit

The last test consists in measuring whether the in-sample fit and out-of-sample fit is significantly different across model specifications all else being equal. Again, the data curation and modelling work is organised as follows:

i Each data set is randomly split into two parts: an estimation sample (70% of the respondents) of the dataset and a validation sample (30% of the respondents).

ii For each estimation sample, different models of interest are estimated

iii The estimated parameters for each model and each dataset are used to measure the fit of the model (log-likelihood) on their respective estimation and validation sample.

iv This process is repeated 10 times for each model and each dataset with different (randomly chosen) estimation and validation samples. This is performed in order to increase the sample size for the meta-regressions as well as because performances could vary across different, randomly selected estimation and validation samples.

The datasets on which these tests are performed are introduced in the next section.

4 Data

4.1 Overview

We use ten SP surveys datasets from five different countries (France, Germany, Poland, United Kingdom and United States of America). The SP surveys vary in terms of design (number of attributes, number of choice scenarios, and number of alternatives). Some SP surveys address market goods (car choice, chicken and fish meat choice, theatre play choice, coffee choice) while others address non-market goods (forest protection, erosive run-offs mitigation, water protection). The data from each of these surveys has been used for published methodological or empirical contributions to the SP literature in 9 cases out of 10 (the remaining case being a report for the German government). For each paper, we describe the SP in terms of design (attributes, alternative, levels) and sampling (number of respondents). One reference is given for each dataset. The data is summarised in Table 2 below.

Table 2: Overview of data

Survey	Ref.	Description	Number of alternatives & choice tasks	Nb. of respondents	Attributes	Levels
Normandy flood protection	(Crastes et al., 2014)	Measuring preferences for flood risk mitigation measures in Normandy (France)	3 alternatives including one status quo; 6 choice tasks per respondent	416	*Agri - Improving agricultural practices against floods*	0, 1
					Infra - Improving protections against floods	0, 1
					Com - Improving communication against floods	0, 1
					Price (in)	0, 12.5, 25, 32.5
University of Nantes coffee machines	(Sandorf et al., 2018)	Preferences for upgrading coffee machines at the University of Nantes (France)	3 alternatives including one status quo; 8 choice tasks per respondent	288	*Cardpay - is it possible to pay by card?*	0, 1
					Organic - Organic coffee	0, 1
					Fairtrade - Fairtrade coffee	0, 1
					Recycle - Can the cup be recycled	0, 1
					Price (in)	0.40, 0.45, 0.50, 0.55
Bialowieza Forest	(Bartczak, 2015)	Management choices for improving the quality of the Bialowieza forest (Poland)	3 alternatives including one status quo; 12 choice tasks per respondent	1202	*Cen - Level of "naturalness" of the commercial part of the forest*	0, 1
					Gos - Level of "naturalness" of the second growth forest	0, 1
					Vis1 - Restrictions on number of visitors per day	0, 1
					Vis2 - Restrictions on number of visitors per day	0, 1
					Fee (in zl)	25, 50, 75, 100
Ecological value of Polish forests	(Czajkowski et al., 2014)	Management choices for improving the quality of Polish forests	4 alternatives including one status quo; 26 choice tasks per respondent	1001	*Nat - Partial and substantial improvement in the protection of the forest*	0,1,2
					Tra - Partial and substantial improvement in the amount of litter found	0,1,2
					Inf - Partial and substantial improvement in tourist infrastructure	0,1,2
					Fee (in zl)	10, 25, 50, 100
Endangered fish	(Campbell, 2008)	Policy preferences for protecting endangered fish species in the Republic of Ireland and Northern Ireland	3 alternatives including one status quo; 16 choice tasks per respondent	754	*Ac - Protect Arctic char*	0,1
					As - Protect Atlantic salmon	0,1
					F - Protect Ferox brown trout	0,1
					G - Protect Gillaroo brown trout	0,1
					S - Protect Sonaghan brown trout	0,1
					Income tax (/year)	3, 6, 12, 24, 48
Warsaw Theatres	(Czajkowski et al., 2017)	Public support for discounted theatre tickets	2 alternatives including one status quo; 12 choice tasks per respondent	1569	*Roz - Entertainment theatre is funded*	0,1
					Sro- Drama theatre is funded	0,1
					Dzi - Children's theatre is funded	0,1
					Eks - Experimental theatre is funded	0,1
					Contribution (in zl)	10, 20, 50, 100
Car choice	(Train and Weeks, 2005)	Choice of different vehicles including electric and hybrid vehicles	3 alternatives; up to 15 choice tasks per respondent	500	*Engine type - Gas, electric or hybrid*	0,1,2
					Operating cost (in $ per month)	From 2.51 to 72.29
					Performance - Low, medium or high	0,1,2
					Range (in hundreds of miles for EV)	Up to 200
					Body type (10 types ranging from mini car to large van)	0,1,2,3,4,5,6,7,8,9
					Purchase price (in $)	From 7,018 to 97,301
Chicken	(Campbell and Doherty, 2013)	Value-added services to chicken meat	3 alternatives including one status quo; 12 choice tasks per respondent	816	*Test - Food testing, standard or enhanced*	0,1
					Trace - Traceability, standard or enhanced	0,1
					Well - Animal health, standard or enhanced	0,1
					Origin - Ireland, Great-Britain, EU	0,1,2
					Price (in GBP)	2,2.5,3,3.5,4,4.5,5
Bluerivers	(Horbat, 2017)	Improving the naturalness of German rivers	3 alternatives including one status quo; 6 choice tasks per respondent	2023	*Aue - Surface of protective flood plains*	10,000 ha, 25,000 ha, 50,000 ha
					Wald - Surface of alluvial forests in the floodplains	10%, 30%, 50%
					Ufer - Near natural banks	1000 km, 2000km, 3000 km
					Fisch - Improved river continuity for fish	50%, 75%, 100%
					Baden - Reduction of the entry of waste waters and fertilisers (bad, medium, good)	0,1,2
					Fee (in per year)	15, 25, 50, 100, 170, 250
Nitrolimit	(Meyerhoff et al., 2014)	Value of water quality improvements in the region Berlin-Brandenburg (Germany)	3 alternatives including one status quo; 12 choice tasks per respondent	754	uhavel - Quality improvement in Lower havel	0,1,2,3
					ohavel - Quality improvement in Upper Havel	0,1,2,3
					stadts - Quality improvement in city stretch spree	0,1,2
					koeps - Quality improvement in spree Kopenick	0,1,2
					dahme - Quality improvement in Dahme - Scharmutzelsee	0,1,2,3
					cost (in per year)	10,25,20,75,100,150

4.2 Sensitivity for the monetary attribute

Further information about each dataset can be derived from the analysis of the choices made by respondents as a function of the level of the monetary attribute. More precisely, and given the central importance given to the monetary attribute and its specification in the current paper, we propose to investigate whether the monetary attribute levels presented to the respondents in each dataset were binding or not. For most normal goods, alternatives with a higher price should be less appealing than alternatives with a lower price and hence should be less likely to be chosen. Glenk et al. (2019), citing Mørkbak et al. (2010) as well as Guy and Willis (1999), report that the highest level of the monetary vector should be chosen 'following' a rule of thumb that the alternatives with the highest level for the monetary attribute should not be selected in more than 5 to 10% of the choice situations where it is present. Although the evidence for such a rule is lacking, it is clear that the amounts presented to the respondents must be credible (Johnston et al., 2017) and that datasets for which the chosen alternatives are more often the alternatives for which the monetary attribute level is the highest suggest the presence of issues (with the experimental design, with the engagement of respondents, etc). We follow Glenk et al. (2019) and plot 'bid acceptance curves' for each dataset, which show the cumulative number of times (expressed in percentage) an alternative has been chosen at its cost level or at a lower cost level. Cumulative acceptance rates should decline as the monetary attribute level increases in all cases. Results are reported in Figure 1 below[2].

[2]Results for the car choice dataset were not plotted because the number of levels for the cost attribute was found to be too large. Detailed results for this dataset are available from the author upon request

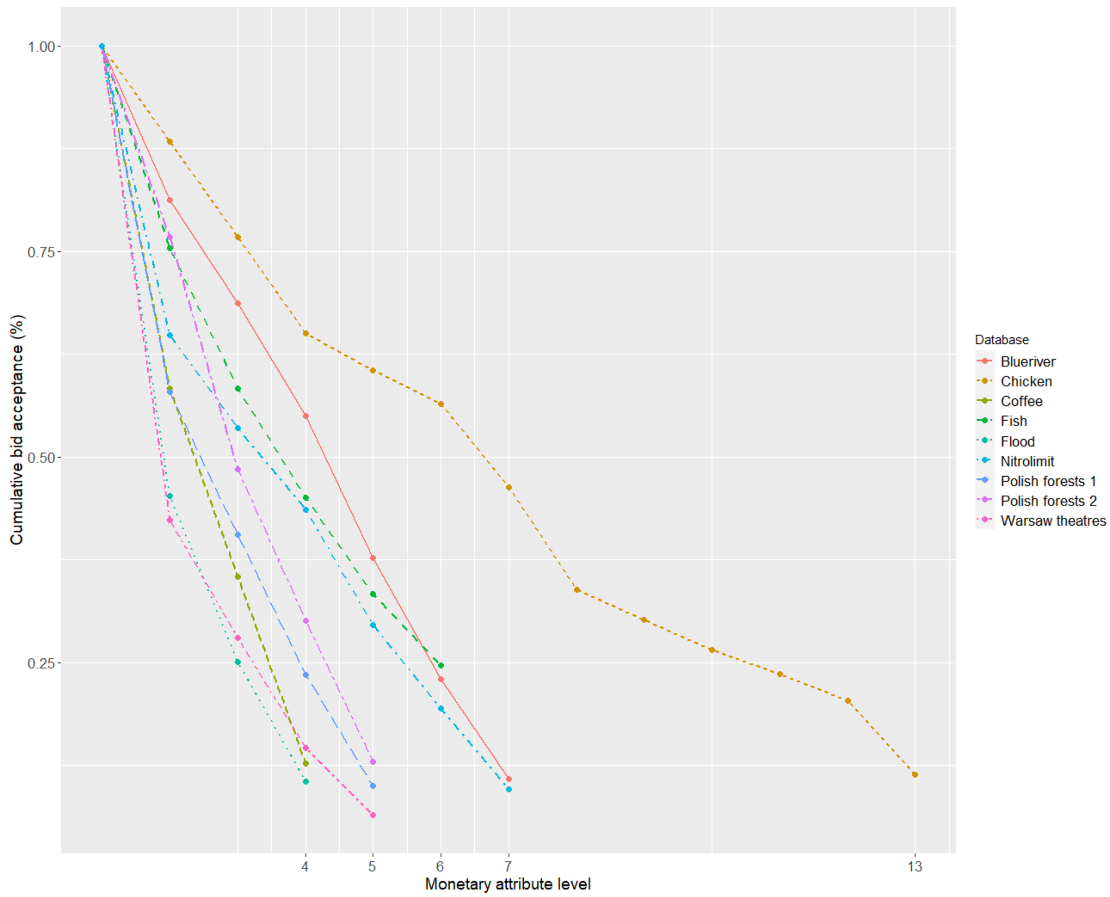


Figure 1: Bid acceptance curves

We find that most datasets yield 'well-behaved' bid acceptance curves. The highest monetary bid is selected between 6 and 12% of the times for all the datasets apart from the *Fish* dataset and the *Chicken* dataset. The most problematic results are found for the *Fish* dataset, where the alternative featuring the highest price bid (six levels in total) is selected 24.64% of the times. This is more than double the proportion found for the other datasets and clearly indicates a form of price insensitivity. The *Chicken* dataset is also problematic because the three highest price bids out of 13 are selected in nearly 25% of the cases, which is disproportionately higher than what is found in the other datasets. In the *Chicken* data case, the choices are about food, where the price can be considered as an indicator of quality (Palma et al., 2016). The *Fish* dataset is about endangered species conservation, which yields difficult choices because of the lack of comparison with existing markets. The econometric treatments to apply in such context are out of the scope of this paper. Despite the fact that they feature clear issues when it comes to price sensitivity, the *Fish* and the *Chicken* datasets are considered in the remainder of the paper in order to illustrate how the proposed modelling approach performs under different circumstances.

5 Results

5.1 Goodness-of-fit and welfare estimates

In this section, we discussed the results obtained for each model and each dataset in terms of goodness-of-fit and moments of the mWTP distributions for each dataset. We only report the mWTP for the most valued attribute (according to the MNL model) for each dataset in an attempt to simplify the presentation of the outputs of our analysis.

Table 3 shows that the models estimated in preference space (excepted the MNL models and the models for which random heterogeneity in preferences for the price is not considered) outperform the WTP space models in terms of goodness-of-fit for all the datasets considered. The largest gap is found when comparing Model E3 (*Fish LN*) and Model E5 (*Fish WTPS*). The log-likelihood for Model E3 is found to be -6740.244, while it is -8042.301 for Model E5, which is a difference of 1302.057 likelihood points. The BIC for model E3 is found to be 16.04% lower than the BIC for Model E5. The smallest difference is found between Model G5 (*Cars WTPS*, for which the log-likelihood is equal to -6326.96) and Model G9 (*Cars κ-shifted*, for which the log-likelihood is found to be -6310.088). The difference is 16.872 likelihood points while the BIC difference is -0.26%.

Overall, the best performing model across datasets in terms of goodness-of-fit is found to be the LN model as defined in Table 1, which is found to be the best fitting dataset in 7 cases out of 10. The 3-parameters shifted LN model is found to be the best model in 2 cases and the shifted model in one case. However, none of these models are adequate for deriving meaningful welfare estimates for policy making. Indeed, in 8 cases out of 10, the shift parameter κ for the 3-parameters shifted LN model is found to be positive, thus preventing WTP moments to be derived. The same result is found for the κ-shifted model in 3 cases, making it unsuitable for welfare analysis. Although it is possible to derive moments for the mWTP distributions from the LN models in all cases, we face issues related to 'exploding' implicit prices as previously discussed. The mWTP estimates derived from the LN models (and, to a smaller extend, from the LU model) are found to be much higher that the estimates derived from all the other models. As an illustration, the mean mWTP for the attribute *as* (*Fish* dataset) derived from Model E3 is found to be 1447135, while it is 1394.496 for Model E4 (LU), which is found to be 4426439.75% and 4165.52% higher respectively, than the value derived from Model E5 (32.69). Given than the maximum price bid in the *Fish* survey design is 48, we conclude that these really high mWTP measures are driven by distributional assumptions rather than genuinely high preferences. This becomes particularly plausible when also looking at the values derived from Model E1 (41.56) and Model E2 (41.77). This issue is also found for the *Normandy flood protection* dataset, the *Polish forests II* dataset, the *Chicken* dataset, the *Bluerivers* dataset and the *Nitrolimit* dataset.

We note that the difference in terms of goodness-of-fit between the LN models, the LU models and the shifted models is marginal in most cases. This is particularly relevant when it comes to the μ-shifted model, for which the difference with the LN model is systematically below 1% of the BIC. At the same time, the μ-shifted model mitigates the issues related to the "exploding ratio" in most cases (and always provides existing moments). As an illustration, the log-likelihood for Model A8 (*Normandy flood protec-*

tion) is found to be -1348.84, while it is -1338.63 for Model A3 and -1341.71 for Model A4 (which is a difference of 10.22 likelihood points maximum). On the other hand, the average mWTP for the attribute *infra* derived from Model A7 is found to be 27.57, while it is 184.74 for the LN model and 59.97 for the LU model. Similar patterns of results are found for most datasets, excepted for the *Fish* dataset, the *Chicken* dataset and the *Bluerivers* datasets, for which the mWTP difference between the μ-shifted and the WTPS model is found to be above 100% (the differences between the MNL specification and the μ-shifted specification are found to be smaller in most cases). However, we note that for the *Bluerivers* dataset, Model I1 (MNL) is found to yield mWTP estimates which are very close to Model I6 (121.75 *versus* 144.81). Moreover, we note that the WTPS models are largely outperformed by the μ-shifted models in terms of goodness-of-fit for both the *Fish* and *Chicken* datasets, which suggests the presence of cost insensitivity which is better captured by the μ-shifted model structure with respect to the WTPS specification.

The WTPS P2 models are found to be an improvement over the WTPSe models in all cases in terms of goodness-of-fit. Moreover, the mWTP measures are found to increase with respect to the corresponding WTPS model in 8 cases out of 10 with the highest increase being equal to 52.41% (Model I6) which corresponds to one of the outliers datasets for which the sensitivity for the price attribute has been found to be extremely low.

Overall, we note that the models which deliver the best goodness-of-fit are also those which yield the highest welfare estimates. It is also clear from Table 3 that the μ-shifted model is much more parsimonious than the LU and the LN models when it comes to their impact on welfare estimates. Indeed, welfare estimates derived from the the μ-shifted model structure are found to be between 9.98% (*Cars choice*) and 99.99% lower (*Fish* dataset) than the LN structure. In other words, the μ-shifted model yields much smaller and much more plausible mWTP (according to expert knowledge and competing models) than the LN and LU specifications while at the same time ensuring a similar goodness-of-fit (and better than the WTPS specification in all cases). In the next section, we further refine our analysis by providing a quantitative assessment of the performances of the μ-shifted model in terms of welfare estimates in comparison to the MNL, the MMNL without random price heterogeneity, the LN model, the LU model and the WTPS model. The κ-shifted specification and the 3-parameters specification are not considered in the remainder of this paper because of their inability to provide moments for the mWTP in all cases.

Table 3: Overview of results

Survey	Goodness-of-fit					WTP moments					
	Model	**LL**	**BIC**	**BIC vs LN (%)**	**BIC vs WTPS (%)**	**Attribute**	**Mean**	**Q50 (Median)**	**Q99**	**Mean vs LN (%)**	**Mean vs WTPS (%)**
	1. MNL	-2323.01	4692.95	69.35	50.92		26.66			-85.57	85.08
	2. Fixed monetary attribute	-1557.72	3201.49	15.53	2.96		18.80	18.81	90.71	-89.82	30.55
	3. LN	-1338.63	2771.12		-10.88		184.74	12.61	2937.03		1182.56
	4. LU	-1341.71	2777.29	0.22	-10.68		59.97	10.76	575.93	-67.54	316.30
A. Normandy flood protection	5. WTPS	-1501.94	3097.74	11.79		*infra*	16.4	16.42	75.50	-92.20	
	6. WTPS P2	-1489.51	3096.36	11.74	-0.04		20.68	13.90	107.98	-88.81	26.14
	7. 3-Param shifted	-1339.07	2779.83	0.31	-10.60						
	8. μ-shifted	-1348.84	2791.56	0.74	-10.22		27.57	15.39	130.55	-85.08	91.40
	9. κ-shifted	-1363.53	2820.93	1.80	-9.28		26.24	18.00	108.57	-85.80	82.18
	1. MNL	-1880.31	3814.81	22.33	17.70		0.13			-61.09	3.21
	2. Fixed monetary attribute	-1589.18	3279.01	5.15	1.17		0.14	0.14	0.43	-59.48	7.49
	3. LN	-1505.03	3118.45		-3.78		0.34	0.14	3.31		165.25
	4. LU	-1505.14	3118.66	0.01	-3.78		0.34	0.15	2.30	-1.30	161.79
B. Coffee machines	5. WTPS	-1568.26	3244.91	4.06		*cardpay*	0.13	0.13	0.44	-62.30	
	6. WTPS P2	-1524.56	3188.49	2.25	-1.74		0.16	0.13	0.68	-54.08	21.80
	7. 3-Param shifted	-1499.42	3114.97	-0.11	-3.89						
	8. μ-shifted	-1504.75	3117.9	-0.02	-3.80		0.21	0.15	0.90	-39.82	59.63
	9. κ-shifted	-1517.18	3142.76	0.78	-3.03		0.19	0.17	0.65	-46.05	43.10
	1. MNL	-14843.4	29753.8	37.86	29.90		44.97			-68.91	24.39
	2. Fixed monetary attribute	-11662.8	23450.09	8.65	2.38		29.29	29.18	188.02	-79.75	-18.99
	3. LN	-10724.7	21583.37		-5.77		144.66	16.28	2567.06		300.10
	4. LU	-10739.5	21613.05	0.14	-5.64		82.53	14.51	964.24	-42.95	128.26
C. Polish forests 1	5. WTPS	-11385.8	22905.67	6.13		*cen*	36.08	36.09	190.85	-75.01	
	6. WTPS P2	-11322	22816.29	5.71	-0.39		48.78	16.78	379.79	-66.28	35.22
	7. 3-Param shifted	-10721.2	21586.02	0.01	-5.76						
	8. μ-shifted	-10762.6	21659.27	0.35	-5.44		40.47	18.37	263.90	-72.02	11.94
	9. κ-shifted	-10756.1	21646.25	0.29	-5.50		50.54	20.88	345.42	-65.06	39.78
	1. MNL	-29708.3	59497.89	61.94	43.95		142.71			-90.13	53.14
	2. Fixed monetary attribute	-21120	42433.19	15.49	2.66		87.91	87.84	311.44	-93.92	-5.67
	3. LN	-18268.8	36740.96		-11.11		1446.46	81.63	22332.47		1452.07
	4. LU	-18350.8	36904.88	0.45	-10.71		294.32	39.82	2736.30	-79.65	215.81
D. Polish forests II	5. WTPS	-20564.8	41332.85	12.50		*tra2*	93.20	93.14	278.44	-93.56	
	6. WTPS P2	-20397.8	41060	11.76	-0.66		86.34	112.83	205.15	-94.03	-7.35
	7. 3-Param shifted	-18284.9	36783.32	0.12	-11.01		143.91	94.42	581.01	-90.05	54.42
	8. μ-shifted	-18260.6	36724.53	-0.04	-11.15		146.26	100.42	563.87	-89.89	56.94
	9. κ-shifted	-18250.5	36704.39	-0.10	-11.20		172.97	149.33	550.47	-88.04	85.60
	1. MNL	-12516.4	25107.92	84.20	54.65		41.56			-99.997	27.12
	2. Fixed monetary attribute	-8213.46	16567.89	21.55	2.05		41.77	41.53	438.72	-99.997	27.76
	3. LN	-6740.24	13630.86		-16.04		1.45E+06	1.85E+01	2.26E+06		4.43E+06
	4. LU	-6754.13	13658.63	0.20	-15.87		1394.50	4.88	27908.93	-99.904	4165.53
E. Fish	5. WTPS	-8042.3	16234.97	19.10		*as*	32.69	32.65	138.58	-99.998	
	6. WTPS P2	-7715.06	15627.48	14.65	-3.74		32.90	10.62	264.39	-99.998	0.64
	7. 3-Param shifted	-6728.52	13616.8	-0.10	-16.13						
	8. μ-shifted	-6749.24	13648.84	0.13	-15.93		106.44	24.15	688.05	-99.993	225.57
	9. κ-shifted	-6741.14	13632.64	0.01	-16.03						

Survey	Goodness-of-fit						WTP moments				
	Model	LL	BIC	BIC vs LN (%)	BIC vs WTPS (%)	Attribute	Mean	Q50 (Median)	Q99	Mean vs LN (%)	Mean vs WTPS (%)
F. Warsaw theaters	1. MNL	-11860.4	23779.83	41.13	37.67	*roz*	36.48			-40.55	7.44
	2. Fixed monetary attribute	-8725.55	17559.46	4.21	1.66		34.27	34.23	115.44	-44.15	0.93
	3. LN	-8365.79	16849.79		-2.45		61.36	23.58	612.05		80.71
	4. LU	-8378.8	16875.81	0.15	-2.30		49.04	25.80	299.43	-20.09	44.41
	5. WTPS	-8577.6	17273.41	2.51			33.96	33.92	110.82	-44.66	
	6. WTPS P2	-8512.87	17183.35	1.98	-0.52		34.45	21.00	170.24	-43.86	1.45
	7. 3-Param shifted	-8370.33	16868.72	0.11	-2.34						
	8. μ-shifted	-8418.69	16955.59	0.63	-1.84		39.33	26.45	177.00	-35.91	15.82
	9. κ-shifted	-8462.7	17043.61	1.15	-1.33		38.51	29.30	155.35	-37.24	13.41
G. Car choice	1. MNL	-6930.28	14021.01	9.01	8.06	*engine_hybrid*	7.74			-40.11	-12.76
	2. Fixed monetary attribute	-6427.59	13167.18	2.37			8.06	8.03	45.12	-37.66	-9.19
	3. LN	-6270.89	12862.69		-0.86		12.93	8.09	109.10		45.67
	4. LU	-6277.87	12876.65	0.11	-0.76		13.67	8.38	104.13	5.76	54.06
	5. WTPS	-6326.96	12974.83	0.87			8.87	8.84	44.94	-31.35	
	6. WTPS P2	-6295.4	13045.42	1.42	0.54		9.44	4.96	65.52	-26.96	6.40
	7. 3-Param shifted	-6271.43	12872.68	0.08	-0.79						
	8. μ-shifted	-6284.83	12890.56	0.22	-0.65		11.65	8.79	68.03	-9.89	31.26
	9. κ-shifted	-6310.09	12941.08	0.61	-0.26		10.85	8.77	60.62	-16.11	22.21
H. Chicken	1. MNL	-8899.49	17872.49	17.89	10.97	*irel*	2.75			-89.97	-1.77
	2. Fixed monetary attribute	-7988.05	19783.25	30.49	22.84		2.98	2.98	16.29	-89.14	6.31
	3. LN	-7506.66	15160.35		-5.87		27.44	3.22	517.66		879.14
	4. LU	-7524.14	15195.3	0.23	-5.65		204.43	5.34	3834.87	645.12	7195.81
	5. WTPS	-7979.06	16105.14	6.23			2.80	2.80	14.99	-89.79	
	6. WTPS P2	-7846.15	15885.27	4.78	-1.37		3.86	1.07	32.74	-85.92	37.87
	7. 3-Param shifted	-7503.98	15164.17	0.03	-5.84						
	8. μ-shifted	-7512	15171.02	0.07	-5.80		6.49	3.25	47.72	-76.34	131.68
	9. κ-shifted	-7533.73	15214.49	0.36	-5.53		6.32	4.23	37.47	-76.95	125.68
I. Bluerivers	1. MNL	-12243.8	24600.46	29.27	24.85	*baden2*	121.75			-91.91	88.92
	2. Fixed monetary attribute	-9783.48	19783.25	3.96	0.41		69.00	68.93	584.26	-95.42	7.07
	3. LN	-9402.13	19029.96		-3.42		1505.24	44.51	30049.69		2235.59
	4. LU	-9428.43	19082.55	0.28	-3.15		612.87	52.91	9193.19	-59.28	850.96
	5. WTPS	-9738.85	19703.39	3.54			64.45	64.39	532.35	-95.72	
	6. WTPS P2	-9403.04	19041.19	0.06	-3.36		98.22	6.83	1041.61	-93.47	52.41
	7. 3-Param shifted	-9712.65	19745.03	3.47	3.70						
	8. μ-shifted	-9421.08	19067.85	0.20	-3.23		144.81	52.54	1134.52	-90.38	124.70
	9. κ-shifted	-9404.4	19034.49	0.02	-3.39						
J. Nitrolimit	1. MNL	-9055.68	18247.99	62.43	57.26	*uhavel3*	98.65			-96.66	77.35
	2. Fixed monetary attribute	-5724.04	11712.24	4.25	0.93		55.67	55.67	112.77	-98.11	0.07
	3. LN	-5480.64	11234.55		-3.18		2.95E+03	6.13E+01	4.10E+04		5.21E+03
	4. LU	-5503.78	11280.83	0.41	-2.78		2070.59	8.29E+01	2.53E+04	-29.85	3.62E+03
	5. WTPS	-5665.35	11603.97	3.29			55.63	55.63	97.80	-98.12	
	6. WTPS P2	-5545.82	11492.43	2.30	-0.96		50.17	52.27	60.25	-98.30	-9.80
	7. 3-Param shifted	-5485.36	11253.1	0.17	-3.02		100.04	52.74	428.34	-96.61	79.84
	8. μ-shifted	-5482.32	11237.91	0.03	-3.15		90.97	62.37	345.95	-96.92	63.53
	9. κ-shifted	-5554.4	11382.06	1.31	-1.91						

5.2 Meta-analysis of welfare estimates

The meta-regression seeks to measure the effect of model specifications on welfare estimates all else being equal. For each dataset, 7 models are considered:

i MNL

ii MMNL without random heterogeneity in price sensitivity (simply labelled as "fixed price")

iii LN

iv LU

v WTPS

vi WTPS P2

vii μ-shifted

The dependent variable corresponds to the natural logarithm of the average mWTP for each attribute, each model and each dataset. For the *Coffee machines* dataset, the welfare measures have been rescaled and expressed in euro-cents in order to prevent the log-transformation from leading to a sign change. In the rare cases for which the average mWTP for a given attribute is found to be negative, we have calculated the negative of the natural logarithm of the absolute value of the average mWTP. We obtain a total of 483 observations (69 attributes times 7 models). The independent variables correspond to a set of dummy variables related to the dataset (and for which the base is the first dataset, *i.e. the Normandy flood protection* dataset) as well as a set of dummy variables which capture the effect of each model. Robust standard errors were calculated at the attribute cluster level and a random effect (normally distributed using 1,000 mlhs draws) is introduced to control for potential correlations for the same attribute. A second model is estimated without the two outlier datasets (the *Fish* dataset and the *Chicken meat* dataset) in an attempt to control for the effect of price insensitivity on welfare estimates as defined in Section 4.2. For both meta-regressions, a few observations have been removed. This only concerns some attributes for which the sign was inconsistent across models and the significance was poor overall. The first meta-regression features 470 observations while the second meta-regression features 407 observations. Model results are reported in Table 4 below.

Table 4: Meta-analysis of welfare estimates

Sample		**All observations**			**No outliers**		
n		470			407		
Variables		**Param.**	**Rob. T**		**Param.**	**Rob. T**	
Parametrisation variables	*MNL*	-0.266	-2.62	***	-0.179	-1.61	.
	Fixed price	-0.384	-9.86	***	-0.377	-9.43	***
	LN	2.187	7.29	***	1.609	8.91	***
	LU	1.389	8.80	***	1.120	7.08	***
	WTPS	-0.455	-7.31	***	-0.353	-7.05	***
	WTPS P2	-0.389	-5.55	***	-0.307	-4.62	***
	μ-shifted	base outcome			base outcome		
Data variables	*Coffee*	base outcome			base outcome		
	Nitrolimit	1.944	7.27	***	1.949	7.32	***
	Polish forests 1	2.239	9.80	***	2.241	9.80	***
	Polish forests 2	0.082	0.10	.	0.080	0.10	.
	Warsaw theaters	0.599	1.91	***	0.601	1.92	***
	Cars	-3.533	-5.65	***	-3.525	-5.62	***
	Chicken	-0.903	-3.73	***			
	Fish	2.994	12.34	***			
	Normandy floods	0.900	3.48	***	0.901	3.48	***
	Bluerivers	2.505	11.20	***	2.507	11.19	***
Misc	*Constant*	2.105	11.05	***	2.184	11.50	***
	Random effect	-1.126	-7.17	***	1.265	7.99	***

* Significant at the 10% level

** Significant at the 5% level

*** Significant at the 1% level

The meta-regression with *all observations*, for which the base outcome related to model parametrisation is the μ-shifted model, shows that the LN and LU models yield mean mWTP estimates which are significantly higher than those derived from the μ-shifted model, while the rest of the models (MNL, MMNL without random heterogeneity in the price attribute, WTPS model and WTPS P2 model) all yield mean welfare estimates which are lower than the μ-shifted model all else being equal. However, when the outlier datasets are excluded from the analysis (*no outliers*), the difference between the MNL model and the μ-shifted model becomes non-significant (even at the 10% level). This is an important result because this shows that, in the context of this paper and the data considered in this meta-regression, the μ-shifted model produces mean welfare estimates which are overall not significantly different than those derived from an MNL while at the same time fitting the data much better, in contrary to the LN and LU approaches which provide significantly higher welfare estimates. A better way to further comment on the differences across models in terms of welfare estimates is to look at the marginal effects, which are reported in Table 5 below.

Table 5 reports the marginal effects of the different model specifications together with confidence intervals (at the 1% level) for three different bases: μ-shifted model, WTPS model and LN model. Looking at the results featuring *all observations*, we find that the μ-shifted model successfully prevents the 'exploding' ratio problem because it provides welfare estimates which are 791% lower than those derived from the LN model and 301% lower than those derived from the LU model. The WTPS model is found to provide results which are not significantly different from the MNL model, the fixed price model and the WTPS P2 model. This is not surprising in the sense that the distributional assumptions are fairly similar across these models. The μ-shifted model provides results which are between 32% and 83% higher than the WTPS model although, as we will see next, this gap is largely reduced when outliers are excluded from the analysis. Finally,

Table 5: Meta-analysis of welfare estimates - Marginal effects

All observations									
	μ-shifted			**WTPS**			**LN**		
	Low.	*Mean*	*Upp.*	*Low.*	*Mean*	*Upp.*	*Low.*	*Mean*	*Upp.*
MNL	-43%	-23%	-3%	-9%	21%	50%	-99%	-91%	-84%
Fixed price	-39%	-32%	-25%	-6%	7%	21%	-99%	-92%	-86%
LN	102%	791%	1480%	47%	1304%	2561%			
LU	138%	301%	464%	222%	532%	842%	-83%	-55%	-27%
WTP	-47%	-37%	-26%				-99%	-93%	-87%
WTP P2	-44%	-32%	-20%	-7%	7%	21%	-99%	-92%	-86%
μ-shifted				32%	58%	83%	-97%	-89%	-80%
Without outliers									
	μ-shifted			**WTPS**			**LN**		
	Low.	*Mean*	*Upp.*	*Low.*	*Mean*	*Upp.*	*Low.*	*Mean*	*Upp.*
MNL	-40%	-16%	8%	-14%	19%	52%	-93%	-83%	-73%
Fixed price	-38%	-31%	-24%	-11%	-2%	6%	-94%	-86%	-79%
LN	167%	400%	632%	237%	611%	985%			
LU	82%	207%	332%	137%	337%	536%	-50%	-39%	-27%
WTP	-39%	-30%	-21%				-93%	-86%	-79%
WTP P2	-39%	-26%	-14%	-10%	5%	20%	-94%	-85%	-77%
μ-shifted				24%	42%	61%	-89%	-80%	-71%

taking the LN model as a base, we find that the μ-shifted model yield welfare estimates which are 89% lower while it is 91% lower for the MNL, 92% lower for the WTPS P2 model and 93% for the WTPS model. On the other hand, the LU model provides results which are only 55% lower than the LN model.

Excluding the *Chicken* and *Fish* datasets from the analysis further bridges the gap between the μ-shifted model and the other models (excluding the LN and LU models). Indeed, the difference between all the models (apart from the LU model) and the LN model in terms of welfare estimates is still very high and ranges between -80% (μ-shifted model) and -85% (WTPS model). The LU model provides welfare estimates which are only 39% lower than those derived from the LN model, even when "badly behaved" observations are removed from the analysis. As previously mentioned, the welfare estimates derived from the μ-shifted model are not longer significantly different than those derived from the MNL model. Removing outliers also reduces the difference between the μ-shifted model and the WTPS model from -37% to -30% (in other words, the welfare estimates derived from the WTP-space approach are now only 30% lower than those derived from the μ-shifted model once outliers have been removed). Another important result is that increasing the flexibility of the WTPS model by using mixture distributions shifts the mWTP values towards the μ-shifted model values. Indeed, the values derived from the WTPS P2 model are 7% higher than those derived from the WTPS model and only 26% lower than those derived from the μ-shifted model.

Altogether, the results of the meta-regressions indicate that the μ-shifted model is providing welfare estimates which are only slightly higher than those derived from the WTPS model and not significantly different than those derived from the MNL model while providing a much better fit. The LN model is found to perform only slightly better than the μ-shifted model in terms of goodness-of-fit while at the same time providing systematically untenable welfare estimates, even when the datasets with extreme price preferences are removed from the analysis. Adding some flexibility to the WTPS model by the means of mixture distributions leads to further bridge the gap between the WTPS model and the μ-shifted model although the μ-shifted model still fits the data better and is more parsimonious in parameters.

5.3 Comparison of distributions

Another way to compare the welfare estimates derived from the new μ-shifted specification to the other specifications analysed in this paper is to consider the whole distribution of the mWTP for each attribute instead of the mean. Such a test does not seek to find whether a given model specification yields higher (or lower) welfare estimates, but to inform whether the mWTP distributions derived from the μ-shifted specification are closer to the distributions derived from the WTP-space specification or the LN specification, for example. In this paper, this is achieved by using the test for measuring the common area of kernel density estimates described in Section 3.2.1.

The specifications considered are the following: fixed price, LN, LU, WTPS, WTPS P2 and μ-shifted. There are 69 non-monetary attributes across the 10 datasets considered. For each one of these attributes, 7 distributions are generated using 1,000 draws each time (higher values have been tested and didn't lead to inconsistencies for this test). For each distribution, we compute the common area of the kernel density estimates of two selected specifications at a time. For example, for the mWTP distributions of the attribute *infra* (*Normandy flood protection dataset*) illustrated in Figure 2, we compute the common area between:

i Fixed price and LN (0.29)

ii Fixed price and LU (0.63)

iii Fixed price and WTPS (0.92)

iv Fixed price and WTPS P2 (0.89)

v Fixed price and μ-shifted (0.86)

vi LN and LU (0.56)

vii LN and WTPS (0.29)

viii LN and WTPS P2 (0.32)

ix LN and μ-shifted (0.33)

x LU and WTPS (0.61)

xi LU and WTPS P2 (0.62)

xii LU and μ-shifted (0.66)

xiii WTPS and WTPS P2 (0.90)

xiv WTPS and μ-shifted (0.83)

xv WTPS P2 and μ-shifted (0.86)

[Figure 2 next page for better clarity]

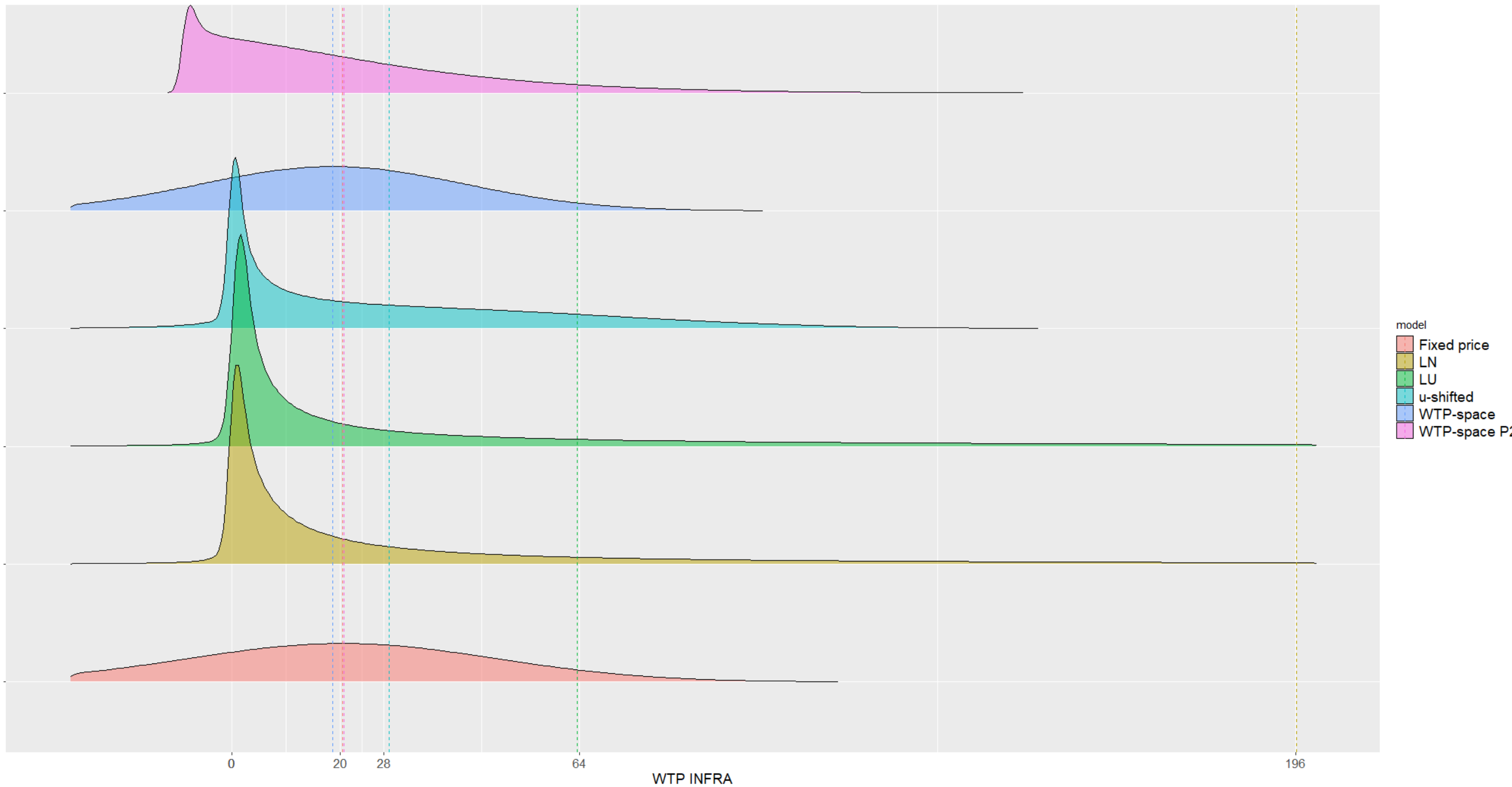


Figure 2: mWTP distributions for *the Normandy flood protection* dataset

The values previously listed as well as Figure 2 indicate for example that the distribution of the mWTP for the attribute *infra* derived from the LN model shares only 29% of its surface with the distribution derived from the WTP space model, while it shares 33% of its surface with the distribution derived from the μ-shifted model. The same measures are computed for the mWTP of the remaining 68 attributes. This leads to a total of 1035 common area measures, which is a sample large enough to allow to perform statistical tests. The mean (and median) common area for each pair of specification across all attributes is reported in Table 6. It is found that the two specifications which yield the most similar mWTP distributions are the fixed price specification and the WTP space specification (mean = 0.84, median = 0.92). This is unsurprising because the mWTP distributions are the same in both cases (in the context of this paper they are generally normally distributed. The parameters of the distributions might be different of course). The specifications which yield the distributions which have the least surface in common are the LN specification and the WTP space specification (mean = 0.40, median = 0.32).

Table 6: Common Area of kernel density estimates of mWTP distributions

		Median					
		Fixed price	***LN***	***LU***	***WTPS***	***WTPS P2***	***μ-shifted***
Mean	***Fixed price***	1	0.322	0.485	0.918	0.828	0.791
	LN	0.412	1	0.638	0.320	0.328	0.328
	LU	0.435	0.675	1	0.444	0.470	0.616
	WTPS	0.842	0.402	0.421	1	0.825	0.752
	WTPS P2	0.743	0.414	0.426	0.783	1	0.750
	μ-shifted	0.694	0.471	0.543	0.670	0.645	1

In the context of this paper, it is of particular interest to find whether the distributions derived from the μ-shifted specification are closer to those derived from the WTP space specification or the LN specification. The average common area between the mWTP distributions derived from the LN specification and the WTP space specification is found to be 0.40 and significantly lower than the average common area between the mWTP distributions derived from the μ-shifted specification and the WTP space specification (0.66) according to a T-test (t = -6.444, degrees of freedom = 136). This confirms that the proposed μ-shifted specification does not only yield mWTP moments which are closer to the WTP space parametrisation with respect to the LN specification, but that the overall mWTP distributions are found to be more similar.

The distributions derived from the WTP-space P2 specification have less area in common with the distributions from the μ-shifted approach than the distributions derived from the WTP-space approach (although this effect is very small and not significant according to a T-test). This is surprising because the WTP-space P2 specification is found to yield welfare estimates which reduce the gap with the μ-shifted specification with respect to the WTP-space approach. This might be due to the fact that although both the μ-shifted and WTP-space P2 specifications generally yield mWTP distributions with longer tails than the WTP-space approach, the distributions derived from from the WTP-space P2 approach exhibit a point-mass which is generally slightly more to the left than the μ-shifted approach, as illustrated by Figure 2. Introducing polynomials of higher order (see Fosgerau and Mabit (2013)) for the WTP-space specification might yield different results.

The results presented above should only be considered as valid in the context of this paper, where the overwhelming majority of the non-monetary attributes are specified as normally distributed as it is often the case in non-market valuation. Running such an analysis using different data sources (for example Value of Travel Time studies) might lead to different results because travel time is an attribute which is typically specified as negative log-normal.

5.4 Out-of-sample fit

The final test performed is an out-of sample fit test. This is a common method to test for over-fitting of a model (or, in the current case, compare whether different models are more or less prone to over-fitting). Each dataset is split into two parts: an estimation sample and a validation sample. 70% of the respondents are allocated to the estimation sample and the remaining 30% are allocated to the validation sample. This is repeated 10 times with different, randomly selected respondents each time. Different models of interest are produced using the estimation samples. For each estimation sample, the resulting parameters are used to compute the log-likelihood of the model on the estimation sample as well as on the corresponding validation sample. Repeating this process with different estimation and validation samples enhances the robustness of the test.

This protocol is slightly different from the protocol proposed by Train and Weeks (2005), where the sampled respondents are divided into two equal-sized sub-samples and the log-likelihood of the estimated models are evaluated on the other sub-sample. Moreover, our protocol is very different from the approach of Sonnier et al. (2007) who included all but one choice situation for each respondent in their estimation sample and computed the log-likelihood of the competing models on the single choice situation remaining for each respondent. For each estimation sample, we estimate three models:

i LN model

ii μ-shifted model

iii WTP space model

The analysis is restricted to these three specifications because the proposed protocol is computationally intensive and requires the estimation of 300 additional models (10 datasets times 10 training sets times 3 specifications). Each model is estimated using

2,000 mlhs draws. As previously stated, it is also clear at this stage of the analysis that, among the three distributions for the price parameter derived from the three-parameter log-normal distribution presented in this paper, the μ-shifted distribution is the only one that always ensures the existence of moments for the underlying mWTP distributions and doesn't suffer from identification issues and thus should be the one to receive increased attention. The LU distribution is also not considered for this test because, as seen in Table 3, it consistently results in an inferior goodness-of-fit with respect to the LN distribution while at the same time performing poorly when it comes to mitigating the "exploding ratio" problem. The WTPS P2 model is not considered because it is quite similar to the WTPS model and features more parameters than the other competing models. Finally, it is of little interest to compare the out-of-sample performances of the MNL model and the MMNL model with a fixed monetary attribute in the context of this paper. Results are reported in Table 7 below.

The LN model is found to exhibit the best average out-of-sample fit in 9 cases out of 10. The μ-shifted model is found to slightly outperform the LN model for the Nitrolimit dataset (-1644.70 *versus* -1649.05). The difference between the LN model and the μ-shifted model is found to be marginal in all cases and rarely exceeds 10 likelihood points. On the other hand, the average out-of-sample log-likelihood is between 1.41% (car choice dataset) and 18.98% (Fish dataset) lower for the WTP space specification compared to the LN specification (the difference is about the same for the μ-shifted specification). The average out-of-sample fit for the WTP-space specification across datasets is 7.41% lower than what is found for the μ-shifted specification (7.64% for the LN specification). However, the out-of-sample fit is not found to be disproportionally lower than the in-sample-fit for all the datasets and specifications considered. This means that the out-of-sample performances of the WTP space approach are not found to be lower than those of the μ-shifted and LN specifications.

Table 7: Out-of-sample fit

Survey	Model	In	Out
Normandy flood protection	*Log-normal*	-934.699755	-406.63
	μ-shifted	-941.660776	-410.16
	WTPS	-1084.43652	-480.85
University of Nantes coffee machines	*Log-normal*	-1048.42563	-456.53
	μ-shifted	-1052.22735	-457.84
	WTPS	-1091.40451	-477.41
Polish forests 1	*Log-normal*	-12863.3804	-5443.86
	μ-shifted	-12828.545	-5454.71
	WTPS	-14334.9336	-6219.88
Polish forests 2	*Log-normal*	-7505.71019	-3221.98
	μ-shifted	-7532.51409	-3231.43
	WTPS	-7960.65166	-3420.21
Fish	*Log-normal*	-4673.77172	-2072.58
	μ-shifted	-4685.04456	-2079.24
	WTPS	-5835.43869	-2558.33
Warsaw theaters	*Log-normal*	-5844.74319	-2531.42
	μ-shifted	-5879.24441	-2542.35
	WTPS	-5987.2186	-2591.02
Car choice	*Log-normal*	-4382.79432	-1895.41
	μ-shifted	-4395.1896	-1896.78
	WTPS	-4439.9235	-1922.60
Chicken	*Log-normal*	-5256.48631	-2257.45
	μ-shifted	-5261.34282	-2261.64
	WTPS	-5567.29331	-2388.85
Bluerivers	*Log-normal*	-6593.36604	-2816.68
	μ-shifted	-6606.03361	-2821.04
	WTPS	-6811.40965	-2929.93
Nitrolimit	*Log-normal*	-3851.67037	-1649.05
	μ-shifted	-3845.04323	-1644.70
	WTPS	-4058.28796	-1760.52

Altogether, these results suggest that there are only very small differences between the three competing specifications in terms of over-fitting. Again, the μ-shifted is not found to feature any specific disadvantage compared to the LN specification while at the same time providing much more reasonable welfare estimates.

6 Conclusion

Mixed multinomial logit models are the most widely used specification for fitting stated choice experiments data in non-market valuation, transport and health, among other fields (Mahieu et al., 2017). The past 20 years have seen a profusion of competing specifications, some of which aim at fitting the data better while other seek to provide more reasonable distributions of mWTP. However, in some cases, improving the fit of a given model leads to implausibly large welfare estimates (and mitigating this issues leads, in turn, to a decrease in the fit of the model). This trade-off, largely illustrated by the opposition between preference space and WTP space models, has led practitioners to suggest finding new modelling strategies for fitting the data better in WTP space and/or mitigating the so-called exploding ratio issue in preference space (Train and Weeks, 2005).

In this paper, we have investigated the usefulness of using shifted log-normal distributions (Sangal and Biswas, 1970) for the price attribute in mixed logit models. In addition to testing the traditional shifted negative log-normal distribution, labelled as the three-parameters log-normal and recently studied by McFadden and Robles (2019), we have introduced two new parametrisations: the κ-shifted distribution and the μ-shifted distribution. We have compared these new approaches to the more common modelling strategies found in the literature. These comparisons have been made following a framework for empirical testing which included a quantitative assessment of the differences across specifications in terms of goodness-of-fit performed by the means of a meta-analysis where 10 datasets and 483 distributions of welfare estimates were considered.

While the three-parameters log-normal and the κ-shifted distributions have been found to be problematic due to the fact that they cannot necessarily ensure the existence of moments for the implicit prices/mWTP distributions, the μ-shifted distribution for the price attribute has been found to systematically deliver welfare estimates which are not significantly different from those derived from the multinomial logit model (providing that the data is well-behaved). Moreover, the μ-shifted approach delivers welfare estimates which are 80% lower than those derived from a model with a (negative) log-normally distributed price parameter, while it is 86 % lower for the WTP space approach. At the same time, the μ-shifted approach fits the data almost as well as the negative log-normal approach (the models are virtually equivalent in most cases, and the μ-shifted approach is found to be a marginal improvement over the log-normal approach in 2 cases out of 10). In comparison, the WTP space approach is outperformed in terms of goodness-of-fit (sometimes largely) in 10 cases out of 10. Improving the flexibility of the WTP space model by introducing mixtures of normal (second order polynomials) marginally reduces the gap between the μ-shifted and the WTP space approach in terms of goodness-of-fit and welfare measures but is costly in terms of parameters.

Despite the large amount of evidence provided in this paper, it is possible that our results are context specific. More precisely, most of the non-monetary attributes featured in the 10 datasets considered have been specified as normally distributed. Different results could have been found if more papers from the transport literature would have been considered since typical attributes such as travel time are often specified as negative log-normal. The ratio of a log-normal distribution and a μ-shifted distribution might yield different results than the ratio of a normal distribution and a μ-shifted distribution. Future research on this topic will hence investigate the usefulness of the proposed method in different empirical contexts. Another area for future research is to focus on less models but compare the μ-shifted approach to models featuring flexible distributions with polynomials of higher order in order to investigate whether highly flexible distributions are further converging towards the results derived from the μ-shifted model.

References

Bartczak, A. (2015). The role of social and environmental attitudes in non-market valuation: An application to the białowieża forest. *Forest Policy and Economics*, 50:357–365.

Cameron, T. A. (1988). A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression. *Journal of environmental economics and management*, 15(3):355–379.

Campbell, D. (2008). Identification and analysis of discontinuous preferences in discrete choice experiments. In *European Association of Environmental and Resource Economists Annual Conference Gothenburg, Sweden*, pages 25–28.

Campbell, D. and Doherty, E. (2013). Combining discrete and continuous mixing distributions to identify niche markets for food. *European Review of Agricultural Economics*, 40(2):287–312.

Crastes, R., Beaumais, O., Arkoun, O., Laroutis, D., Mahieu, P.-A., Rulleau, B., Hassani-Taibi, S., Barbu, V. S., and Gaillard, D. (2014). Erosive runoff events in the european union: Using discrete choice experiment to assess the benefits of integrated management policies when preferences are heterogeneous. *Ecological economics*, 102:105–112.

Czajkowski, M., Giergiczny, M., and Greene, W. H. (2014). Learning and fatigue effects revisited: Investigating the effects of accounting for unobservable preference and scale heterogeneity. *Land Economics*, 90(2):324–351.

Czajkowski, M., Vossler, C. A., Budziński, W., Wiśniewska, A., and Zawojska, E. (2017). Addressing empirical challenges related to the incentive compatibility of stated preferences methods. *Journal of Economic Behavior & Organization*, 142:47–63.

Daly, A., Hess, S., and Train, K. (2012). Assuring finite moments for willingness to pay in random coefficient models. *Transportation*, 39(1):19–31.

Fiebig, D. G., Keane, M. P., Louviere, J., and Wasi, N. (2010). The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science*, 29(3):393–421.

Fosgerau, M. and Mabit, S. L. (2013). Easy and flexible mixture distributions. *Economics Letters*, 120(2):206–210.

Giergiczny, M., Valasiuk, S., Czajkowski, M., De Salvo, M., and Signorello, G. (2012). Including cost income ratio into utility function as a way of dealing with explodingimplicit prices in mixed logit models. *Journal of Forest Economics*, 18(4):370–380.

Glenk, K., Meyerhoff, J., Akaichi, F., and Martin-Ortega, J. (2019). Revisiting cost vector effects in discrete choice experiments. *Resource and Energy Economics*, 57:135–155.

Guy, G. and Willis, K. G. (1999). Economic valuation of the environment: Methods and case studies.

Hess, S., Daly, A., Dekker, T., Cabral, M. O., and Batley, R. (2017). A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research. *Transportation Research Part B: Methodological*, 96:126–149.

Hess, S. and Palma, D. (2019). Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of choice modelling*, 32:100170.

Hess, S., Train, K. E., and Polak, J. W. (2006). On the use of a modified latin hypercube sampling (mlhs) method in the estimation of a mixed logit model for vehicle choice. *Transportation Research Part B: Methodological*, 40(2):147–163.

Horbat, A. (2017). Der wert naturnaher flsse und auen in deutschland.

Hoyos, D. (2010). The state of the art of environmental valuation with discrete choice experiments. *Ecological economics*, 69(8):1595–1603.

Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W. M., Hanley, N., Ryan, M., Scarpa, R., et al. (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, 4(2):319–405.

Mahieu, P.-A., Andersson, H., Beaumais, O., dit Sourd, R. C., Hess, S., and Wolff, F.-C. (2017). Stated preferences: a unique database composed of 1657 recent published articles in journals related to agriculture, environment, or health. *Review of Agricultural, Food and Environmental Studies*, 98(3):201–220.

Martínez-Camblor, P., De Una-Alvarez, J., and Corral, N. (2008). k-sample test based on the common area of kernel density estimators. *Journal of Statistical Planning and Inference*, 138(12):4006–4020.

McFadden, D. and Robles, P. (2019). Instability in mixed logit demand models for analysis of concentrated markets. *2019 California Econometrics Conference.*

Meyerhoff, J., Boeri, M., and Hartje, V. (2014). The value of water quality improvements in the region berlin–brandenburg as a function of distance and state residency. *Water Resources and Economics*, 5:49–66.

Mørkbak, M. R., Christensen, T., and Gyrd-Hansen, D. (2010). Choke price bias in choice experiments. *Environmental and resource economics*, 45(4):537–551.

Palma, D., de Dios Ortúzar, J., Rizzi, L. I., Guevara, C. A., Casaubon, G., and Ma, H. (2016). Modelling choice when price is a cue for quality: a case study with chinese consumers. *Journal of choice modelling*, 19:24–39.

Sandorf, E. D., dit Sourd, R. C., and Mahieu, P.-A. (2018). The effect of attribute-alternative matrix displays on preferences and processing strategies. *Journal of choice modelling*, 29:113–132.

Sangal, B. and Biswas, A. K. (1970). The 3-parameter lognormal distribution and its applications in hydrology. *Water Resources Research*, 6(2):505–515.

Scarpa, R., Franceschinis, C., and Thiene, M. (2020). Logit mixed logit under asymmetry and multimodality of wtp: A monte carlo evaluation. *American Journal of Agricultural Economics*.

Scarpa, R., Franceschinis, C., and Thiene, M. (2021). Logit mixed logit under asymmetry and multimodality of wtp: a monte carlo evaluation. *American Journal of Agricultural Economics*, 103(2):643–662.

Scarpa, R., Thiene, M., and Train, K. (2008). Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the alps. *American Journal of Agricultural Economics*, 90(4):994–1010.

Sonnier, G., Ainslie, A., and Otter, T. (2007). Heterogeneity distributions of willingness-to-pay in choice models. *Quantitative Marketing and Economics*, 5(3):313–331.

Svenningsen, L. S. and Jacobsen, J. B. (2018). Testing the effect of changes in elicitation format, payment vehicle and bid range on the hypothetical bias for moral goods. *Journal of choice modelling*, 29:17–32.

Train, K. (2016). Mixed logit with a flexible mixing distribution. *Journal of choice modelling*, 19:40–53.

Train, K. and Weeks, M. (2005). Discrete choice models in preference space and willingness-to-pay space. In *Applications of simulation methods in environmental and resource economics*, pages 1–16. Springer.

7 Appendix

7.1 Detailed model results

Table 8: Normandy flood protection model results

		1.MNL		**2. Fixed mon. attr.**		**3. LN**		**4. LU**		**5. WTPS**		**6. WTPS P2**		**7. 3-param. Shifted**		**8. μ-shifted**		**9. κ-shifted**	
LL		-2323.007		-1557.723		-1338.625		-1341.708		-1501.937		-1489.513		-1339.067		-1348.843		-1363.532	
AIC		4658.01		3137.45		2701.25		2707.42		3027.87		3009.03		2704.13		2721.69		2751.06	
BIC		4692.95		3201.49		2771.12		2777.29		3097.74		3096.36		2779.83		2791.56		2820.93	
R2		0.1507		0.4279		0.5075		0.5063		0.4479		0.4513		0.5069		0.5037		0.4984	
		Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.
μ	ASC1	-0.491	-3.69	-0.256	-0.88	2.401	7.27	2.394	7.34	0.025	7.67	0.022	23.38	2.410	7.13	2.090	7.40	1.797	6.72
	ASC2	-0.659	-4.96	-0.695	-2.49	2.000	6.11	1.995	6.12	0.001	0.36	0.004	2.75	2.004	5.97	1.693	5.89	1.412	5.04
	Agri	1.021	13.88	2.383	9.73	2.295	9.64	2.247	10.76	0.144	63.15	0.149	117.00	2.290	9.67	2.239	10.60	2.126	10.85
	Infra	1.091	16.26	2.580	11.09	2.434	10.66	2.385	12.28	0.164	47.20	0.139	96.50	2.434	10.92	2.350	12.05	2.239	12.39
	Com	0.536	6.69	1.361	6.35	1.347	7.10	1.336	7.64	0.071	28.37	0.064	198.02	1.379	7.13	1.323	7.57	1.256	7.56
	Price	-4.093	-11.58	-12.668	-9.58	2.544	19.18	5.647	25.03	4.722	12.10	4.971	10.59	2.566	12.40	1.400	10.15	.	.
$\sigma1$	ASC1	.	.	1.196	5.14	-0.172	-0.75	-0.138	-0.78	-0.084	-44.86	-0.099	-514.54	-0.320	-1.15	0.098	0.59	0.072	0.59
	ASC2	.	.	-0.600	-1.25	-0.262	-0.37	-0.251	-0.78	-0.047	-31.19	-0.062	-314.38	0.175	0.61	0.262	0.77	0.129	0.39
	Agri	.	.	3.920	11.45	-1.670	-4.97	1.819	6.09	0.267	61.83	0.268	265.07	1.775	4.92	1.769	5.74	1.648	5.40
	Infra	.	.	3.598	11.63	1.395	5.49	1.349	6.82	0.255	59.98	0.247	320.27	1.412	5.77	1.417	6.54	1.372	6.74
	Com	.	.	-2.600	-9.02	-1.316	-5.14	-0.977	-3.12	0.161	55.18	0.162	213.84	1.149	3.48	-0.988	-2.57	-0.932	-2.55
	Price	.	.	.	.	2.161	11.75	-6.144	-12.94	2.189	6.64	2.360	7.93	2.140	8.31	3.335	13.81	5.288	12.43
$\sigma2$	Agri	.	.	.	.	.	.	.	.	.	.	0.039	86.37	.	.	.	.	.	.
	Infra	.	.	.	.	.	.	.	.	.	.	0.068	206.66	.	.	.	.	.	.
	Com	.	.	.	.	.	.	.	.	.	.	0.017	98.26	.	.	.	.	.	.
κ		.	.	.	.	.	.	.	.	.	.	.	.	0.210	0.17	.	.	-4.958	-5.00

Table 9: University of Nantes coffee machines model results

		1.MNL		2. Fixed mon. attr.		3. LN		4. LU		5. WTPS		6. WTPS P2		7. 3-param. Shifted		8. μ-shifted		9. κ-shifted	
LL		-1880.305		-1589.18		-1505.03		-1505.135		-1566.306		-1524.561		-1499.415		-1504.753		-1517.183	
AIC		3774.61		3204.36		3118.45		3118.66		3241.01		3085.12		3028.83		3037.51		3062.37	
BIC		3814.81		3279.01		3038.06		3038.27		3160.61		3188.49		3114.97		3117.9		3142.76	
R2		0.2544		0.367		0.3999		0.3998		0.3757		0.3906		0.4017		0.4		0.3951	
		Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.
μ	ASC1	-0.007	-0.05	-0.264	-1.24	0.250	1.11	0.176	0.81	-0.010	-1.29	0.00	-0.96	0.257	1.10	0.222	1.01	0.148	0.68
	ASC2	-0.137	-1.12	-0.281	-1.59	0.252	1.36	0.215	1.17	-0.008	-1.12	0.00	-6.21	0.261	1.38	0.232	1.25	0.142	0.78
	cardpay	1.372	13.87	3.076	10.13	3.536	8.1	3.754	9.94	0.130	17.70	0.13	50.82	3.636	9.73	3.453	10.11	3.194	10.03
	organic	0.669	9.41	1.482	8.83	1.530	7.38	1.571	8.37	0.065	6.52	0.04	16.62	1.556	8.02	1.515	8.52	1.445	8.84
	fairtrade	0.398	4.52	1.254	5.94	1.373	4.55	1.561	5.15	0.060	6.35	0.07	22.57	1.430	5.08	1.385	5.25	1.230	4.99
	recycle	1.068	14.99	2.000	10.77	2.110	9.33	2.200	10.75	0.081	12.00	0.06	14.92	2.184	10.47	2.089	10.33	1.951	10.64
	Price	-0.102	-11.49	-22.001	-9.27	2.948	20.53	1.078	3.56	3.733	13.91	4.19	24.43	3.410	10.06	2.119	15.58	.	.
$\sigma1$	ASC1	.	.	0.750	3.29	0.835	3.36	0.872	3.83	-0.018	-2.61	0.04	30.63	0.838	3.11	0.829	3.11	0.663	2.13
	ASC2	.	.	-0.660	-3.57	0.474	1.32	0.507	1.63	0.031	6.90	0.01	18.25	-0.581	-2.13	0.494	1.64	0.570	2.46
	cardpay	.	.	2.742	9.3	3.069	6.99	3.213	8.95	0.132	13.87	0.16	25.16	3.125	8.91	2.816	8.48	2.635	8.70
	organic	.	.	1.596	7.73	1.616	7.1	1.620	8.13	0.070	6.88	0.07	19.35	1.663	7.78	1.567	7.12	1.448	7.30
	fairtrade	.	.	-1.183	-3.87	1.936	4.76	1.908	7.63	0.043	5.39	0.03	13.00	2.007	5.91	1.885	5.53	1.671	5.53
	recycle	.	.	1.934	10.08	-1.892	-7.25	1.892	8.94	0.085	10.79	0.11	31.68	1.906	7.94	1.807	8.56	-1.798	-7.82
	Price	.	.	.	.	-1.113	-8.98	3.661	9.36	-1.072	-3.90	-1.61	-9.67	-0.815	-3.94	1.700	8.37	3.035	18.47
$\sigma2$	cardpay	.	.	.	.	.	.	.	.	.	.	0.03	17.22	.	.	.	.	.	.
	organic	.	.	.	.	.	.	.	.	.	.	0.04	14.57	.	.	.	.	.	.
	fairtrade	.	.	.	.	.	.	.	.	.	.	-0.02	-14.63	.	.	.	.	.	.
	recycle	.	.	.	.	.	.	.	.	.	.	0.04	29.07	.	.	.	.	.	.
	κ	.	.	.	.	.	.	.	.	.	.	.	.	9.994	1.13	.	.	-13.197	-6.16

Table 10: Bialowieza forest model results

		1.MNL		**2. Fixed mon. attr.**		**3. LN**		**4. LU**		**5. WTPS**		**6. WTPS P2**		**7. 3-param. Shifted**		**8. μ-shifted**		**9. κ-shifted**	
LL		-14843.38		-11662.8		-10724.65		-10739.49		-11385.8		-11321.96		-10721.18		-10762.6		-10756.09	
AIC		29700.76		23351.59		21477.29		21506.97		22799.6		22679.91		21472.37		21553.2		0.3203	
BIC		29753.8		23450.09		21583.37		21613.05		22905.67		22816.29		21586.02		21659.27		21540.17	
R2		0.0629		0.2632		0.3223		0.3214		0.2806		0.2844		0.3225		0.3208		21646.25	
		Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.
μ	ASC1	-0.222	-3.58	0.252	2.67	1.175	14.31	1.151	13.93	0.091	2.67	0.06	2.33	1.165	13.61	1.187	14.20	1.118	13.27
	ASC2	-0.316	-5.15	0.083	0.88	1.037	12.86	1.013	12.48	0.005	0.12	-0.01	-0.41	1.026	12.25	1.043	12.71	0.977	11.87
	gos	0.568	17.95	0.728	12.56	0.790	16.09	0.774	16.12	0.247	6.34	0.25	10.66	0.787	15.57	0.822	16.14	0.803	15.74
	cen	0.526	17.18	0.555	10.07	0.674	14.59	0.670	14.67	0.361	8.78	0.25	7.97	0.671	14.66	0.700	14.82	0.683	14.54
	vis1	0.037	0.97	-0.078	-1.46	-0.009	-0.20	-0.013	-0.29	-0.020	-0.95	-0.26	-6.20	-0.013	-0.28	-0.007	-0.14	-0.006	-0.14
	vis2	0.156	4.30	0.078	1.28	0.157	3.24	0.148	3.08	0.043	1.81	-0.17	-3.81	0.147	3.00	0.176	3.57	0.162	3.33
	fee	-1.262	-18.17	-2.489	-19.77	0.982	14.24	-1.766	-10.88	-0.715	-5.73	1.14	13.91	1.117	10.44	0.011	0.16	.	.
$\sigma 1$	ASC1	.	.	1.292	8.22	0.217	2.47	-0.199	-2.38	0.287	7.24	0.22	9.47	-0.240	-3.36	-0.250	-3.21	0.235	3.13
	ASC2	.	.	1.241	7.29	-0.016	-0.40	-0.057	-1.60	-0.285	-5.60	-0.22	-14.62	-0.056	-1.43	-0.058	-0.74	-0.001	-0.02
	gos	.	.	1.698	18.57	-1.045	-18.94	1.035	19.76	0.672	14.56	0.67	12.12	-1.053	-18.79	-1.087	-19.40	1.064	19.03
	cen	.	.	-1.671	-20.58	-1.071	-20.47	1.064	20.23	0.666	17.98	0.63	15.20	-1.078	-20.21	1.105	20.54	-1.079	-20.20
	vis1	.	.	1.044	9.37	0.698	10.56	0.675	10.33	0.372	7.13	0.08	2.09	0.693	10.37	-0.757	-11.65	0.711	10.73
	vis2	.	.	1.304	13.98	0.791	11.70	0.790	12.00	0.544	12.74	0.34	5.68	0.790	11.52	0.862	12.95	-0.835	-12.11
	fee	.	.	.	.	-1.782	-23.12	5.476	27.36	3.678	16.90	1.34	15.52	1.633	13.03	2.796	20.34	2.668	23.96
$\sigma 2$	gos	.	.	.	.	.	.	.	.	.	.	0.07	2.15	.	.	.	.	.	.
	cen	.	.	.	.	.	.	.	.	.	.	0.16	5.27	.	.	.	.	.	.
	vis1	.	.	.	.	.	.	.	.	.	.	0.30	6.34	.	.	.	.	.	.
	vis2	.	.	.	.	.	.	.	.	.	.	0.27	8.49	.	.	.	.	.	.
	κ	.	.	.	.	.	.	.	.	.	.	.	.	0.244	1.01	.	.	-0.733	-5.79

Table 11: Ecological value of Polish forests model results

		1.MNL		2. Fixed mon. attr.		3. LN		4. LU		5. WTPS		6. WTPS P2		7. 3-param. Shifted		8. μ-shifted		9. κ-shifted	
LL		-29708.28		-21120.01		-18268.81		-18350.77		-20564.76		-20397.83		-18284.91		-18260.6		-18250.53	
AIC		59432.55		42278.02		36577.62		36741.54		41169.51		40847.66		36611.82		36561.2		36541.06	
BIC		59497.89		42433.19		36740.96		36904.88		41332.85		41060		36783.32		36724.53		36704.39	
R2		0.1764		0.4141		0.4931		0.4908		0.4295		0.4339		0.4926		0.4933		0.4936	
		Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.
μ	asc1	2.003	23.50	-1.378	-16.02	0.267	1.98	0.235	1.52	-0.456	-7.84	-0.46	-11.95	0.170	1.10	0.206	1.38	0.218	1.74
	asc2	2.003	23.50	-1.270	-14.55	0.350	2.51	0.309	2.04	-0.433	-8.71	-0.44	-11.78	0.287	1.86	0.306	2.08	0.309	2.43
	asc3	2.003	23.50	-1.484	-16.87	0.188	1.35	0.153	1.03	-0.484	-9.38	-0.49	-12.46	0.109	0.69	0.137	0.89	0.132	1.02
	nat1	0.798	25.60	1.063	16.32	1.212	24.33	1.222	22.39	0.371	10.82	0.27	7.73	1.242	22.56	1.231	21.65	1.248	23.55
	nat2	1.174	27.68	1.694	14.86	1.866	24.81	1.836	20.91	0.573	8.83	0.49	7.64	1.843	24.06	1.914	26.93	1.861	24.76
	tra1	1.434	38.24	1.788	19.67	1.878	34.75	1.832	34.20	0.605	12.18	0.70	12.41	1.891	34.46	1.878	31.94	1.869	35.55
	tra2	1.919	40.12	2.613	18.86	2.688	35.79	2.791	32.43	0.932	7.62	1.13	14.01	2.732	32.55	2.741	35.02	2.809	34.81
	inf1	0.653	24.63	0.947	21.70	0.822	21.29	0.811	20.20	0.295	14.70	0.14	8.18	0.829	20.82	0.826	20.96	0.835	21.27
	inf2	1.052	35.12	1.289	22.85	1.402	30.36	1.393	31.33	0.439	15.64	0.25	9.33	1.374	30.57	1.387	28.01	1.393	30.94
	fee	-1.345	-21.06	-2.973	-26.65	0.973	11.63	-2.084	-3.87	1.341	19.10	1.32	23.27	0.132	0.64	-0.064	-0.54	.	.
$\sigma1$	asc1	.	.	0.697	8.06	0.436	6.10	-0.514	-7.28	0.076	0.56	0.03	0.61	0.529	8.40	0.500	9.64	0.449	8.80
	asc2	.	.	-0.513	-7.52	-0.295	-5.32	-0.305	-4.42	0.046	1.59	0.03	0.88	-0.063	-0.27	-0.109	-0.48	-0.244	-3.41
	asc3	.	.	0.662	9.21	0.482	9.06	-0.501	-9.44	0.128	3.73	0.09	3.73	-0.501	-9.21	-0.508	-9.19	0.488	10.70
	nat1	.	.	1.585	15.99	0.806	11.55	0.772	5.62	0.539	17.20	0.45	12.32	-0.883	-8.96	0.875	10.49	0.890	10.87
	nat2	.	.	2.398	19.46	1.373	13.95	1.303	10.55	0.763	12.19	0.67	12.43	1.467	13.20	1.471	15.49	1.487	13.02
	tra1	.	.	1.887	18.33	0.839	11.32	0.881	10.02	0.589	12.02	0.86	16.81	0.899	10.39	-0.843	-5.76	-0.844	-11.18
	tra2	.	.	2.852	25.89	1.445	15.74	1.424	17.09	0.795	14.78	0.99	14.16	1.542	15.61	1.494	7.16	1.412	15.96
	inf1	.	.	0.656	8.46	-0.335	-3.25	0.200	0.76	0.136	2.66	0.04	1.01	-0.348	-4.58	-0.386	-5.37	-0.400	-6.48
	inf2	.	.	-1.079	-17.40	-0.789	-17.09	-0.774	-16.81	-0.311	-8.94	-0.28	-6.33	0.766	16.45	-0.787	-11.57	-0.790	-17.27
	fee	.	.	.	.	2.313	26.70	7.624	12.87	-0.711	-12.31	-0.80	-24.24	2.915	11.00	3.340	30.24	3.209	48.06
$\sigma2$	nat1	.	.	.	.	.	.	.	.	.	.	0.20	6.81	.	.	.	.	.	.
	nat2	.	.	.	.	.	.	.	.	.	.	0.21	12.81	.	.	.	.	.	.
	tra1	.	.	.	.	.	.	.	.	.	.	-0.28	-15.29	.	.	.	.	.	.
	tra2	.	.	.	.	.	.	.	.	.	.	-0.26	-3.20	.	.	.	.	.	.
	inf1	.	.	.	.	.	.	.	.	.	.	0.16	8.32	.	.	.	.	.	.
	inf2	.	.	.	.	.	.	.	.	.	.	0.23	9.28	.	.	.	.	.	.
κ		.	.	.	.	.	.	.	.	.	.	.	.	-0.896	-4.87	.	.	-0.981	-7.74

Table 12: Endangered Fish model results

		1.MNL		2. Fixed mon. attr.		3. LN		4. LU		5. WTPS		6. WTPS P2		7. 3-param. Shifted		8. μ-shifted		9. κ-shifted	
LL		-12516.37		-8213.458		-6740.244		-6754.13		-8042.301		-7715.059		-6728.515		-6749.237		-6741.136	
AIC		25048.74		16456.92		13512.49		13540.26		16116.6		15472.12		13491.03		13530.47		13514.27	
BIC		25107.92		16567.89		13630.86		13658.63		16234.97		15627.48		13616.8		13648.84		13632.64	
R2		0.055		0.341		0.4902		0.4892		0.392		0.4163		0.491		0.4896		0.4902	
		Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.
μ	asc1	-0.197	-2.16	0.467	3.26	2.160	13.31	2.155	13.20	0.105	2.86	0.10	7.55	2.160	13.36	2.132	13.58	2.176	13.70
	asc2	-0.243	-2.69	0.372	2.52	2.032	12.78	2.053	12.93	0.103	2.59	0.10	7.41	2.056	13.01	2.025	13.27	2.067	13.30
	ac	0.265	10.07	0.455	7.73	0.520	10.62	0.518	10.09	0.098	7.14	-0.02	-2.76	0.498	10.08	0.518	10.60	0.515	9.98
	as	0.572	15.73	1.040	10.71	1.111	14.26	1.089	13.64	0.327	10.34	-0.07	-5.72	1.068	13.44	1.101	14.51	1.070	13.99
	f	0.260	10.77	0.483	9.37	0.481	10.89	0.490	11.05	0.112	6.72	-0.03	-2.97	0.477	10.84	0.500	11.47	0.474	10.72
	g	0.322	11.63	0.615	10.00	0.646	13.12	0.636	12.98	0.105	7.68	-0.04	-5.11	0.629	12.86	0.641	13.10	0.629	12.89
	S	0.457	12.16	0.803	9.95	0.916	11.69	0.926	11.93	0.170	3.08	-0.06	-6.58	0.929	11.84	0.961	12.36	0.925	11.53
	cost	-1.376	-9.20	-4.251	-12.65	0.419	1.21	-5.122	-7.22	1.947	15.79	2.18	22.43	0.485	2.93	-0.659	-2.97	.	.
$\sigma 1$	asc1	.	.	1.695	5.65	0.516	5.23	0.555	5.10	0.309	11.29	0.32	20.31	-0.573	-6.74	-0.559	-6.73	-0.547	-6.40
	asc2	.	.	1.785	5.30	0.447	3.79	0.476	6.06	0.299	9.80	0.33	23.62	0.402	3.82	0.402	3.95	0.458	5.04
	ac	.	.	1.277	11.50	0.760	11.65	0.756	12.67	0.246	4.39	-0.20	-11.19	0.735	11.78	0.743	12.40	-0.752	-12.32
	as	.	.	2.342	14.05	1.281	17.46	1.337	15.50	0.454	7.19	-0.10	-3.66	1.288	16.87	1.304	16.77	1.273	17.04
	f	.	.	1.012	10.08	-0.529	-7.92	0.556	8.16	0.210	4.57	0.11	11.79	0.549	8.19	0.538	8.64	0.541	8.61
	g	.	.	1.164	12.33	0.627	10.69	0.667	11.37	0.242	5.97	0.10	8.27	0.656	11.09	0.662	11.46	-0.604	-10.84
	S	.	.	2.191	12.63	1.378	15.73	1.337	17.21	0.401	5.00	-0.06	-5.00	1.350	16.66	1.372	16.84	1.336	16.00
	cost	.	.	.	.	4.518	15.13	13.113	13.17	-0.890	-12.48	1.36	16.86	4.219	19.45	5.628	18.62	4.824	30.80
$\sigma 2$	ac	.	.	.	.	.	.	.	.	.	.	0.17	10.79	.	.	.	.	.	.
	as	.	.	.	.	.	.	.	.	.	.	0.40	15.56	.	.	.	.	.	.
	f	.	.	.	.	.	.	.	.	.	.	0.14	14.94	.	.	.	.	.	.
	g	.	.	.	.	.	.	.	.	.	.	0.22	13.65	.	.	.	.	.	.
	S	.	.	.	.	.	.	.	.	.	.	0.32	14.79	.	.	.	.	.	.
κ		.	.	.	.	.	.	.	.	.	.	.	.	0.710	2.64	.	.	0.473	2.18

Table 13: Warsaw theatres model results

		1.MNL		**2. Fixed mon. attr.**		**3. LN**		**4. LU**		**5. WTPS**		**6. WTPS P2**		**7. 3-param. Shifted**		**8. μ-shifted**		**9. κ-shifted**	
LL		-11860.36		-8725.551		-8365.793		-8378.801		-8577.599		-8512.87		-8370.331		-8418.693		-8462.7	
AIC		23732.73		17473.1		16755.59		16781.6		17179.2		0.3514		16766.66		16861.39		16949.4	
BIC		23779.83		17559.46		16849.79		16875.81		17273.41		17057.74		16868.72		16955.59		17043.61	
R2		0.0976		0.3356		0.3629		0.3619		0.3468		17183.35		0.3625		0.3589		0.3556	
		Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.
μ	SQ	0.283	6.41	0.482	5.02	-0.114	-0.97	0.160	1.48	0.057	1.99	0.042	2.22	-0.072	-0.69	-0.262	-2.66	-0.234	-2.46
	roz	0.785	22.15	1.644	20.33	2.025	20.94	1.950	20.25	0.340	23.72	0.196	5.92	2.015	20.46	1.976	21.28	1.883	21.25
	sro	0.489	16.02	1.064	16.35	1.291	16.36	1.261	16.43	0.213	17.39	0.114	1.94	1.281	16.20	1.267	16.65	1.212	16.84
	dzi	0.286	10.24	0.557	9.52	0.763	10.73	0.750	10.76	0.106	7.79	-0.019	-0.84	0.756	10.75	0.720	10.79	0.673	10.54
	eks	0.238	9.54	0.494	8.84	0.563	8.56	0.530	8.17	0.094	8.60	0.206	9.67	0.560	8.43	0.565	8.92	0.543	8.87
	cost	-2.152	-32.86	-4.795	-28.48	1.898	32.97	0.186	2.17	2.054	28.55	2.090	34.18	2.064	22.64	0.922	20.43	.	.
$\sigma1$	SQ	.	.	2.709	26.29	2.481	13.81	2.817	21.37	0.527	23.81	0.52	28.18	2.581	20.62	2.265	17.66	2.262	20.13
	roz	.	.	1.674	17.18	1.681	15.94	1.642	15.63	0.331	33.07	-0.304	-6.93	1.705	14.35	1.595	15.67	1.563	15.49
	sro	.	.	1.078	10.30	-1.204	-10.34	1.141	10.11	0.226	18.83	0.183	2.87	1.140	10.20	1.099	10.11	-1.073	-10.66
	dzi	.	.	0.960	9.35	1.221	11.40	1.225	10.92	0.172	14.68	-0.061	-3.72	1.213	11.83	-1.131	-11.37	-1.073	-10.74
	eks	.	.	0.974	9.16	1.187	11.20	1.126	9.58	0.205	9.99	0.003	0.13	1.160	10.35	-1.084	-9.68	1.025	9.62
	cost	.	.	.	.	1.184	13.75	3.160	23.32	1.027	11.01	1.081	13.99	0.980	11.81	2.128	23.33	2.862	29.07
$\sigma2$	roz	.	.	.	.	.	.	.	.	.	.	0.149	12.22	.	.	.	.	.	.
	sro	.	.	.	.	.	.	.	.	.	.	0.100	1.84	.	.	.	.	.	.
	dzi	.	.	.	.	.	.	.	.	.	.	0.130	5.44	.	.	.	.	.	.
	eks	.	.	.	.	.	.	.	.	.	.	-0.124	-4.94	.	.	.	.	.	.
κ		.	.	.	.	.	.	.	.	.	.			1.136	2.12	.	.	-3.089	-18.49

Table 14: Car choice model results

		1.MNL		2. Fixed mon. attr.		3. LN		4. LU		5. WTPS		6. WTPS P2		7. 3-param. Shifted		8. μ-shifted		9. κ-shifted	
LL		-6930.278		-6427.591		-6270.892		-6277.871		-6326.96		-6295.403		-6271.427		-6284.827		-6310.088	
AIC		13896.56		12925.18		12613.78		12627.74		12725.92		12692.81		12616.85		12641.65		12692.18	
BIC		14021.01		13167.18		12862.69		12876.65		12974.83		13045.42		12872.68		12890.56		12941.08	
R2		0.1495		0.2089		0.228		0.2271		0.2211		0.2231		0.2278		0.2263		0.2232	
		Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.
μ	ASC1	0.064	1.86	0.076	1.82	0.082	1.82	0.076	1.70	0.005	1.02	0.009	2.17	0.079	1.74	0.080	1.74	0.070	0.58
	ASC2	0.126	3.70	0.191	4.58	0.204	4.41	0.194	4.23	0.022	4.09	0.023	4.59	0.202	4.36	0.199	4.33	0.193	2.56
	op_cost	-0.265	-12.86	1.013	9.64	1.076	9.77	1.060	8.93	-1.069	-9.88	-1.308	-15.52	1.094	10.60	1.062	9.48	1.049	8.34
	range	0.453	5.60	-0.672	-3.00	-0.611	-2.87	-0.537	-2.76	-2.579	-15.82	-2.711	-14.61	-0.494	-2.50	-0.466	-2.52	-0.457	-1.80
	engine_electric	-1.448	-10.54	-2.053	-10.43	-2.044	-9.98	-1.989	-10.03	-0.262	-10.62	-0.260	-9.12	-2.033	-9.96	-2.013	-9.61	-2.050	-4.76
	Engine_hybrid	0.412	5.73	0.608	6.26	0.803	7.74	0.816	7.81	0.088	7.39	0.049	5.58	0.767	7.50	0.811	7.71	0.741	7.25
	perf_h	0.444	9.39	-0.843	-5.55	-0.776	-4.68	-0.726	-4.95	-2.961	-22.56	-3.149	-20.29	-0.776	-4.83	-0.783	-5.09	-0.762	-4.49
	perf_m	0.322	7.64	-0.960	-6.15	-0.928	-5.45	-0.802	-5.55	-3.454	-15.60	-3.446	-20.70	-0.805	-5.25	-0.890	-5.84	-0.849	-1.24
	car_mini	-1.763	-13.60	-2.707	-13.18	-3.008	-12.83	-2.973	-12.49	-0.361	-13.84	-0.327	-10.86	-2.992	-12.95	-2.959	-13.07	-2.922	-5.35
	car_small	-0.786	-7.23	-1.223	-8.18	-1.334	-8.02	-1.332	-8.00	-0.160	-5.97	-0.073	-4.74	-1.364	-8.02	-1.342	-7.96	-1.307	-7.40
	car_large	-0.307	-2.73	-0.513	-3.34	-0.448	-2.55	-0.496	-2.82	-0.072	-3.59	-0.006	-0.22	-0.467	-2.61	-0.425	-2.50	-0.473	-2.15
	suv_small	-0.458	-4.15	-0.728	-5.33	-0.769	-4.80	-0.764	-4.72	-0.080	-4.48	-0.071	-4.07	-0.785	-4.86	-0.770	-4.72	-0.754	-2.15
	suv_mid	0.200	1.92	0.261	2.01	0.347	2.28	0.349	2.27	0.040	2.65	0.084	5.88	0.313	2.00	0.358	2.34	0.317	1.71
	suv_large	0.011	0.08	-0.404	-2.18	-0.192	-0.80	-0.144	-0.65	-0.059	-2.18	-0.122	-4.75	-0.186	-0.80	-0.183	-0.80	-0.201	-0.19
	pickup_compact	-0.808	-7.20	-1.210	-7.82	-1.221	-7.10	-1.243	-7.19	-0.156	-10.02	-0.198	-5.48	-1.267	-7.42	-1.276	-7.41	-1.234	-2.91
	pickup_fs	-0.424	-3.59	-0.750	-4.47	-0.716	-3.72	-0.683	-3.74	-0.103	-4.21	-0.244	-10.27	-0.727	-3.79	-0.747	-3.97	-0.707	-0.61
	minivan	-0.188	-1.56	-0.445	-3.06	-0.488	-2.61	-0.500	-2.67	-0.064	-3.44	-0.221	-7.98	-0.489	-2.55	-0.464	-2.47	-0.438	-1.43
	price	-0.532	-21.17	-7.571	-20.01	2.094	35.47	0.768	6.77	2.267	32.07	2.233	33.84	2.321	12.67	1.249	20.74	.	.
$\sigma1$	ASC1	.	.	-0.335	-3.17	-0.320	-3.11	-0.352	-3.43	-0.020	-2.38	0.003	0.30	-0.300	-2.87	0.361	3.39	0.353	0.79
	ASC2	.	.	0.323	3.57	-0.302	-2.82	-0.260	-2.50	-0.016	-1.57	-0.027	-3.77	0.304	3.03	0.331	3.39	0.350	0.90
	op_cost	.	.	0.837	12.97	-0.903	-9.77	0.872	7.61	0.921	15.82	-0.324	-8.78	0.831	13.87	-0.895	-11.42	0.851	1.68
	range	.	.	0.494	4.97	0.465	4.25	0.333	2.24	-0.576	-7.77	-0.150	-3.43	0.268	1.86	0.009	0.06	-0.005	0.00
	engine_electric	.	.	1.237	11.90	1.246	12.48	1.236	10.26	-0.102	-9.92	0.170	9.86	1.308	11.45	1.307	12.98	1.334	11.54
	Engine_hybrid	.	.	1.208	13.87	1.171	11.71	1.203	11.27	0.155	13.04	0.156	12.93	1.196	11.79	1.168	12.07	1.192	5.79
	perf_h	.	.	0.914	10.78	0.893	7.09	0.886	7.83	-0.857	-14.93	-1.297	-5.10	0.927	8.30	0.917	10.77	0.876	4.09
	perf_m	.	.	0.545	4.73	0.595	3.40	-0.332	-2.01	0.797	7.36	-0.247	-1.10	0.368	1.28	0.535	4.08	0.442	0.37
	car_mini	.	.	2.057	8.19	-1.963	-7.47	1.755	4.88	0.269	10.35	0.327	11.30	1.817	5.84	-2.045	-7.84	1.905	2.23
	car_small	.	.	1.171	5.30	1.088	3.94	-1.114	-3.99	-0.162	-3.75	-0.016	-0.76	1.019	3.74	1.174	4.09	1.125	0.92
	car_large	.	.	-1.372	-5.73	1.164	4.66	1.202	4.53	-0.108	-6.06	-0.153	-5.46	1.235	3.95	-1.014	-3.52	-1.279	-0.99
	suv_small	.	.	0.764	2.94	0.802	3.13	-0.796	-3.13	-0.086	-4.16	0.109	9.04	0.658	2.21	0.695	2.06	0.521	0.10
	suv_mid	.	.	0.818	2.98	-0.814	-2.55	0.932	3.68	-0.096	-8.35	0.051	2.67	-1.030	-4.36	0.702	1.83	-0.738	-1.51
	suv_large	.	.	-2.122	-6.80	-1.648	-5.30	1.365	4.36	-0.223	-3.46	-0.141	-8.30	-1.700	-4.48	-1.676	-5.36	-1.767	-1.09
	pickup_compact	.	.	0.855	2.70	0.635	1.24	0.963	3.25	0.141	9.24	0.161	4.20	-0.878	-3.04	-1.098	-4.44	0.828	0.70
	pickup_fs	.	.	1.615	6.13	1.660	5.38	1.515	5.68	0.228	7.12	-0.050	-2.32	-1.476	-4.94	1.590	5.72	1.518	1.38
	minivan	.	.	1.422	6.39	1.528	5.89	-1.665	-6.08	-0.173	-10.25	0.033	3.67	-1.576	-6.30	1.414	5.47	1.395	0.95
	price	.	.	.	.	0.731	15.42	2.541	16.53	-0.910	-9.62	-1.026	-11.10	0.598	6.24	1.191	17.40	1.959	6.32
$\sigma2$	op_cost	.	.	.	.	.	.	.	.	.	.	0.364	14.90	.	.	.	.	.	.
	range	.	.	.	.	.	.	.	.	.	.	0.109	3.22	.	.	.	.	.	.
	engine_electric	.	.	.	.	.	.	.	.	.	.	-0.004	-0.40	.	.	.	.	.	.
	Engine_hybrid	.	.	.	.	.	.	.	.	.	.	0.045	5.41	.	.	.	.	.	.
	perf_h	.	.	.	.	.	.	.	.	.	.	-0.177	-1.96	.	.	.	.	.	.
	perf_m	.	.	.	.	.	.	.	.	.	.	0.071	1.10	.	.	.	.	.	.
	car_mini	.	.	.	.	.	.	.	.	.	.	-0.081	-5.23	.	.	.	.	.	.
	car_small	.	.	.	.	.	.	.	.	.	.	-0.096	-7.18	.	.	.	.	.	.
	car_large	.	.	.	.	.	.	.	.	.	.	-0.107	-4.75	.	.	.	.	.	.
	suv_small	.	.	.	.	.	.	.	.	.	.	-0.044	-6.49	.	.	.	.	.	.
	suv_mid	.	.	.	.	.	.	.	.	.	.	-0.083	-8.51	.	.	.	.	.	.
	suv_large	.	.	.	.	.	.	.	.	.	.	0.080	11.12	.	.	.	.	.	.
	pickup_compact	.	.	.	.	.	.	.	.	.	.	0.010	0.32	.	.	.	.	.	.
	pickup_fs	.	.	.	.	.	.	.	.	.	.	0.139	12.18	.	.	.	.	.	.
	minivan	.	.	.	.	.	.	.	.	.	.	0.176	10.29	.	.	.	.	.	.
κ		.	.	.	.	.	.	.	.	.	.	.	.	1.987	1.23	.	.	-5.007	-3.13

Table 15: Chicken meat model results

		1.MNL		**2. Fixed mon. attr.**		**3. LN**		**4. LU**		**5. WTPS**		**6. WTPS P2**		**7. 3-param. Shifted**		**8. μ-shifted**		**9. κ-shifted**	
LL		-8899.485		-7988.05		-7506.658		-7524.137		-7979.055		-7846.145		-7503.976		-7511.996		-7533.731	
AIC		17814.97		16113.94		15160.35		15195.3		16105.14		15734.29		15164.17		15171.02		0.2982	
BIC		17872.49		16006.1		15045.32		15080.27		15990.11		15885.27		15041.95		15055.99		15214.49	
R2		0.172		0.2561		0.3007		0.2991		0.2568		0.2687		0.3009		0.3002		15099.46	
		Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.
μ	asc1	1.630	13.54	2.240	16.15	2.910	12.97	2.804	12.66	0.075	11.13	0.070	11.32	2.894	12.18	2.885	13.12	2.727	12.92
	asc2	1.579	13.14	2.131	15.69	2.834	12.53	2.732	12.26	0.070	10.80	0.067	12.79	2.818	11.78	2.814	12.74	2.651	12.47
	test	0.262	8.88	0.405	8.09	0.380	8.69	0.359	8.04	0.014	6.99	-0.007	-3.80	0.369	8.43	0.382	8.77	0.364	8.50
	trace	0.220	7.76	0.233	4.93	0.263	6.68	0.242	6.07	0.009	5.00	-0.002	-1.80	0.254	6.34	0.263	6.72	0.250	6.30
	wel	0.315	11.03	0.403	8.33	0.406	9.92	0.383	9.03	0.014	7.27	-0.004	-2.33	0.393	9.45	0.410	10.02	0.391	9.45
	irel	0.503	11.01	0.828	9.39	0.834	9.83	0.854	9.90	0.028	6.61	-0.013	-4.97	0.845	9.75	0.820	9.93	0.818	9.72
	gb	0.138	3.97	0.147	2.41	0.199	3.65	0.228	4.20	0.006	2.60	-0.010	-3.77	0.214	3.82	0.199	3.60	0.212	3.87
	cost	-0.183	-10.14	-27.794	-9.09	2.583	12.46	-3.041	-2.24	3.458	26.10	3.720	29.01	3.083	11.04	1.845	9.23	.	.
$\sigma 1$	asc1	.	.	1.126	5.48	-0.426	-4.33	-0.452	-4.52	0.037	4.75	0.041	8.23	-0.434	-4.44	0.445	4.24	0.411	4.03
	asc2	.	.	1.099	5.44	0.381	3.07	-0.382	-3.21	0.036	4.31	-0.042	-9.14	0.379	2.99	0.353	2.79	0.359	2.97
	test	.	.	0.987	13.19	0.664	10.16	0.672	10.56	0.032	7.74	0.004	0.64	0.663	10.13	0.661	10.28	0.665	10.55
	trace	.	.	0.916	13.57	0.553	8.96	0.573	9.05	0.030	7.45	0.021	8.69	0.551	8.92	0.562	9.33	-0.546	-8.83
	wel	.	.	0.960	14.35	0.611	9.92	-0.619	-9.88	0.032	8.09	-0.018	-6.85	0.614	9.87	0.599	10.03	0.602	10.26
	irel	.	.	1.590	15.35	1.365	12.49	1.439	12.00	0.052	7.87	0.000	0.09	1.373	12.50	1.351	12.61	1.312	11.82
	gb	.	.	1.101	12.80	0.756	9.12	0.706	7.89	0.037	7.37	0.012	4.78	0.758	9.08	0.750	8.81	0.713	7.69
	cost	.	.	.	.	1.709	9.70	8.754	6.22	0.422	4.48	-0.796	-8.61	1.378	7.36	2.177	16.22	-3.436	-16.14
$\sigma 2$	test	.	.	.	.	.	.	.	.	.	.	0.024	8.37	.	.	.	.	.	.
	trace	.	.	.	.	.	.	.	.	.	.	0.011	8.50	.	.	.	.	.	.
	wel	.	.	.	.	.	.	.	.	.	.	0.021	9.11	.	.	.	.	.	.
	irel	.	.	.	.	.	.	.	.	.	.	0.051	9.42	.	.	.	.	.	.
	gb	.	.	.	.	.	.	.	.	.	.	0.016	6.96	.	.	.	.	.	.
κ		.	.	.	.	.	.	.	.	.	.	.	.	9.704	1.79	.	.	-9.282	-3.23

Table 16: Bluerivers model results

		1.MNL		**2. Fixed mon. attr.**		**3. LN**		**4. LU**		**5. WTPS**		**6. WTPS P2**		**7. 3-param. Shifted**		**8. μ-shifted**		**9. κ-shifted**	
LL		-12243.8		-9783.479		-9402.132		-9428.427		-9738.848		-9712.645		-9403.041		-9421.075		-9404.396	
AIC		24511.61		19612.96		18852.26		18904.85		19525.7		19493.29		18856.08		18890.15		18856.79	
BIC		24600.46		19783.25		19029.96		19082.55		19703.39		19745.03		19041.19		19067.85		19034.49	
R2		0.0809		0.2646		0.2931		0.2912		0.2679		0.2691		0.293		0.2917		0.293	
		Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.
μ	ASC	-0.463	-7.24	-2.485	-13.57	-2.515	-13.24	-2.782	-13.90	-4.266	-9.59	-4.056	-10.38	-2.600	-13.79	-2.564	-12.98	-2.562	-13.76
	aue1	0.275	7.46	0.557	10.23	0.581	9.60	0.571	9.30	0.760	9.65	0.764	10.43	0.586	9.65	0.576	9.76	0.585	9.62
	aue2	0.241	6.32	0.624	10.15	0.772	11.07	0.761	10.74	0.859	10.14	0.606	7.74	0.774	11.01	0.747	11.07	0.781	11.16
	wald1	0.242	7.91	0.418	9.46	0.625	12.28	0.628	12.32	0.557	9.25	0.520	10.08	0.632	12.38	0.585	11.85	0.641	12.71
	wald2	0.251	7.60	0.511	10.39	0.657	11.07	0.658	11.01	0.700	10.82	0.954	10.87	0.665	11.06	0.614	10.89	0.676	11.38
	ufer1	0.028	0.62	0.131	2.18	0.245	3.61	0.276	3.97	0.297	4.00	0.350	5.28	0.249	3.64	0.224	3.42	0.255	3.75
	ufer2	0.168	3.74	0.470	7.20	0.559	7.51	0.566	7.55	0.775	9.38	0.874	11.78	0.566	7.54	0.529	7.35	0.572	7.66
	fisch1	0.167	3.32	0.611	8.51	0.552	7.01	0.492	6.17	0.819	9.48	0.402	3.91	0.538	6.67	0.591	7.67	0.517	6.39
	fisch2	0.145	2.79	0.803	9.68	0.738	8.21	0.713	7.86	1.146	11.05	1.290	11.38	0.718	7.77	0.783	8.80	0.714	7.72
	baden1	0.302	4.20	0.505	3.53	1.824	7.62	1.749	10.64	0.399	1.90	-1.166	-4.63	1.749	6.94	1.563	6.94	1.835	9.83
	baden2	0.544	10.43	0.522	6.65	0.833	10.12	0.832	10.14	0.646	7.02	-0.002	-0.02	0.819	9.82	0.835	10.18	0.826	9.87
	preis	-0.447	-18.87	-0.755	-20.15	-0.309	-3.29	2.806	24.84	-0.170	-2.84	-0.088	-1.36	-0.180	-1.19	-1.243	-15.08	.	.
$\sigma 1$	ASC	.	.	4.372	20.44	2.898	10.25	4.330	17.11	6.604	12.77	6.767	13.37	3.183	12.81	2.894	12.03	3.076	13.63
	aue1	.	.	-0.048	-1.60	0.013	0.36	-0.023	-0.71	0.022	0.31	-0.022	-0.74	0.010	0.33	0.015	0.50	0.030	0.90
	aue2	.	.	-0.640	-6.41	-0.721	-6.80	-0.772	-7.30	0.500	2.24	-0.200	-2.16	0.750	7.16	0.725	7.17	-0.743	-7.30
	wald1	.	.	-0.023	-1.56	0.026	1.23	0.012	0.55	-0.039	-1.44	0.023	0.46	0.014	0.53	0.017	0.81	0.006	0.30
	wald2	.	.	-0.092	-0.89	0.488	3.93	-0.477	-3.75	-0.378	-2.74	-0.073	-0.73	0.459	3.23	0.371	2.32	-0.473	-3.47
	ufer1	.	.	0.016	1.11	-0.003	-0.07	-0.039	-0.91	0.036	0.96	0.032	0.80	0.028	0.82	0.006	0.24	-0.044	-0.94
	ufer2	.	.	0.006	0.47	-0.031	-0.62	0.068	1.10	-0.007	-0.25	-0.030	-1.19	0.032	0.56	0.009	0.22	0.090	1.32
	fisch1	.	.	-0.811	-6.38	1.006	7.15	1.031	7.31	0.575	4.55	-0.256	-2.50	1.002	7.14	0.917	6.65	-1.030	-7.36
	fisch2	.	.	0.008	0.16	-0.022	-0.19	0.020	0.18	0.196	1.96	0.076	0.97	0.047	0.41	0.037	0.17	-0.073	-0.70
	baden1	.	.	1.832	9.42	0.988	1.63	-0.750	-1.89	2.078	8.96	-0.969	-9.43	-0.633	-0.63	-0.445	-0.38	-0.746	-1.42
	baden2	.	.	1.674	23.78	-1.354	-16.07	1.300	15.93	2.015	17.16	2.190	18.72	1.353	16.01	1.406	17.51	1.341	15.89
	preis	.	.	.	.	-2.285	-22.10	-6.709	-18.60	-0.882	-11.31	1.051	13.75	-2.081	-14.86	-2.974	-32.13	-2.011	-32.04
$\sigma 2$	aue1	.	.	.	.	.	.	.	.	.	.	-0.009	-0.41	.	.	.	.	.	.
	aue2	.	.	.	.	.	.	.	.	.	.	0.281	5.37	.	.	.	.	.	.
	wald1	.	.	.	.	.	.	.	.	.	.	0.017	0.67	.	.	.	.	.	.
	wald2	.	.	.	.	.	.	.	.	.	.	-0.196	-2.67	.	.	.	.	.	.
	ufer1	.	.	.	.	.	.	.	.	.	.	-0.020	-1.08	.	.	.	.	.	.
	ufer2	.	.	.	.	.	.	.	.	.	.	0.001	0.03	.	.	.	.	.	.
	fisch1	.	.	.	.	.	.	.	.	.	.	0.433	5.33	.	.	.	.	.	.
	fisch2	.	.	.	.	.	.	.	.	.	.	-0.177	-3.59	.	.	.	.	.	.
	baden1	.	.	.	.	.	.	.	.	.	.	1.735	11.71	.	.	.	.	.	.
	baden2	.	.	.	.	.	.	.	.	.	.	0.982	10.40	.	.	.	.	.	.
κ		.	.	.	.	.	.	.	.	.	.	.	.	0.093	1.17	.	.	0.170	3.68

Table 17: Nitrolimit model results

		1.MNL		2. Fixed mon. attr.		3. LN		4. LU		5. WTPS		6. WTPS P2		7. 3-param. Shifted		8. μ-shifted		9. κ-shifted	
LL		-9055.677		-5724.038		-5480.638		-5503.78		-5665.348		-5545.818		-5485.362		-5482.321		-5554.396	
AIC		18141.35		11506.08		11021.28		11067.56		11390.7		11179.64		11032.72		11024.64		11168.79	
BIC		18247.99		11712.24		11234.55		11280.83		11603.97		11492.43		11253.1		11237.91		11382.06	
R2		0.0863		0.394		0.4449		0.4426		0.4263		0.4369		0.4443		0.4447		0.4375	
		Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.	Param.	Rob. T.
	ASC	1.207	11.98	0.288	2.84	0.374	3.29	0.406	3.50	1.087	2.19	-4.051	-22.60	0.347	3.00	0.374	3.41	-1.441	-5.31
	uhavel1	0.005	0.09	1.413	14.24	1.588	13.49	1.545	13.39	4.036	10.59	1.982	15.32	1.539	13.84	1.569	13.98	0.418	3.66
	uhavel2	0.650	12.02	1.859	16.78	1.957	14.22	1.840	13.68	5.561	13.18	3.383	27.05	1.891	14.94	1.941	15.03	1.610	13.27
	uhavel3	0.781	14.65	0.112	1.03	0.178	1.91	0.214	2.34	0.080	0.27	5.226	51.60	0.163	1.76	0.123	1.33	2.016	14.12
	ohavel1	0.157	3.27	1.171	11.28	1.187	10.11	1.202	10.13	3.209	9.84	-0.161	-1.89	1.183	10.06	1.150	9.90	0.152	1.59
	ohavel2	0.747	12.71	1.459	14.17	1.547	14.63	1.543	14.52	4.124	10.17	1.832	22.14	1.557	11.61	1.550	14.51	1.205	9.54
	ohavel3	0.879	14.52	0.283	3.69	0.391	4.60	0.384	4.46	0.963	3.67	4.292	40.45	0.369	4.58	0.378	4.62	1.575	13.88
μ	stadts1	-0.051	-1.20	0.781	9.26	0.720	8.19	0.662	7.50	2.235	7.35	1.127	17.79	0.739	8.59	0.748	8.87	0.402	4.67
	stadts2	0.132	2.96	0.782	6.96	0.709	5.48	0.578	4.65	2.291	4.83	1.819	23.21	0.747	5.78	0.752	5.99	0.760	8.26
	koeps1	0.441	6.50	0.821	13.22	0.905	12.15	0.886	12.07	2.246	9.23	1.280	9.46	0.891	12.44	0.898	12.44	0.752	5.64
	koeps2	0.566	13.85	0.826	6.84	0.731	5.06	0.619	4.44	2.440	2.76	1.367	19.76	0.770	5.28	0.812	5.60	0.927	12.38
	dahme1	0.334	5.39	1.323	9.85	1.142	7.09	0.949	6.14	3.700	3.85	1.386	10.29	1.179	7.45	1.237	7.77	0.778	5.02
	dahme2	0.541	6.15	1.107	13.46	1.194	12.89	1.194	12.76	2.918	6.99	2.155	11.06	1.160	13.26	1.186	13.10	1.183	6.76
	dahme3	0.669	13.60	-0.334	-14.97	-1.346	-10.77	-7.067	-17.80	-0.977	-9.70	1.721	21.49	-1.696	-4.25	-2.240	-21.43	1.219	13.11
	cost	-0.079	-13.66	-0.129	-0.33	-1.041	-4.08	-1.070	-4.14	-1.220	-2.15	-0.164	-1.66	-1.207	-4.63	-1.310	-5.28	.	.
	ASC	.	.	0.832	6.51	-0.671	-4.61	-0.455	-1.25	1.594	2.06	20.567	34.23	0.638	3.63	-0.582	-3.50	0.552	3.48
	uhavel1	.	.	0.118	0.42	-0.062	-0.11	-0.376	-1.62	-0.626	-1.05	0.536	11.64	-0.055	-0.16	0.399	2.25	0.404	2.43
	uhavel2	.	.	0.819	8.52	-0.997	-9.38	1.050	9.26	1.811	5.03	0.536	9.41	0.940	6.17	-0.995	-8.73	1.066	8.64
	uhavel3	.	.	-0.243	-0.20	0.223	0.80	0.456	3.42	1.179	0.97	0.818	16.97	0.021	0.01	0.517	3.68	0.440	2.22
	ohavel1	.	.	0.905	3.24	0.974	7.23	-0.980	-6.49	1.739	1.59	0.825	16.08	0.992	7.14	-0.912	-6.59	-0.979	-6.44
	ohavel2	.	.	-1.442	-13.09	1.482	12.22	1.422	12.42	-3.376	-4.56	1.197	17.95	1.441	13.13	1.409	13.41	-1.421	-11.90
	ohavel3	.	.	0.105	0.62	0.164	1.73	-0.031	-0.25	-0.393	-1.52	-1.523	-18.59	0.114	0.27	0.179	1.38	-0.224	-2.05
$\sigma 1$	stadts1	.	.	-0.656	-5.94	0.714	6.65	0.651	4.88	-1.502	-3.86	-0.288	-7.16	0.639	5.31	0.699	6.62	0.627	5.15
	stadts2	.	.	-0.315	-1.78	-0.199	-1.37	0.102	0.83	-0.056	-0.02	-0.646	-12.88	-0.134	-1.06	-0.047	-0.30	-0.230	-1.17
	koeps1	.	.	-0.552	-4.87	-0.656	-6.22	-0.717	-6.81	-1.111	-1.97	0.110	2.40	-0.685	-7.14	0.696	7.12	0.756	7.62
	koeps2	.	.	-0.483	-3.76	0.467	2.65	0.393	2.33	-1.075	-3.49	-0.836	-17.18	0.384	2.08	0.499	3.28	-0.457	-1.40
	dahme1	.	.	-0.121	-0.25	0.471	2.48	0.592	3.73	0.865	1.30	-0.061	-0.94	0.510	3.10	0.349	1.95	0.471	1.36
	dahme2	.	.	0.867	8.75	1.028	9.62	-0.979	-8.96	1.790	6.21	0.691	10.88	0.942	9.79	0.993	8.97	-1.011	-9.15
	dahme3	.	.	6.540	15.82	2.709	19.75	10.419	16.80	8.371	9.51	1.294	23.40	3.760	5.50	3.878	26.34	2.615	11.16
	cost	.	.	.	.	3.639	9.19	2.942	10.43	0.722	11.58	1.444	13.61	3.029	8.57	3.295	9.61	3.657	11.50
	ASC	.	.	.	.	.	.	.	.	.	.	-1.859	-2.42	.	.	.	.	.	.
	uhavel1	.	.	.	.	.	.	.	.	.	.	-0.528	-10.28	.	.	.	.	.	.
	uhavel2	.	.	.	.	.	.	.	.	.	.	-0.022	-0.66	.	.	.	.	.	.
	uhavel3	.	.	.	.	.	.	.	.	.	.	-0.209	-6.06	.	.	.	.	.	.
	ohavel1	.	.	.	.	.	.	.	.	.	.	0.028	0.95	.	.	.	.	.	.
	ohavel2	.	.	.	.	.	.	.	.	.	.	0.541	13.86	.	.	.	.	.	.
	ohavel3	.	.	.	.	.	.	.	.	.	.	-1.412	-22.93	.	.	.	.	.	.
$\sigma 2$	stadts1	.	.	.	.	.	.	.	.	.	.	0.043	2.17	.	.	.	.	.	.
	stadts2	.	.	.	.	.	.	.	.	.	.	0.406	9.24	.	.	.	.	.	.
	koeps1	.	.	.	.	.	.	.	.	.	.	0.183	4.95	.	.	.	.	.	.
	koeps2	.	.	.	.	.	.	.	.	.	.	0.095	3.72	.	.	.	.	.	.
	dahme1	.	.	.	.	.	.	.	.	.	.	0.303	11.45	.	.	.	.	.	.
	dahme2	.	.	.	.	.	.	.	.	.	.	0.581	7.40	.	.	.	.	.	.
	dahme3	.	.	.	.	.	.	.	.	.	.	0.981	22.43	.	.	.	.	.	.
	κ	.	.	.	.	.	.	.	.	.	.	.	.	-0.079	-2.90	.	.	0.049	1.28

Sinh-arcsinh distributions: a broad family giving rise to powerful tests of normality and symmetry

Article *in* Biometrika · October 2009

DOI: 10.1093/BIOMET%2FASP053 · Source: RePEc

Arthur Pewsey
Universidad de Extremadura

Sinh-arcsinh distributions

Journal Article

Sinh-arcsinh distributions: a broad family giving rise to powerful tests of normality and symmetry

M.C. Jones

The Open University, UK

and Arthur Pewsey

University of Extremadura, Spain

Summary. We introduce the 'sinh-arcsinh transformation' and thence, by applying it to random variables from some 'generating' distribution with no further parameters beyond location and scale (which we take for most of the paper to be the normal), a new family of 'sinh-arcsinh distributions'. This four parameter family has both symmetric and skewed members and allows for tailweights that are both heavier and lighter than those of the generating distribution. The 'central' place of the normal distribution in this family affords likelihood ratio tests of normality that appear to be superior to the state-of-the-art because of the range of alternatives against which they are very powerful. Likelihood ratio tests of symmetry are also available and very successful. Three-parameter symmetric and asymmetric subfamilies of the full family are of interest too. Heavy-tailed symmetric sinh-arcsinh distributions behave like Johnson S_U distributions while light-tailed symmetric sinh-arcsinh distributions behave like Rieck and Nedelman's sinh-normal distributions, the sinh-arcsinh family allowing a seamless transition between the two, via the normal, controlled by a single parameter. The sinh-arcsinh family is very tractable and many properties are explored. Likelihood inference is pursued, including an attractive reparametrisation. A multivariate version is considered. Options and extensions are discussed.

Keywords: Heavy tails; Johnson's S_U distribution; Light tails; Sinh-normal distribution; Skew-normal distribution; Skewness; Transformation.

Address for correspondence: M.C. Jones, Department of Mathematics & Statistics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK
E-mail: m.c.jones@open.ac.uk

1. Introduction

Families of distributions with four parameters, accounting for location, scale and, in some appropriate senses, skewness and tailweight, cover many of the most important aspects of any unimodal distribution on $\mathbb{R}$. They can be used to accommodate the random parts of regression-type models where, typically, they allow potentially complex modelling of the location (and perhaps scale) parameters while acting robustly with respect to asymmetry and weight of tails. Subsets of the Pearson and Johnson families of distributions are famous examples (Johnson *et al.*, 1994, Chapter 12); stable laws (Samorodnitsky and Taqqu, 1994), generalised hyperbolic distributions (Barndorff-Nielsen, 1978), two-piece distributions (Fernandez and Steel, 1998), generalised distributions of order statistics (Jones, 2004) and a very popular class of skew distributions in which a symmetric density is perturbed by a rescaled symmetric distribution function (Azzalini, 1985, Genton, 2004) are among other examples. Many more families live on finite or semi-infinite support.

Broadly speaking, most of these families of distributions have the normal distribution as a special, often a limiting, case with other members of the families having heavier tails than the normal. In this paper, we propose a novel relatively simple and tractable four-parameter family of distributions on $\mathbb{R}$ with the normal distribution 'situated centrally' and other members having *both lighter and heavier tails*. This has practical benefits especially in affording excellent tests of the appropriateness of the normal distribution.

To describe the new distributions, consider their canonical case in which location $\mu \in \mathbb{R}$ and scale $\sigma > 0$ are removed; they can be reinstated for practical work in the usual way by utilising $\sigma^{-1} f_{\epsilon,\delta}(\sigma^{-1}(x-\mu))$ where $f_{\epsilon,\delta}(x)$ is the density of a member of the new family. Here, $\epsilon \in \mathbb{R}$ will turn out to be a skewness parameter and $\delta > 0$ will control tailweight. Associate random variables Z and $X_{\epsilon,\delta}$ with the standard normal density ϕ and $f_{\epsilon,\delta}$, respectively. Then, we propose to define $f_{\epsilon,\delta}$ by what we shorthandedly call the 'sinh-arcsinh transformation'

$$Z = S_{\epsilon,\delta}(X_{\epsilon,\delta}) \equiv \sinh\{\epsilon + \delta \sinh^{-1}(X_{\epsilon,\delta})\}. \tag{1}$$

It follows that the density of the 'sinh-arcsinh distribution' is given by

$$f_{\epsilon,\delta}(x) = \frac{1}{\sqrt{2\pi}} \frac{\delta C_{\epsilon,\delta}(x)}{\sqrt{1+x^2}} \exp\left\{-\frac{1}{2} S^2_{\epsilon,\delta}(x)\right\}, \tag{2}$$

where $C_{\epsilon,\delta}(x) = \cosh(\epsilon + \delta \sinh^{-1}(x)) = \sqrt{1 + S^2_{\epsilon,\delta}(x)}$. Of course, $f_{0,1}(x) = \phi(x)$. Examples of densities (2) can be seen in Fig. 1 to follow. We note in passing that, unlike some other families of distributions, no special functions appear in the definition of the density of the sinh-arcsinh distribution above.

Properties of the full family (2) are considered in Section 2 and further properties of the three-parameter symmetric subfamily thereof (corresponding to $\epsilon = 0$) in Section 3. A considerable degree of tractability is evident in the provision of distribution and quantile functions, unimodality and moments. Tailweights are also considered. It is shown that ϵ (Section 2.2) and δ (in the symmetric case; Section 3.1) are skewness and kurtosis parameters in the sense of van Zwet (1964). A three-parameter subfamily of 'skew-normal' distributions is briefly described in Section 2.5. In Section 3.3 it is shown how, in the symmetric case, the small δ (heavy-tailed) members of family (2) behave like Johnson's (1949) S_U distributions while the large δ (light-tailed) members behave like Rieck and Nedelman's (1991) sinh-normal distributions. In this sense, the symmetric sinh-arcsinh distributions form a seamless combination of the two, the single-parameter δ controlling the transition from one to the other via the normal distribution ($\delta = 1$).

Likelihood fitting of the sinh-arcsinh distribution in the form of (2) with location and scale parameters introduced is considered in Section 4. Asymptotic properties are considered in Section 4.1, leading to a useful reparametrisation in Section 4.2. Although these subsections concentrate on the three-parameter symmetric subfamily of sinh-arcsinh distributions, we employ (and recommend) the same reparametrisation for use in fitting the full four-parameter family (Section 4.3). An example illustrating the modelling flexibility of the full sinh-arcsinh family is presented in Section 4.4.

Likelihood ratio tests (LRTs) of normality are immediately available within the sinh-arcsinh family: $H_0 : \epsilon = 0, \delta = 1$. The performance of these tests is investigated in a substantial simulation study reported in Section 5. We actually consider testing for normality against either symmetric or asymmetric alternatives and against alternatives both within and beyond the sinh-arcsinh family. We compare performance with that of seven of the best performing omnibus tests of normality and conclude that our LRTs appear to provide the best tests of normality.

A similar large simulation study of the sinh-arcsinh LRT for testing symmetry ($H_0 : \epsilon = 0$) was undertaken and is reported in Section 6. Again, we observe excellent performance and show that it outperforms two competing omnibus tests chosen as representing the 'state-of-the-art'.

There is an immediate and straightforward extension of the univariate distributions above to the multivariate case by marginal transformation of a multivariate normal distribution. The resulting multivariate distributions are considered relatively briefly in Section 7 with some emphasis on their dependence properties.

In Section 8, we consider three ways in which the sinh-arcsinh distribution (2) might be/have been formulated differently. In Section 8.1, we discuss the choice of transformation function within the class of transformations of the form $H(\epsilon + \delta H^{-1}(X))$. In Section 8.2, we investigate alternative options to the normal for the role of the 'central' symmetric distribution in the family. And in Section 8.3, we explore a different approach to skewing the (same) symmetric members of the family. While there prove to be a number of interesting considerations and alternatives, the end result is a justification — for most general use — of the choices made in (2).

We close with discussion in Section 9.

2. Properties of family (2)

2.1. Basic properties

We begin by noting several equivalent formulations of transformation (1):

$$\begin{aligned} S_{\epsilon,\delta}(X) &= \frac{1}{2}\left\{e^{\epsilon}\exp(\delta\sinh^{-1}(X)) - e^{-\epsilon}\exp(-\delta\sinh^{-1}(X))\right\}, \\ &= \frac{1}{2}\left\{e^{\epsilon}(\sqrt{X^2+1}+X)^{\delta} - e^{-\epsilon}(\sqrt{X^2+1}+X)^{-\delta}\right\} \qquad (3) \\ &= \frac{1}{2}\left\{e^{\epsilon}(\sqrt{X^2+1}+X)^{\delta} - e^{-\epsilon}(\sqrt{X^2+1}-X)^{\delta}\right\}. \qquad (4) \end{aligned}$$

Also, $\sinh^{-1}(Z) = \epsilon + \delta\sinh^{-1}(X_{\epsilon,\delta})$ or $X_{\epsilon,\delta} = \sinh[\delta^{-1}\{\sinh^{-1}(Z) - \epsilon\}]$. Random variate generation is immediate using the latter formula.

Second, the distribution function associated with density (2) is readily written as

$$F_{\epsilon,\delta}(x) = \Phi(S_{\epsilon,\delta}(x)),$$

where Φ is the standard normal distribution function.

Third, since $S_{\epsilon,\delta}^{-1}(z) = S_{-\epsilon/\delta,1/\delta}(z)$, the quantile function associated with density (2) is

$$Q_{\epsilon,\delta}(u) = S_{-\epsilon/\delta,1/\delta}(\Phi^{-1}(u)), \qquad 0 < u < 1. \qquad (5)$$

In particular, the median of the distribution is $-\sinh(\epsilon/\delta)$.

Fourth, density (2) is always unimodal. To see this, the first derivative of $\log f_{\epsilon,\delta}(x)$ is of the form

$$-\frac{x}{1+x^2} - \frac{\delta S^3_{\epsilon,\delta}(x)}{\sqrt{1+x^2}\ C_{\epsilon,\delta}(x)}.$$

Any point x_0 for which this derivative is zero satisfies

$$\frac{\delta S^3_{\epsilon,\delta}(x_0)}{\sqrt{1+S^2_{\epsilon,\delta}(x_0)}} = -\frac{x_0}{\sqrt{1+x_0^2}}.$$

But the left-hand side of this equation is a monotonically increasing function of x_0 taking all real values while the right-hand side is a monotonically decreasing function of x_0 taking values from 1 to -1. It follows that there can only be one crossing point and so the density is unimodal. Of course, when $\epsilon = 0$, $x_0 = 0$, else $x_0 \neq 0$.

2.2. Skewness

First, in this subsection, let us note that $f_{-\epsilon,\delta}(x) = f_{\epsilon,\delta}(-x)$.

We can show that, for fixed δ, ϵ acts as a skewness parameter in the sense of van Zwet's (1964) skewness ordering. This ordering defines $G_1 \leq_2 G_2$ if $G_2^{-1}(G_1)$ is convex for all x. So now let $G_1 = F_{\epsilon_1,\delta}$ and $G_2 = F_{\epsilon_2,\delta}$ for $\epsilon_1 > \epsilon_2$. Then $F^{-1}_{\epsilon_2,\delta}(F_{\epsilon_1,\delta}(x)) = S_{c,1}(x)$, where $c = (\epsilon_1 - \epsilon_2)/\delta > 0$, and

$$\frac{d^2 F^{-1}_{\epsilon_2,\delta}(F_{\epsilon_1,\delta}(x))}{d^2x} = \frac{\sqrt{1+S^2_{c,1}(x)}}{1+x^2}\left(\frac{S_{c,1}(x)}{\sqrt{1+S^2_{c,1}(x)}} - \frac{x}{\sqrt{1+x^2}}\right),$$

which is positive because $S_{c,1}(x) > x$ for $c > 0$. Note that distribution (2) is parametrised in such a way that, while the absolute value of skewness increases with increasing $|\epsilon|$, positive skewness corresponds to negative ϵ.

This attractive result about monotonicity of skewness allows us to calculate the limits to the achievable range of skewness values in family (2). Consider the Bowley skewness (e.g. Bowley, 1937) defined by

$$B_{\epsilon,\delta} \equiv \frac{Q_{\epsilon,\delta}(3/4) - 2Q_{\epsilon,\delta}(1/2) + Q_{\epsilon,\delta}(1/4)}{Q_{\epsilon,\delta}(3/4) - Q_{\epsilon,\delta}(1/4)}.$$

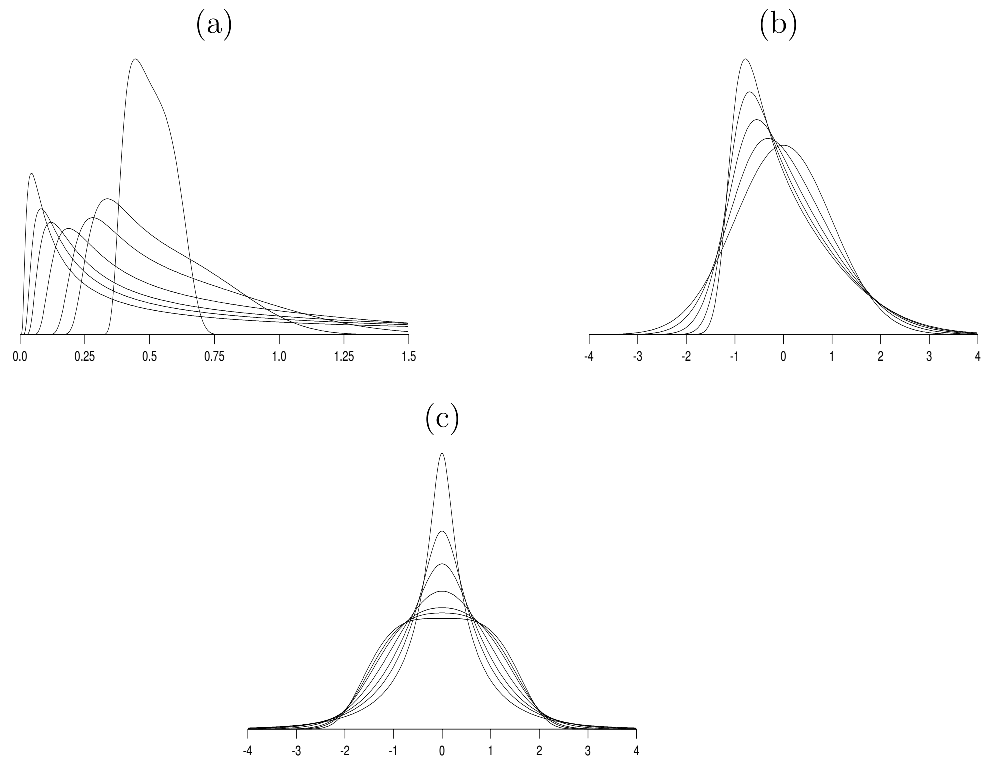


Figure 1: (a) densities $f_{-\infty,\delta}$ for, reading from left to right, $\delta = 0.5, 0.625, 0.75, 1, 1.5, 2, 5$; (b) normalised densities $\sigma_{\epsilon,1} f_{\epsilon,1}(\sigma_{\epsilon,1} x + \mu_{\epsilon,1})$ for, in increasing degree of skewness, $\epsilon = 0, -0.25, -0.5, -0.75, -1$; (c) scaled densities $\sigma_{0,\delta} f_{0,\delta}(\sigma_{0,\delta} x)$ for, in decreasing value of $\sigma_{0,\delta} f_{0,\delta}(0)$, $\delta = 0.5, 0.625, 0.75, 1, 1.5, 2, 5$.

This measure is monotone in ϵ because the distribution follows van Zwet's skewness ordering (Groeneveld and Meeden, 1984) and, in general, can take any values between -1 and 1. It is easy to show that, as $\epsilon \to \pm\infty$, $B_{\epsilon,\delta} \to \mp(k_\delta - 1)/(k_\delta + 1)$ where $k_\delta \equiv \exp(\sinh^{-1}(\Phi^{-1}(3/4))/\delta) \approx \exp(0.6316/\delta)$.

It is possible to identify the limiting densities $f_{\epsilon,\delta}$ as $\epsilon \to \pm\infty$. For concreteness, let us work with negative ϵ (positive skewness) and call the limiting densities $f_{-\infty,\delta}$. Employing suitable normalisation of mean and location, the limiting densities turn out to be

$$f_{-\infty,\delta}(y) = \frac{1}{\sqrt{2\pi}} \frac{\delta \cosh(\delta \log 2y)}{y} \exp\left\{ -\frac{1}{2} \sinh^2(\delta \log 2y) \right\},$$

with support $y > 0$. These are the densities of $Y = \exp(\sinh^{-1}(Z)/\delta)/2$, where Z is standard normal, and are plotted in Fig. 1(a) for a range of values of δ. (The reader might prefer to look first at the less extreme members of family (2) shown in Fig. 1(b) and Fig. 1(c).) Note that all the densities in Fig. 1(a) have median 1/2! Density $f_{5,-\infty}$ — which we shall shortly confirm is associated with very light tails — is not very skew, but most of the others are. Any limitations associated with the range of available skewness values determined above seem mild.

Similar consideration of the kurtosis role of δ is delayed until consideration of the symmetric subfamily in Section 3.1.

2.3. Tailweight

As $|x| \to \infty$, $S_{\epsilon,\delta}(x) \sim 2^{\delta-1}\text{sgn}(x)\exp(\text{sgn}(x)\epsilon)|x|^\delta$ and $C_{\epsilon,\delta}(x) \sim 2^{\delta-1}\exp(\text{sgn}(x)\epsilon)|x|^\delta$. It follows that, retaining the position of ϵ (but not other constants) in asymptotic formulae even though it does not affect rates,

$$f_{\epsilon,\delta}(|x|) \sim \exp(\text{sgn}(x)\epsilon)|x|^{\delta-1}\exp(-e^{\text{sgn}(x)2\epsilon}|x|^{2\delta}). \tag{6}$$

Such tails are closely related to Weibull and 'semi-heavy' tails for small δ, being heavier than exponentially decaying tails and lighter than tails decreasing as a power of $|x|$. We also see the effect of ϵ, through $\exp(\pm\epsilon)$, on the relative scales of the tails of the distribution. This is a major contributory factor to the way in which ϵ controls skewness.

2.4. Moments

The moments — which necessarily all exist as a consequence of the tail behaviour given by (6) — are available for family (2). Using the version of (3) associated with the inverse sinh-arcsinh transformation, we have

$$\begin{aligned} E(X^r_{\epsilon,\delta}) &= \frac{1}{2^r}E\left[\left\{e^{-\epsilon/\delta}\left(Z+\sqrt{Z^2+1}\right)^{1/\delta} - e^{\epsilon/\delta}\left(Z+\sqrt{Z^2+1}\right)^{-1/\delta}\right\}^r\right] \\ &= \frac{1}{2^r}\sum_{i=0}^{r}\binom{r}{i}(-1)^i\exp\left((r-2i)\frac{\epsilon}{\delta}\right)P_{(r-2i)/\delta} \end{aligned}$$

where

$$P_q = E\left\{\left(Z+\sqrt{Z^2+1}\right)^q\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(x+\sqrt{x^2+1}\right)^q e^{-x^2/2}dx$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{8\pi}} \int_0^\infty w^q \left(1+\frac{1}{w^2}\right) \exp\left\{-\frac{1}{8}\left(w-\frac{1}{w}\right)^2\right\} dw \\
&= \frac{e^{1/4} 8^{(q+1)/2}}{\sqrt{32\pi}} \int_0^\infty z^{(q-1)/2} \left(1+\frac{1}{8z}\right) \exp\left\{-\left(z+\frac{1}{64z}\right)\right\} dz \\
&= \frac{e^{1/4}}{\sqrt{8\pi}} \left\{K_{(q+1)/2}(1/4) + K_{(q-1)/2}(1/4)\right\},
\end{aligned}
$$

using (3.471.12) of Gradshteyn and Ryzhik (1994). A property of the modified Bessel function is that $K_{-\nu}(z) = K_\nu(z)$. It follows that $P_{-q} = P_q$, which confirms that odd moments of X are, indeed, zero in the symmetric case where $\epsilon = 0$.

In particular, we have for the mean

$$
\begin{aligned}
\mu_{\epsilon,\delta} &\equiv \mathrm{E}(X_{\epsilon,\delta}) = -\sinh(\epsilon/\delta) P_{1/\delta} \\
&= -\sinh(\epsilon/\delta) \frac{e^{1/4}}{\sqrt{8\pi}} \left(K_{(1+\delta)/(2\delta)}(1/4) + K_{(1-\delta)/(2\delta)}(1/4)\right)
\end{aligned}
$$

and for the variance,

$$
\begin{aligned}
\sigma^2_{\epsilon,\delta} &\equiv \mathrm{Var}(X_{\epsilon,\delta}) = \frac{1}{2}\left(\cosh(2\epsilon/\delta) P_{2/\delta} - 1\right) - \mu^2_{\epsilon,\delta} \\
&= \cosh(2\epsilon/\delta) \frac{e^{1/4}}{\sqrt{32\pi}} \left(K_{(2+\delta)/(2\delta)}(1/4) + K_{(2-\delta)/(2\delta)}(1/4)\right) - \frac{1}{2} - \mu^2_{\delta,\epsilon}.
\end{aligned}
$$

When $\epsilon = 0, \delta = 1$, it can be readily checked that $\mathrm{Var}(X_{0,1}) = 1$.

2.5. An asymmetric subfamily

There may also be some specific interest in the particular three-parameter subfamily of (2) in which $\delta = 1$. In this case transformation (1) has, through (4), the attractively simple form

$$
S_{\epsilon,1}(X) = \sinh(\epsilon)\ \sqrt{1+X^2} + \cosh(\epsilon)\ X.
$$

These densities, some of which are displayed in Fig. 1(b), are true 'skew-normal' distributions in the sense of admitting normality as well as asymmetry and, by (6), retaining two normal-like tails. They share this property with 'two-piece' normal distributions (Fechner, 1897, Fernandez and Steel, 1998, Mudholkar and Hutson, 2000) in which two differentially scaled halves of a normal distribution are joined together. However, unlike the two-piece normal, density (2) is infinitely differentiable at all $x \in \mathbb{R}$. The current and two-piece densities differ from the popular skew-normal distribution with density $2\phi(x)\Phi(\lambda x)$ (Azzalini, 1985, Genton, 2004) for which a side-effect of introducing the skewness parameter λ is a change to the weight in one of the tails.

3. The symmetric subfamily

When $\epsilon = 0$ in transformation (1), density (2) is symmetric about 0. The properties discussed in Section 2 translate to the current special case in a straightforward way. (Inter alia, the mean, median and mode, of course, all reduce to 0 in this case.) In addition, computations strongly suggest that the tails of $f_{0,\delta}$ are sufficiently light for $f_{0,\delta}$ to be log-concave for all $\delta \geq 1$.

3.1. Kurtosis

We can show that, for $\epsilon = 0$, δ acts as a kurtosis parameter in the sense of van Zwet's (1964) ordering, which defines $G_1 \leq_S G_2$ for distributions G_1 and G_2 symmetric about zero if the function $G_2^{-1}(G_1)$ is convex for $x > 0$. In our case, let $G_1 = F_{0,\delta_1}$ and $G_2 = F_{0,\delta_2}$ for $\delta_1 > \delta_2$. Then $F_{0,\delta_2}^{-1}(F_{0,\delta_1}(x)) = S_{0,\delta}(x)$ where $\delta = \delta_1/\delta_2 > 1$. Then

$$\frac{d^2 F_{0,\delta_2}^{-1}(F_{0,\delta_1}(x))}{d^2 x} = \frac{\delta\sqrt{1+S_{0,\delta}^2(x)}}{1+x^2}\left(\frac{\delta S_{0,\delta}(x)}{\sqrt{1+S_{0,\delta}^2(x)}} - \frac{x}{\sqrt{1+x^2}}\right),$$

two d's have been changed to δ's here which is positive because $\delta > 1$ and, correspondingly, $S_{0,\delta}(x) > x$ for $x > 0$.

From Section 2.4, we find

$$\begin{aligned} E(X_{0,\delta}^4) &= \frac{e^{1/4}}{\sqrt{512\pi}}\left\{K_{(4+\delta)/(2\delta)}(1/4) + K_{(4-\delta)/(2\delta)}(1/4)\right. \\ &\qquad \left. - 4\left(K_{(2+\delta)/(2\delta)}(1/4) + K_{(2-\delta)/(2\delta)}(1/4)\right)\right\} + \frac{3}{8}\,. \end{aligned}$$

It can be checked that $E(X_{0,1}^4) = 3$. Given that $f_{0,\delta}$ obeys van Zwet's ordering, the classical kurtosis measure $\beta_2 = E(X_{0,\delta}^4)/\sigma_{0,\delta}^4$ must be monotone decreasing in δ.

3.2. Graphs of density

A range of symmetric members of family (2) is plotted in Fig. 1(c). The densities have been scaled to unit variance by use of the formula for the variance in Section 2.4 (with $\epsilon = 0$). Because of this, the densities are in the reverse order at 0 to what they would have been unscaled, for then $f_{0,\delta}(0) = \delta/\sqrt{2\pi}$. Unimodality, tailweight and kurtosis properties from above are well illustrated by this picture. Notice how the densities vary from the heavy tailed when δ is small, through the normal when $\delta = 1$, to 'wide-bodied'/light tailed densities when δ is large.

3.3. Links to Johnson S_U and sinh-normal distributions

Consider again transformations of a standard normal random variable Z of the form $Z = T_\delta(X)$ for some odd function T_δ generating symmetric distributions for X also on $\mathbb{R}$. Again, δ controls tailweight. This paper, of course, concerns the transformation $T_\delta(X) = S_{0,\delta}(X) = \sinh(\delta \sinh^{-1}(X))$. The two 'component parts' of transformation $S_{0,\delta}(X)$, the sinh and arcsinh transformations, have each previously been employed separately in the same manner. First, when $T_\delta(X) = \delta \sinh^{-1}(X)$, we have the symmetric members of Johnson's (1949) S_U distributions, part of the famous family of transformation-based distributions which also have members on $\mathbb{R}^+$ and $[0, 1]$. See Johnson et al. (1994, Section 12.4.3). These distributions all have tails that are heavier than those of the normal. Second, when Z is normal and $T_{\delta'}(X) = \delta' \sinh(X)$, we have Rieck and Nedelman's (1991) sinh-normal distributions. These symmetric distributions all have tails that are lighter than those of the normal. Indeed, as noted by Rieck and Nedelman (using different notation) the sinh-normal distribution is log-concave for $\delta' \geq 1$, but there is a problem for $\delta' < 1$: the distribution is then bimodal. This is unattractive both because of the form of the bimodality which seems unlikely to be of practical interest and because we feel it better to model bi- and multi-modality through interpretable mixtures of unimodal components.

Now, when δ is small, it is immediate from (2) that

$$f_{0,\delta}(x) \simeq \frac{1}{\sqrt{2\pi}} \frac{\delta}{\sqrt{1+x^2}} \exp\left[-\frac{1}{2}\{\delta \sinh^{-1}(x)\}^2\right].$$

This is precisely the symmetric Johnson S_U density. It can also be shown that, suitably scaled, the limiting form of $f_{0,\delta}$ when $\delta \to \infty$ is

$$f_{0,\infty}(x) = \frac{1}{\sqrt{2\pi}} \cosh(x) \exp\left\{-\frac{1}{2}\sinh^2(x)\right\}.$$

This is the unimodal special case of the sinh-normal distribution with $\delta' = 1$.

These results are very gratifying. They show that by the use of transformation $\sinh(\delta \sinh^{-1}(X))$, we have achieved a 'seamless' family of distributions which 'centre on' the normal distribution, behave very much like Johnson's S_U distributions for tailweights heavier than normal, and like Rieck and Nedelman's sinh-normal distributions for tailweights lighter than normal. Furthermore, recalling that the normal distribution corresponds to '$\delta' = \infty$', the correspondence with the sinh-normal distribution only goes 'down as far as' Rieck and Nedelman's $\delta' = 1$, i.e. automatically stopping just before bimodality kicks in!

Similar reasoning shows why the dual transformation $T_{\delta''}(X) = \sinh^{-1}(\delta'' \sinh(X))$ is not to be recommended for further investigation. For small δ'', $T_{\delta''}(X) \simeq \delta'' \sinh(X)$ which, again, affords Rieck and Nedelman's (1991) sinh-normal distributions. However, these correspond to small $\delta' = \delta''$ cases of the sinh-normal distribution and hence to bimodality.

4. On maximum likelihood estimation

For fitting to one-sample data, family (2) is expanded to a four-parameter family by the addition of location, μ, and scale, σ, parameters in the usual way i.e. by fitting $\sigma^{-1} f_{\epsilon,\delta}(\sigma^{-1}(x-\mu))$. The theoretical work to follow in Sections 4.1 and 4.2 concentrates specifically on the symmetric, $\epsilon = 0$, case. However, this work informs our fitting of the full model also, as described in Section 4.3. Note also that one-sample considerations generalise readily to the important wide class of regression situations in which the sinh-arcsinh distribution can be used to provide a general family of response conditional distributions and location (and possibly one or more other parameters) is modelled as a simple parametric, e.g. linear, function of covariates.

4.1. Maximum likelihood asymptotics in the symmetric case

Manipulations to derive the score equations and elements of the observed information matrix are standard if tedious, and are not given here. We move straight to consideration of the expected information matrix which is n times the matrix made up of values of $\iota_{\eta\xi} = E\{-(\partial^2 \ell / \partial\eta\partial\xi)(Y)\}$, $\eta, \xi = \{\mu, \sigma, \delta\}$. We find we have the special structure

$$\iota_{\mu\mu} = f_m(\delta)/\sigma^2, \qquad \iota_{\mu\sigma} = 0, \qquad \iota_{\mu\delta} = 0,$$

$$\iota_{\sigma\sigma} = f_s(\delta)/\sigma^2, \qquad \iota_{\sigma\delta} = f_c(\delta)/\sigma, \qquad \iota_{\delta\delta} = f_d(\delta),$$

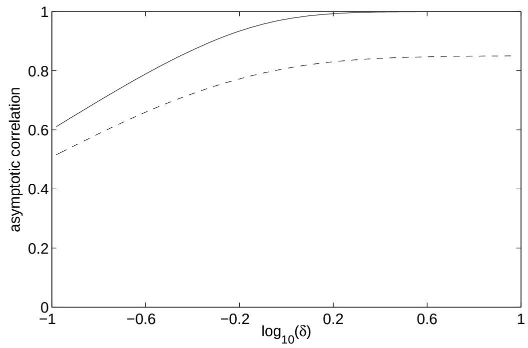


Figure 2: The asymptotic correlations between $\hat{\sigma}$ and $\hat{\delta}$ (solid line) and between $\hat{\sigma}_\delta$ and $\hat{\delta}$ (dashed line) in the symmetric case, plotted as a function of $\log_{10}\delta$.

say, where the f functions are all independent of μ and σ. This structure is a consequence of the symmetry of the fitted model. In fact, we have

$$f_m(\delta) = E\left[\frac{\delta^2 Z^2(3+2Z^2)}{C^2_{0,1/\delta}(Z)(1+Z^2)} - \frac{\delta S_{0,1/\delta}(Z)Z^3}{C^3_{0,1/\delta}(Z)\sqrt{1+Z^2}} + \frac{\{1-S^2_{0,1/\delta}(Z)\}}{C^4_{0,1/\delta}(Z)}\right],$$

$$f_s(\delta) = E\left[S^2_{0,1/\delta}(Z)\left\{\frac{\delta^2 Z^2(3+2Z^2)}{C^2_{0,1/\delta}(Z)(1+Z^2)} - \frac{\delta S_{0,1/\delta}(Z)Z^3}{C^3_{0,1/\delta}(Z)\sqrt{1+Z^2}} + \frac{\{1-S^2_{0,1/\delta}(Z)\}}{C^4_{0,1/\delta}(Z)}\right\}\right]+1,$$

$$f_c(\delta) = -E\left[\frac{S_{0,1/\delta}(Z)Z^2}{C_{0,1/\delta}(Z)(1+Z^2)}\left\{Z\sqrt{1+Z^2}+(3+2Z^2)\sinh^{-1}(Z)\right\}\right],$$

and

$$f_d(\delta) = \frac{1}{\delta^2}\left(1+E\left[\frac{Z^2(3+2Z^2)}{(1+Z^2)}\{\sinh^{-1}(Z)\}^2\right]\right),$$

where $Z \sim N(0,1)$.

It is immediately clear that the location and scale parameters are asymptotically independent as are the location and shape (δ) parameters. However, because $\iota_{\sigma\delta} \neq 0$, the scale and shape parameters are not asymptotically independent. In fact, $\mathrm{Corr}(\hat{\sigma},\hat{\delta})$, which does not depend on (μ or) σ asymptotically, equals

$$-\frac{\iota_{\sigma\delta}}{\sqrt{\iota_{\sigma\sigma}\iota_{\delta\delta}}} = -\frac{f_c(\delta)}{\sqrt{f_s(\delta)f_d(\delta)}}.$$

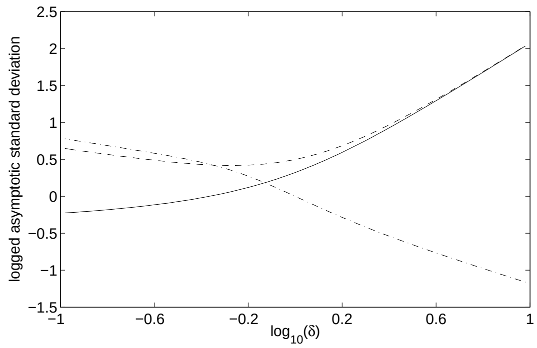


Figure 3: Log_{10} of the asymptotic standard deviation plus $\frac{1}{2}\log_{10} n$ for $\hat{\delta}$ (solid line), $\hat{\mu}$ (dot-dashed line) and $\hat{\sigma}$ (dashed line) in the symmetric case plotted as a function of $\log_{10}\delta$. The solid curve is actually the log standard deviation minus $\log_{10}\delta$ while the other two curves depict the log standard deviation minus $\log_{10}\sigma$.

It is clear that $\iota_{\sigma\delta} < 0$ and hence that the asymptotic correlation between $\hat{\sigma}$ and $\hat{\delta}$ is positive. This correlation can be plotted as a function of δ (solid line in Fig. 2). The correlation is very high for almost all δ. At first, this is disappointing, but it proves to be a standard property of scale/tailweight families of symmetric distributions and reflects the fact that one cannot really tell the difference between changing scale and changing tailweight at all easily in practice.

It is also the case that the asymptotic variance of $\hat{\delta}$ does not depend on σ; it is given by n^{-1} times

$$\frac{\iota_{\sigma\sigma}}{\iota_{\sigma\sigma}\iota_{\delta\delta} - \iota_{\sigma\delta}^2} = \frac{f_s(\delta)}{f_s(\delta)f_d(\delta) - f_c^2(\delta)}.$$

The logged *relative* asymptotic standard deviation (plus $\frac{1}{2}\log_{10} n$) is plotted as the solid curve in Fig. 3; it is necessarily rather large. (See Section 4.3 for comments on the practical effect of this.) While the location parameter μ is in the happy position of being estimated asymptotically independently of σ and δ, the asymptotic variances of the estimates of each are of the form $n^{-1}\sigma^2 h_i(\delta)$ where $i = \mu, \sigma$. So, reasonably enough, both standard deviations increase in direct proportion to the value of σ. We have that

$$h_\mu(\delta) = \frac{1}{f_m(\delta)} \quad \text{and} \quad h_\sigma(\delta) = \frac{f_d(\delta)}{f_s(\delta)f_d(\delta) - f_c^2(\delta)}.$$

These two functions are also shown, square rooted and logged, in Fig. 3 as dotted and dashed lines, respectively.

4.2. *Reparametrisation*

In principle, at least, it is possible to provide an orthogonal parametrisation of the form $(\mu, \sigma F(\delta), \delta)$. Since the correlation between $\hat{\sigma} F(\hat{\delta})$ and $\hat{\delta}$ is proportional to $(\log F)'(\delta) f_s(\delta) - f_c(\delta)$, this would be achieved by setting $(\log F)'(\delta) = f_c(\delta)/f_s(\delta)$. Unfortunately, this is insufficiently tractable to provide a workable formula. However, as shown in Fig. 2, the asymptotic correlation between $\hat{\sigma}$ and $\hat{\delta}$, which we are trying to alleviate via reparametrisation, is highest for large δ. This suggests seeking a large δ approximation to the above.

To this end, we find that, for large δ, $f_c(\delta) \simeq -(C+S)/\delta$ and $f_s(\delta) \simeq 1+S$ where

$$C = E\left\{\frac{Z^3 \sinh^{-1}(Z)}{\sqrt{1+Z^2}}\right\} \quad \text{and} \quad S = E\left[\frac{Z^2\{\sinh^{-1}(Z)\}^2(3+2Z^2)}{1+Z^2}\right].$$

Numerically, we find that $C \approx 1$, at least correct to 7 decimal places (we have been unable to prove exact equality to unity). We then find that $F(\delta) \simeq \delta^{-1}$, so suggesting a simple reparametrisation in which σ is replaced by $\sigma_\delta \equiv \sigma/\delta$. The asymptotic correlation between $\hat{\sigma}_\delta$ and $\hat{\delta}$ is

$$-\frac{\delta f_c(\delta) + f_s(\delta)}{\sqrt{f_s(\delta)\{\delta^2 f_d(\delta) + 2\delta f_c(\delta) + f_s(\delta)\}}}.$$

This is plotted as the dashed line in Fig. 2. It is clear that we have achieved a general lowering of the asymptotic correlation to less extreme values. We have not achieved the very small correlation for large δ that might have been expected because the variance of $\hat{\sigma}_\delta$ tends to zero alongside the covariance for large δ. However, the reduction in correlation that we have achieved proves to make a considerable difference in practice.

4.3. *Practical implementation in the general case*

We employ (and recommend) the reparametrisation just derived in fitting the full four-parameter sinh-arcsinh distribution (as well as its symmetric subfamily) to data, i.e. utilising $\{\mu, \sigma_\delta, \epsilon, \delta\}$ and then setting $\hat{\sigma} = \hat{\sigma}_\delta \hat{\delta}$. This solved severe numerical problems encountered in the original parametrisation when $\delta > 1$. We made use of the Nelder and Mead (1965) simplex

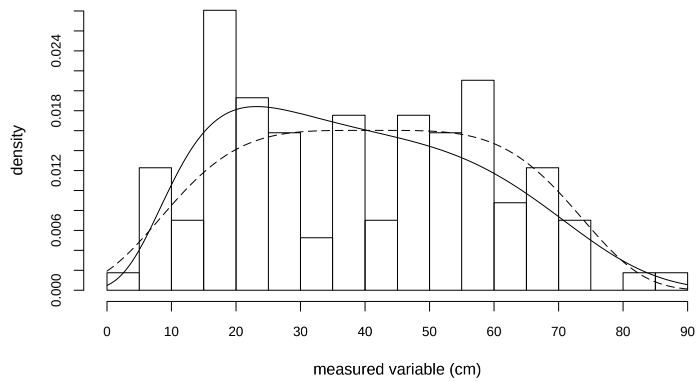


Figure 4: Histogram of the snow depth data together with the fitted densities for family (2) (solid line) and its symmetric subfamily (dashed line).

algorithm to perform maximisation of the log-likelihood. Using this direct search approach, it proves helpful to optimise over $\mu/\sqrt{1+\mu^2} \in (-1,1)$, $\sigma_\delta/(1+\sigma_\delta) \in (0,1)$, $\epsilon/\sqrt{1+\epsilon^2} \in (-1,1)$ and $\delta/(1+\delta) \in (0,1)$ and then back-transform. In practice, we have not come across examples of multiple maxima occurring on the log-likelihood surface. However, as is generally the case when using numerical optimisation techniques, it is advisable to try a range of different starting values in an attempt to ensure that the global maximum is identified. We find that each of μ, σ_δ and ϵ is estimated well but large δ-values are not estimated so precisely. The log-likelihood surface remains flat when δ is large, corresponding to the large asymptotic variance of $\hat{\delta}$ shown in Fig. 3.

4.4. Example

In order to briefly illustrate the modelling flexibility of family (2), we present an analysis of $n = 114$ measurements of the depth of snow (in cm) taken on an ice floe in the eastern Asmundsen Sea, Antarctica, in March 2003. See Banks (2006, Chapter 6) for details, noting that these data pertain to "Floe 2" and Banks's analysis included preliminary use of a symmetric sinh-arcsinh distribution. A histogram of the data appears in Fig. 4. Results for the maximum likelihood fits of family (2) and its normal ($\delta = 1, \epsilon = 0$), normal-tailed ($\delta = 1$) and symmetric ($\epsilon = 0$) submodels are given in Table

Table 1: Parameter estimates for the fits to the snow depth data of, reading from right to left, family (2) and its symmetric ($\epsilon = 0$), normal-tailed ($\delta = 1$) and normal ($\delta = 1, \epsilon = 0$) submodels. The maximised log–likelihood (l_{max}), AIC and BIC values, and p-value for the chi–squared goodness-of-fit test, are included as fit diagnostics.

	Model			
Parameter	Normal	Normal tails	Symmetric	Family (2)
μ	39.24	24.66	40.49	-52.91
σ	20.20	17.88	349139.3	34.27
δ	(1)	(1)	14028.5	3.99
ϵ	(0)	-0.52	(0)	-6.75
l_{max}	-504.39	-502.50	-497.72	-494.98
AIC	1012.78	1011.00	1001.44	997.96
BIC	1018.25	1019.21	1009.65	1008.90
p-value	0.016	0.004	0.228	0.213

1. All three likelihood-based diagnostics in Table 1 indicate that the fit for the full family, with its lighter than normal tails ($\delta > 1$) and positive asymmetry ($\epsilon < 0$), is best, followed by that for its symmetric subfamily. The p-values of likelihood-ratio tests for normality, normal tails and symmetry, calculated using the usual asymptotic chi-squared approximation, are 0.000, 0.000 and 0.019, respectively. So, the fit for the full family appears to offer a significantly better fit than any of its three submodels. Table 1 also contains the p-values for the chi-squared goodness-of-fit test performed using the class intervals of the histogram shown, some of which were combined to obtain expected values of at least 5. The p-values support the adequacy of the two best fits and rule out the normal and normal-tailed submodels. The densities of the two best fits are superimposed on the histogram in Fig. 4. Comparing them with the histogram, there is perhaps some indication of multimodality in the data. However, this could be an artifact of the binning used and the rounding of the data to the nearest whole cm during measurement. It would certainly be difficult to conceive of a better unimodal fit to the data.

In the next two sections we present a substantial practical investigation of the use of sinh-arcsinh distributions in testing normality and symmetry.

5. Testing normality

The central position of the normal distribution within family (2) allows testing of normality within the family via standard likelihood ratio tests (LRTs). However, since family (2) is sufficiently broad to, at some level, provide an approximation to any unimodal distribution, we propose that sinh-arcsinh-based LRTs of normality also be used as general purpose tests of normality. To that end, in this section, we explore the size and power of sinh-arcsinh-based LRTs of normality both within and beyond the sinh-arcsinh family of distributions.

5.1. Testing normality against symmetric alternatives I: size

It is probably most usual to test for normality in a situation where one is willing to assume symmetry of the distribution of interest. In that case, the appropriate LRT is a statistic of the form $L = -2\log(\ell_0/\ell_1)$ where ℓ_0 represents the maximum of the log-likelihood function for an assumed normal distribution and ℓ_1 the maximum of the log-likelihood function assuming that the sample was drawn from a *symmetric* sinh-arcsinh distribution i.e. $\sigma^{-1}f_{0,\delta}(\sigma^{-1}(x-\mu))$ where $f_{\epsilon,\delta}(x)$ is given by (2). ℓ_1 has to be calculated numerically. This being a regular problem, the asymptotic distribution of L is, of course, χ^2_1 (the single degree of freedom being associated with setting $\delta = 1$ in the symmetric subfamily to achieve normality). For testing normality against asymmetric alternatives, see Section 5.3.

We investigated the distribution of 10,000 values of L based on samples generated from the standard normal distribution. The χ^2 approximation to the sampling distribution of L is, as expected, poor for small sample sizes, rapidly improving with increasing n such that for $n \geq 50$ it would appear to provide a very good approximation to the true sampling distribution. However, the tails of the χ^2 approximation to the simulated sampling distribution are in reasonably close agreement even for small n. Indeed, in addition to being adequate for large n, the asymptotic critical values for the test are closest to the simulated critical values in certain cases associated with small n! Overall, we consider it reasonable, as well as simplest, to use the critical values of the asymptotic χ^2_1 distribution when analysing samples of any size. This is further vindicated in the studies that follow in the next section.

5.2. Testing normality against symmetric alternatives II: power

There exists a well-established literature addressing the problem of testing univariate data for normality. Renewed recent interest in this inferential problem can be found in the papers of Zhang and Wu (2005) and Thadewald and Büning (2007), amongst others. In the light of the findings presented in those two papers, we conducted a simulation study designed to compare the performance of the LRT of normality with those of the following seven competitive tests for a nominal significance level of 5% (the description of each test is preceded by the abbreviation we will use when referring to it):

JB. The (one-sided) test of Jarque and Bera (1980), the test statistic of which is a function of the coefficients of skewness and kurtosis. We used the corrected critical values for this test presented in Table 2 of Thadewald and Büning (2007).

D. The (two-sided) test of D'Agostino (1971, 1972). Up to a constant, the test statistic is the ratio of Downton's (1966) linear estimator of the standard deviation to the sample standard deviation. The critical values for this test are given in D'Agostino's papers.

AD. The (one-sided) empirical distribution function (EDF)-based test of Anderson and Darling (1952). We used the corrected critical values for this test presented under the name CMW in Table 2 of Thadewald and Büning (2007). (Those authors do not seem to realise that their CMW statistic is in fact the A^2 statistic of Anderson and Darling.)

CM. The (one-sided) Cramér-von Mises EDF-based test with statistic identified as CM in Thadewald and Büning (2007). We used the corrected critical values given in their Table 2.

SW. The (one-sided) test of Shapiro and Wilk (1965). We used Algorithm AS R94 of Royston (1995) to compute the test statistic and its p-value.

ZA and ZC. The (one-sided) nonparametric likelihood-ratio-based tests with test statistics Z_A and Z_C of Zhang and Wu (2005). We used the corrected critical values for these tests given in Tables 1 and 2, respectively, of that paper.

Our power simulations concern two sets of alternative distributions; here is the first. For each combination of $n = 10, 20, 50, 100, 200$ and $\delta = 0.2, 0.4, 0.6, 1, 2, 10$, $10,000$ random samples of size n were simulated from the symmetric subfamily of (2) with $\epsilon = 0$. All eight tests were then applied to each simulated sample. (Setting $\mu = 0$ and $\sigma = 1$ throughout is appropriate since all tests are location/scale invariant.) In Fig. 5, the proportion of the 10,000 samples for which the null hypothesis of normality was rejected in a nominally 5% test is plotted against $\lambda = \delta/(1+\delta)$. Remember that $\delta = 1$, i.e. $\lambda = 0.5$, corresponds to a normal distribution, so this figure provides information concerning both the true size and power of the different tests.

Considering the content of Fig. 5, we can draw the following conclusions. Firstly, all seven rival tests maintain the nominal significance level of 5% very closely. So does the LRT, in general, although it is slightly conservative for $n = 10$ (size $\simeq 0.03$) and slightly liberal for $n = 50$ (size $\simeq 0.08$). Secondly, the LRT and D tests have the best overall power characteristics; the LRT and CM tests are the most powerful against alternatives with $0 < \lambda < 0.45$ (i.e. distributions with tails that are far heavier than normal), and, for $n \geq 20$, the LRT is the most powerful against alternatives with lighter than normal tails ($\lambda > 0.5$). D has second best power for these latter alternatives; it also has high power for λ-values in the region of 0.6–0.8, as does JB. However, for alternatives with lighter than normal tails, JB has the worst power signature. Indeed, even for samples as large as 100, its power generally lies below the nominal significance level. As is to be expected, the power of all eight tests generally increases with n for fixed δ. The increase in power with n is particularly noteworthy against the alternatives with lighter than normal tails ($\lambda > 0.5$). Note that these results provide interesting further information concerning the relative performance of the competing test to complement the findings of Zhang and Wu (2005) and Thadewald and Büning (2007), particularly concerning the ZA and ZC tests in the former and the Jarque-Bera test in the latter.

Of course, testing *within* the symmetric sinh-arscinh family, as just considered, is the situation for which our LRT was designed and for which it must be expected to be particularly strong as, gratifyingly, it proved. The second set (of three) alternative distributions are not members of class (2). These are: (i) the very heavy-tailed t distribution on two degrees of freedom (t_2); (ii) the fairly-heavy-tailed logistic distribution; and (iii) a light-tailed distribution due to M.L. Tiku with density $16(1+x^2/4)^2\phi(x)/27$ (e.g. Tiku, Islam and Selcuk, 2001). The results of our power simulations against these alternative distributions are given in Fig. 6.

From Fig. 6(a), corresponding to the t_2 alternative distribution, it can be seen that all eight tests are relatively powerful. For larger values of n there is little difference in their powers; for $n = 20, 50$, JB and D tend to dominate.

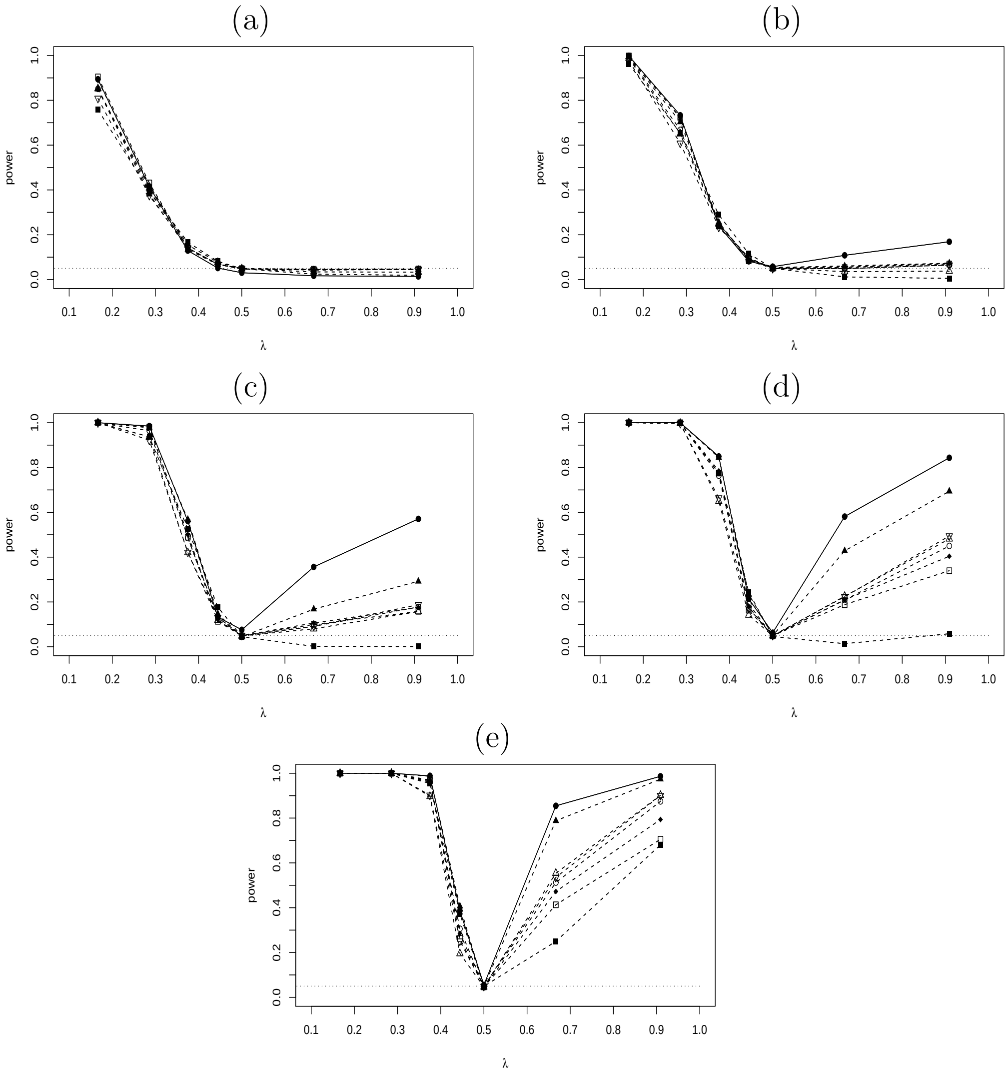


Figure 5: The proportion of samples for which the null hypothesis of normality was rejected in a nominally 5% test, plotted against $\lambda = \delta/(1+\delta)$. The proportions were calculated using 10,000 random samples from alternative sinh-arcsinh distributions with $\epsilon = 0$, $\delta = 0.2, 0.4, 0.6, 1, 2, 10$ and sample sizes of: (a) $n = 10$; (b) $n = 20$; (c) $n = 50$; (d) $n = 100$; (e) $n = 200$. The solid lines connect the results of the LRT, and the dashed lines those for the other seven tests: JB (solid square); D (solid triangle); AD (solid diamond); CM (open square); SW (open circle); ZA (open triangle); ZC (open inverted triangle). The dotted line is at the nominal level of 0.05.

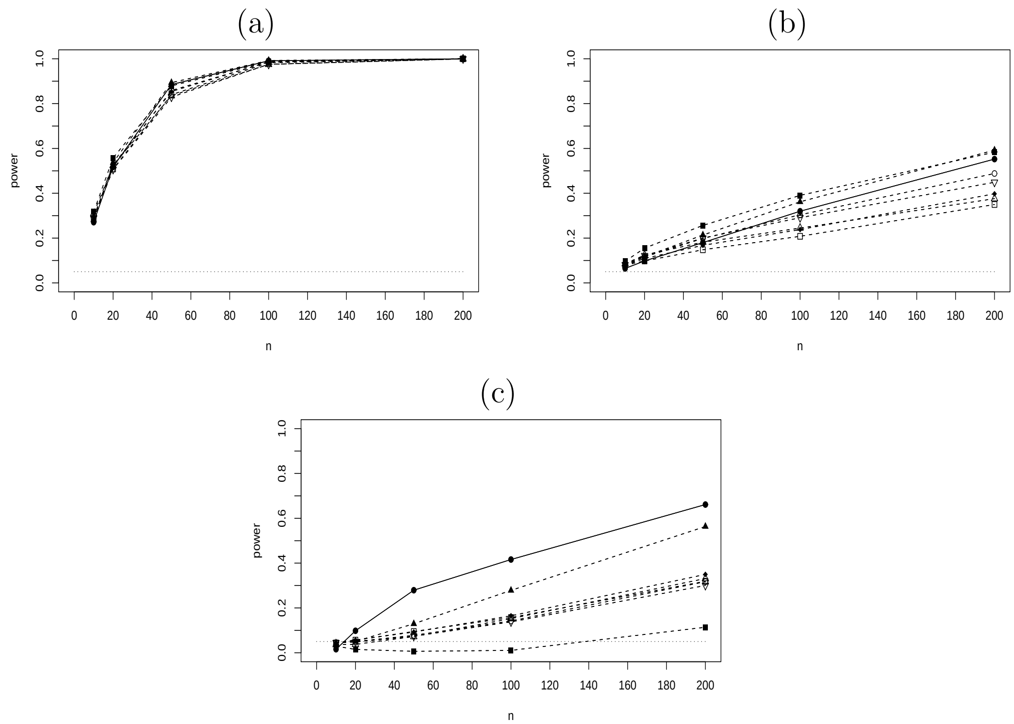


Figure 6: The proportion of samples for which the null hypothesis of normality was rejected in a nominally 5% test plotted against n. The proportions were calculated using 10,000 random samples of size $n = 10, 20, 50, 100, 200$ from the: (a) t_2; (b) logistic; and (c) Tiku short-tailed distributions. The solid lines connect the results of the LRT, and the dashed lines those for the other seven tests: JB (solid square); D (solid triangle); AD (solid diamond); CM (open square); SW (open circle); ZA (open triangle); ZC (open inverted triangle). The dotted line is at the nominal level of 0.05.

Fig. 6(b) portrays the equivalent results for the logistic distribution. Clearly, none of the tests is very powerful against this alternative. The JB test has the best overall performance. D also performs relatively well, particularly for larger values of n. The LRT performs relatively poorly for samples of size $n \leq 50$ but its relative performance improves with increasing n. The performance of CM is worst overall. Finally, the results for Tiku's short-tailed distribution are displayed in Fig. 6(c). Again, none of the tests is particularly powerful. Indeed, for samples of size 10 and 20 the power lies below, and bobs around, respectively, the nominal level of the tests. LRT is clearly the most powerful, followed by D. The powers of five of the other six tests are very similar, with the JB test being very poor.

Overall, the LRT seems very competitive in most symmetric situations with the best of existing tests which would appear to be D'Agostino's test D.

5.3. Testing normality against asymmetric alternatives

If one is not willing to assume symmetry, testing for normality can still be accomplished from within the full four-parameter sinh-arcsinh family. The appropriate LRT now compares the maximised log-likelihood function for an assumed normal distribution with the maximum of the log-likelihood assuming the sample was drawn from $\sigma^{-1} f_{\epsilon,\delta}(\sigma^{-1}(x-\mu))$, the asymptotic distribution of the LRT statistic now being χ^2_2. Simulations from the normal distribution yielded results in keeping with the test's ability, as in the symmetric case, to maintain its nominal significance level using its asymptotic sampling distribution.

Because symmetry is no longer being assumed in constructing the test statistic, the power of the 'asymmetric LRT' is necessarily a little lower than that of the previous 'symmetric LRT' when normality is tested within a truly symmetric situation. The effect is quite small and the overall performance of the asymmetric LRT remains excellent. For example, if the powers of the symmetric LRT in Fig. 5 were replaced by those of the asymmetric LRT: (a) for large δ (light tails), the previous superiority of the symmetric LRT is reduced to a performance essentially on a par with the second-based method, namely D; (b) for very small δ (the heaviest tails), the LRT remains almost as good as the (otherwise best) CM test; (c) for other $\delta < 1$ (tails heavier than normal), the LRT continues to have a quality of performance which is in the middle of the pack of tests considered. These observations are also reflected for the non-sinh-arcsinh symmetric alternatives of Fig. 6 in accordance with their relative tail weights.

Fig. 7 shows simulated powers for the asymmetric LRT and the same set of seven competing tests for normality within a set of asymmetric distributions, namely sinh-arcsinh densities with $\epsilon = 1$. Overall, the LRT is best. Indeed, the ordering of the power performances of the tests is by and large the same as in Fig. 7's symmetric counterpart (Fig. 5) with two notable exceptions: (i) the D test, which was previously competitive with the LRT, is very badly affected by the presence of asymmetry; and (ii) the LRT maintains its 'first place' even for alternatives with slightly heavier tails than those of the normal. The performance of D is particularly poor for a middle range of values of δ including fairly heavy tails when n is small, normal tails when n is small and moderate, and fairly light (but not the lightest) tails even when $n = 200$. We note also that, for small n, the combination of non-light tails ($\delta \leq 1$) and skewness makes for a greater disparity in power performance between the best and the poorest tests. In further experiments with a number of asymmetric alternatives outside the sinh-arcsinh class, the relative performances of tests described here — including the mostly leading performance of the LRT and the many poor performances of the D test — were upheld, again in accordance with their levels of tail weight.

5.4. Conclusion

Taking both symmetric and asymmetric alternatives into account, the LRT seems to be the best of the options considered here (and its competitors have been chosen because of claims of leading performance elsewhere).

6. Testing symmetry

We can also test for symmetry (about an unknown centre) by employing an LRT within the full sinh-arcsinh family of the null hypothesis that $\epsilon = 0$. The asymptotic null distribution of the test, which we shall use, is, again, χ^2_1.

We will compare the size and power performance of our LRT of symmetry (again, for a nominal significance level of 0.05) with those of two other general tests of symmetry. These particular tests were chosen because they were found to perform well in extensive simulation comparisons reported in Cabilio and Masaro (1996). They are:

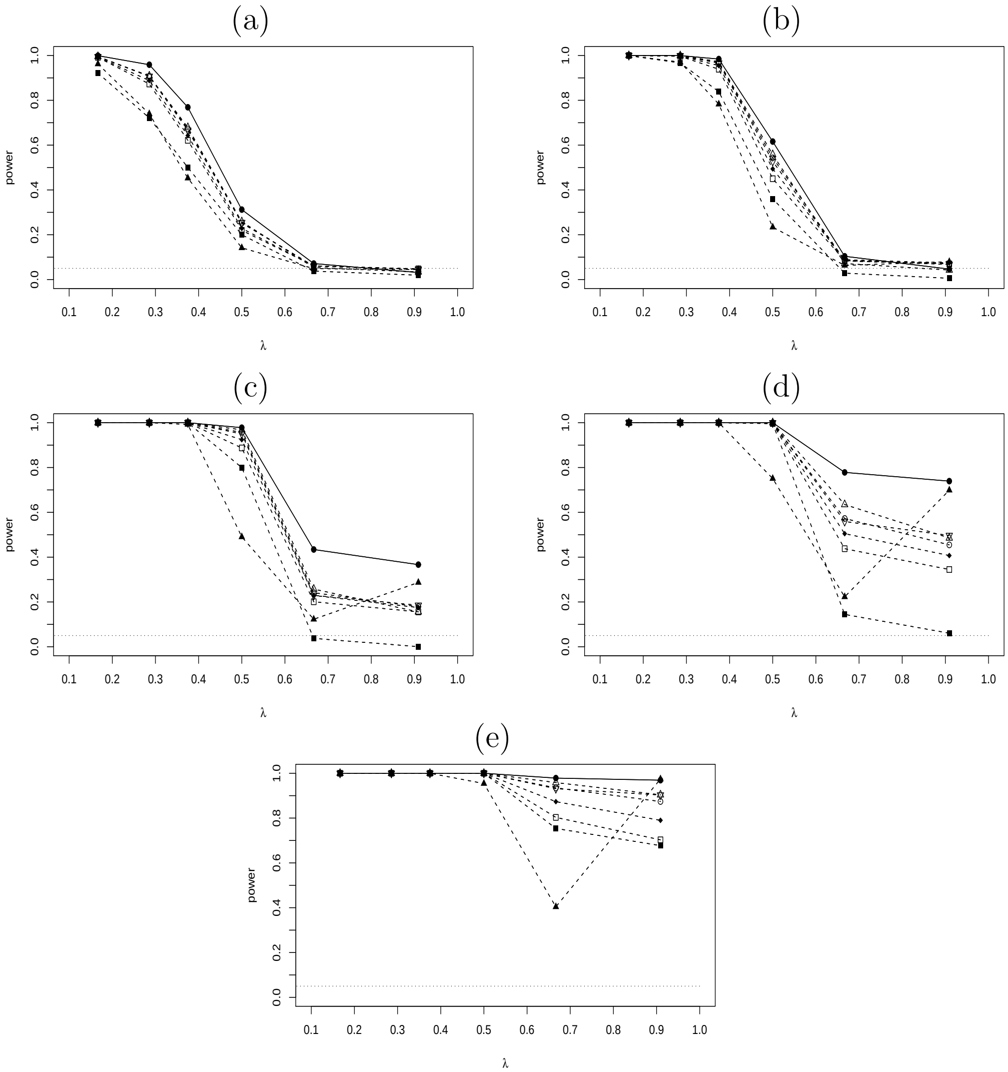


Figure 7: The proportion of samples for which the null hypothesis of normality was rejected in a nominally 5% test, plotted against $\lambda = \delta/(1 + \delta)$. The proportions were calculated using 10,000 random samples from asymmetric sinh-arcsinh distributions with $\epsilon = 1$, $\delta = 0.2, 0.4, 0.6, 1, 2, 10$ and sample sizes of: (a) $n = 10$; (b) $n = 20$; (c) $n = 50$; (d) $n = 100$; (e) $n = 200$. The solid lines connect the results of the LRT, and the dashed lines those for the other seven tests: JB (solid square); D (solid triangle); AD (solid diamond); CM (open square); SW (open circle); ZA (open triangle); ZC (open inverted triangle). The dotted line is at the nominal level of 0.05.

SK. The test of Cabilio and Masaro (1996), the test statistic of which is the simple function $S_K = \sqrt{n}(\bar{X} - m)/s$ where $\bar{X}$, m and s denote the sample mean, median and standard deviation (with divisor n), respectively. We used the critical values for this test presented in Table 1 of Cabilio and Masaro (1996) which were calibrated against the normal distribution.

TN. The second test statistic was the more involved one of Boos (1982) which is based on the Hodges-Lehmann estimator. We used the critical values for this test presented in Table 1 of Boos (1982) which were calibrated against the logistic distribution.

6.1. Testing symmetry I: size

In Fig. 8, simulated values of the size of each of the LRT, SK and TN tests are presented for a variety of symmetric members of the sinh-arcsinh family. It can be seen that the LRT is by far the best test in terms of its overall ability to maintain the nominal significance level. SK tends to be very liberal when the distribution is either heavy- or light-tailed. TN is extremely liberal when the distribution is heavy-tailed and marginally liberal when it is light-tailed. When the underlying distribution is normal ($\lambda = 0.5$), all three tests maintain the nominal level increasingly well with increasing n.

We also computed the size of the tests of symmetry for data simulated from the t_2, logistic and Tiku distributions used above as symmetric alternatives to the normal distribution in Fig. 6. Summarising our results: (a) for the heavy-tailed t_2 distribution, SK holds the nominal level best, whilst the LRT and especially TN are markedly liberal; (b) for the logistic distribution, TN holds the nominal level well, SK is marginally conservative and the LRT marginally liberal; and (c) for Tiku's light-tailed distribution, all three tests hold the nominal level pretty well, with the LRT and TN holding it best, SK being rather liberal. These results collectively chime with earlier observations: Cabilio and Masaro (1996) observed that the size of their test, SK, is inflated when the underlying distribution is uniform or Cauchy, while both Boos (1982) and Cabilio and Masaro (1996) noted that TN can be extremely sensitive to heavy-tailed distributions, tending to mistakenly confuse such tails with asymmetry.

6.2. Testing symmetry II: power

We start this section by investigating the power of the LRT, SK and TN tests against alternative, asymmetric, distributions taken from the sinh-arcsinh

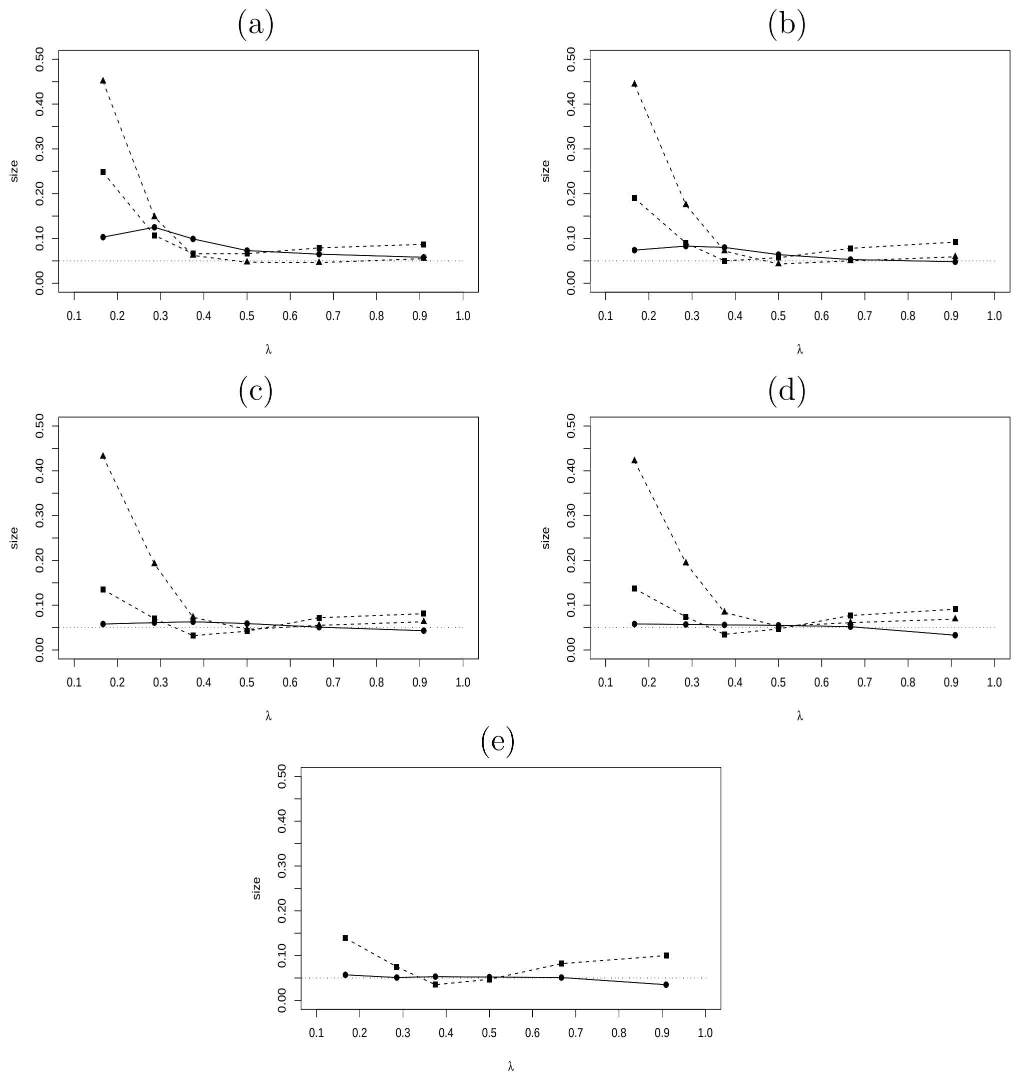


Figure 8: The proportion of samples for which the null hypothesis of symmetry was rejected in a nominally 5% test, plotted against $\lambda = \delta/(1+\delta)$. The proportions were calculated using 10,000 random samples from symmetric sinh-arcsinh distributions ($\epsilon = 0$), with $\delta = 0.2, 0.4, 0.6, 1, 2, 10$ and sample sizes of: (a) $n = 10$; (b) $n = 20$; (c) $n = 50$; (d) $n = 100$; (e) $n = 200$. The solid lines connect the results of the LRT, and the dashed lines those for the other two tests: SK (square); TN (triangle). The dotted line is at the nominal level of 0.05. The results for TN are missing from panel (e) as the computational burden (in terms of storage) proved too much for our programs to handle.

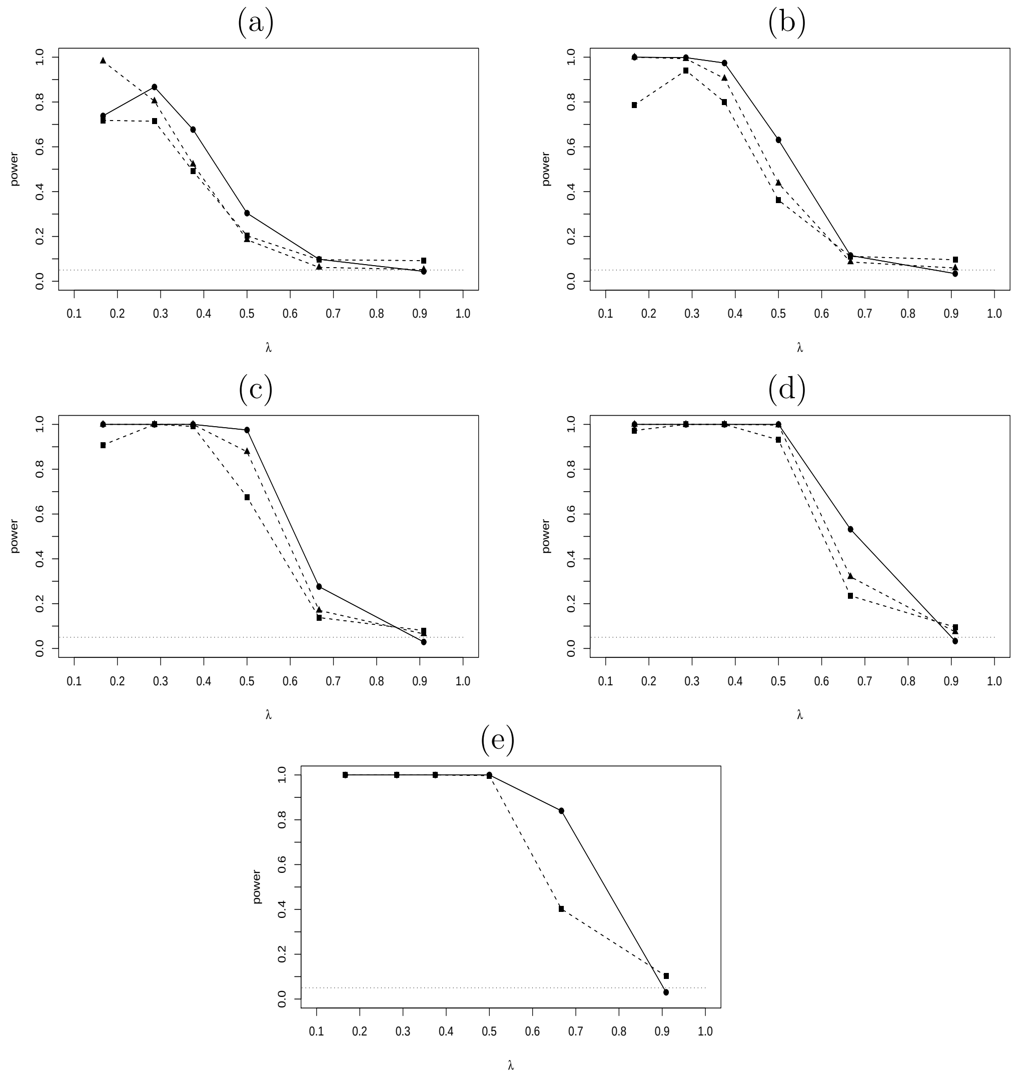


Figure 9: The proportion of samples for which the null hypothesis of symmetry was rejected in a nominally 5% test, plotted against $\lambda = \delta/(1+\delta)$. The proportions were calculated using 10,000 random samples from asymmetric sinh-arcsinh distributions with $\epsilon = 1$, $\delta = 0.2, 0.4, 0.6, 1, 2, 10$ and sample sizes of: (a) $n = 10$; (b) $n = 20$; (c) $n = 50$; (d) $n = 100$; (e) $n = 200$. The solid lines connect the results of the LRT, and the dashed lines those for the other two tests: SK (square); TN (triangle). The dotted line is at the nominal level of 0.05. The results for TN are again missing from panel (e).

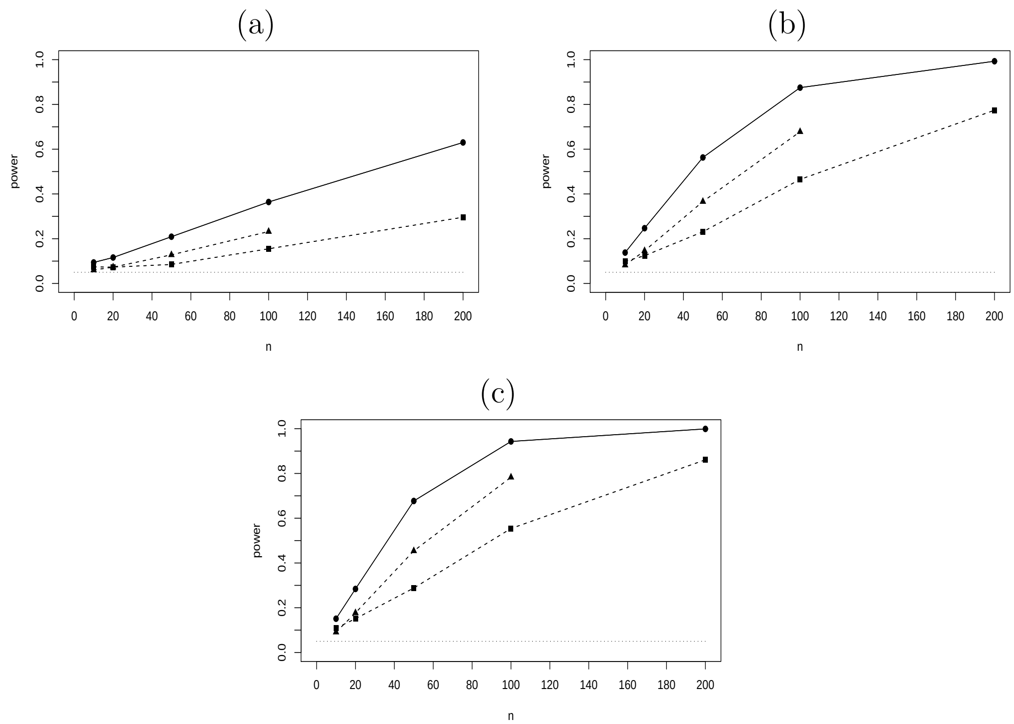


Figure 10: The proportion of samples for which the null hypothesis of symmetry was rejected in a nominally 5% test plotted against n. The proportions were calculated using 10,000 random samples of size $n = 10, 20, 50, 100, 200$ from the log F distribution with: (a) 4 and 2; (b) 16 and 2; and (c) 64 and 2 degrees of freedom. The solid lines connect the results of the LRT, and the dashed lines those for the other two tests: SK (square); TN (triangle). The dotted line is at the nominal level of 0.05. The results for TN are again missing for $n = 200$.

family with $\epsilon = 1$. See Fig. 9 for results. It can be seen that, overall, the LRT is the most powerful of the tests. For the smallest-sized sample from the heaviest-tailed distribution, TN has higher power but it should be recalled that this test is unable to maintain the nominal level for heavy-tailed symmetric distributions within family (2). The LRT also performs quite poorly against extremely light-tailed alternatives, but all three tests have very low power in such cases.

We also made extensive investigations of the power of these three tests to detect asymmetry outside the sinh-arcsinh family. To this end, we investigated — in order of increasing tail weight — the extreme value distribution and a range of skew-normal (Azzalini, 1985), log F (e.g. Baghdachi and Balakrishnan, 2008) and skew t (Jones and Faddy, 2003) distributions. Results for each distribution were broadly similar: the LRT was most powerful, followed by the TN test and then the SK test. Fig. 10 shows these results for a range of log F distributions, those for extreme value and skew-normal alternatives looking very similar. Only in the case of the heavy-tailed skew t distributions was the power performance of the LRT closely matched by that of TN.

6.3. Conclusion

The sinh-arcsinh-based LRT clearly outperforms two omnibus tests for symmetry that we chose for comparison as being 'state-of-the-art', both (unsurprisingly) within the sinh-arcsinh family but also (much less necessarily) in a wide range of situations outside the sinh-arcsinh family too.

7. The multivariate case

Multivariate extensions of the univariate distributions arise naturally and immediately by transforming the univariate marginals of a standardised (but correlated) multivariate normal distribution. By so doing, we choose to model skewness and/or tailweight variations directly on the original scales of the variables. So, in d dimensions, let R be a correlation matrix and define the vector X by $Z_i = S_{\epsilon_i,\delta_i}(X_i)$, $i = 1, ..., d$, where $Z \sim N_d(0, R)$, so that

$$f_{\epsilon,\delta}(x) = \frac{1}{\sqrt{(2\pi)^d |R|}} \prod_{i=1}^{d} \left\{ \frac{\delta_i C_{\epsilon_i,\delta_i}(x_i)}{\sqrt{1+x_i^2}} \right\} \exp\left(-\frac{1}{2} S_{\epsilon,\delta}(x)' R^{-1} S_{\epsilon,\delta}(x) \right). \quad (7)$$

In an abuse of notation, the vector z has been written $S_{\epsilon,\delta}(x)$.

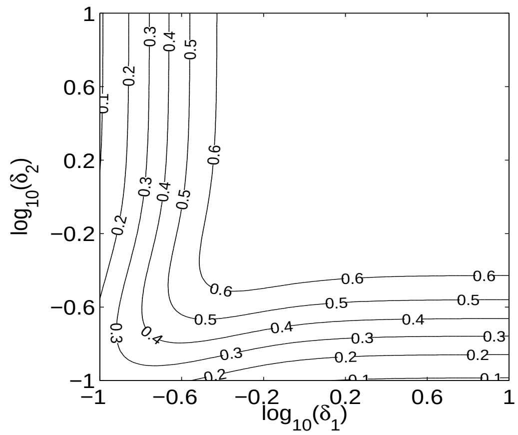


Figure 11: The correlation between X and Y in the symmetric marginals case, plotted as a function of $\log_{10}\delta_1$ and $\log_{10}\delta_2$: here, $\rho = 0.7$.

The univariate marginals of this distribution are sinh-arcsinh distributions by construction. If z is partitioned into (z_1, z_2) and x, X and R are partitioned conformably, $X_1|x_2$ is the distribution of $S^{-1}_{\epsilon_1,\delta_1}(Z_1)|z_2 = S_{\epsilon_2,\delta_2}(x_2)$ where $Z_1|z_2 \sim N(R_{12}(R_{22})^{-1}z_2, R_{11} - R_{12}^T(R_{22})^{-1}R_{12})$. Notice that now the transformation is applied to an unstandardized normal distribution, which means that conditional distributions are members of a wider (and not very tractable) family of distributions that will not be pursued further. The maintenance of unimodality in univariate distributions augurs well for the unimodality of the multivariate case, and we have no counterexamples from our limited experience with these distributions. All moments of the distribution, of course, exist.

The covariance between any two elements of X is not generally tractable. It is, however, plotted in the symmetric marginals ($\epsilon_1 = \epsilon_2 = 0$) case in Fig. 11 as a function of δ_1, the parameter in the x-direction, and δ_2, the parameter in the y-direction, for $\rho = 0.7$. A number of properties of the multivariate distribution are illustrated by this plot. First, the sign of $\rho_{12} = \mathrm{Corr}(S^{-1}_{\epsilon_1,\delta_1}(Z_1), S^{-1}_{\epsilon_2,\delta_2}(Z_2))$ is the same as the sign of ρ for all $\epsilon_1, \delta_1, \epsilon_2, \delta_2$. This follows because of the positive (negative) quadrant dependence of the bivariate normal distribution with $\rho >\ (<)\ 0$ and the strictly increasing nature of the marginal transformations (see, for example, results in Joe, 1997). Second, $|\rho_{12}| \leq |\rho|$. This inequality can be found in literature stemming from Gebelein (1941), see, for example, Koyak (1987) and references therein.

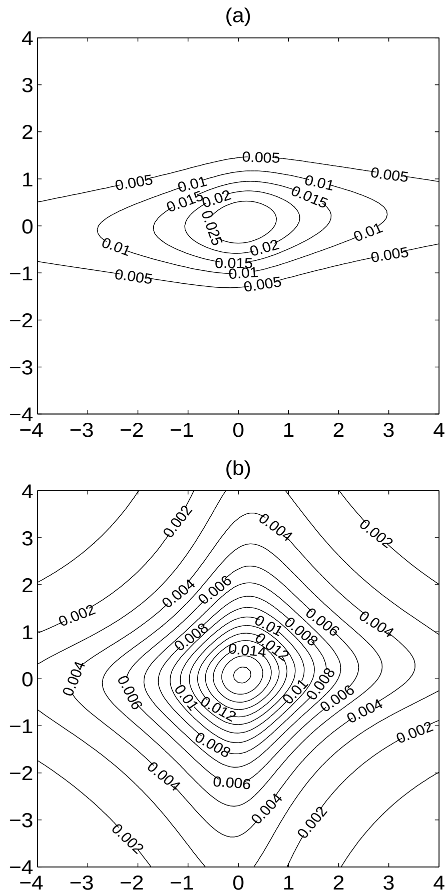


Figure 12: The bivariate sinh-arcsinh density with $\rho = 0.7$ and (a) $\delta_1 = 0.135$, $\delta_2 = 1$, (b) $\delta_1 = \delta_2 = 0.27$.

Concentrating on the symmetric marginals case as in Fig. 11, we note that: (i) the value of the correlation is $\rho = 0.7$ only at the point $\delta_1 = \delta_2 = 1$ and is lower elsewhere; (ii) the value of the correlation remains close to $\rho = 0.7$ for all $\delta_1, \delta_2 \geq 1$ i.e. lighter tails; and (iii) the absolute value of the correlation decreases as one or both tails get heavier. In particular, this makes sense in the case $\delta_1 < 1, \delta_2 = 1$ where the density is spread much more in the x-direction than in the y-direction. For an illustration of this, see the density plotted in Fig. 12(a); (iii) this last effect is reduced somewhat if both tails get heavier. The density for $\delta_1 = \delta_2 = 0.27$ is plotted in Fig. 12(b).

It may also be of interest to consider the local dependence function defined as $\gamma(x,y) = \partial^2 \log f_{\epsilon,\delta}(x,y)/\partial x \partial y$. This was introduced as a continuous analogue of the local log odds ratio by Holland and Wang (1987) and alternatively justified as a localised correlation coefficient by Jones (1996). Either directly, or by noting that $\gamma(x,y) = \rho/(1-\rho^2)$ for the normal distribution and that γ transforms in the same way as density functions, in our case we have

$$\gamma(x,y) = \frac{\rho}{1-\rho^2} \frac{\delta_1 C_{\epsilon_1,\delta_1}(x)}{\sqrt{1+x^2}} \frac{\delta_2 C_{\epsilon_2,\delta_2}(y)}{\sqrt{1+y^2}}.$$

Note that $\gamma(x,y)$ has the same sign as ρ for all x, y. The way that ρ affects only the overall size of local dependence and is otherwise divorced from the influence of the other parameters is a nice feature of this transformation approach. Also, x- and y-dependence are separated out, so we consider, say, $L_{\epsilon,\delta}(x) \equiv \delta C_{\epsilon,\delta}(x)/\sqrt{1+x^2}$ only. In the symmetric case, $L_{0,\delta}(0) = \delta$ and it can readily be shown that $L_{0,\delta}$ symmetrically decreases (increases) towards zero (infinity) if $\delta <\ (>)\ 1$. In the general case, $L_{\epsilon,\delta}(0) = \delta \cosh \epsilon$, while both 'tails' of $L_{\epsilon,\delta}$ still go to zero (infinity) if $\delta <\ (>)\ 1$.

8. Options and extensions

Readers may be discomfited by some of the specific choices that have been made in this paper so far. In particular, a question that we have been asked more than once is: "why is the sinh function at the heart of this methodology rather than some other monotone function?" Second, it is clear that the normal distribution is only one of a number of possible choices for the 'central distribution' in this approach. And there is a third, perhaps less obvious, question that concerns the way in which skewness has been introduced into our model. In this section, we address each of these issues in turn.

8.1. Which transformation function?

Introduce a one-to-one onto function $H : \mathbb{R} \to \mathbb{R}$ with $H(0) = 0$ and write $h(x) = H'(x) > 0\ \forall x$. Consider transformations of the form

$$Z = T_{\epsilon,\delta}(X_{\epsilon,\delta}) \equiv H\{\epsilon + \delta H^{-1}(X_{\epsilon,\delta})\}. \tag{8}$$

This formulation, involving both H and H^{-1}, is key to setting the normal distribution at the centre of the transformed family and allowing both heavier and lighter tails. This is most easily seen when $\epsilon = 0$: for small δ, $T(X) \sim$

$\delta h(0)H^{-1}(x)$, and for large δ, $T(X/\delta) \sim H(x/h'(0))$, division of X by δ being the 'suitable scaling' employed in Section 3.3.

Anticipating the main consideration of Section 8.2, replace the normal density as the object of transformation by a generic simple symmetric distribution with distribution function G. Apply the 'H-arcH' transformation to obtain the transformed family of distributions with distribution function

$$\mathcal{G}(x) \equiv G(H(\epsilon + \delta H^{-1}(x))). \tag{9}$$

The following result concerns the conditions required on H so that ϵ and δ act as skewness and kurtosis parameters in the sense of van Zwet (1964).

Theorem. The parameters ϵ (for fixed δ) and δ (for $\epsilon = 0$) in (9) act as a pair of skewness and kurtosis parameters in the sense of van Zwet (1964) if and only if $\log h$ is either a convex or a concave function of x.

Proof. Let $\mathcal{G}_i$ denote $\mathcal{G}$ when the parameters are $\epsilon_i, \delta_i, i = 1, 2$. Then

$$\mathcal{G}_2^{-1}(\mathcal{G}_1(x)) = H(c + dH^{-1}(x))$$

(independently of G) where $c = (\epsilon_1 - \epsilon_2)/\delta_2$ and $d = \delta_1/\delta_2$. Then,

$$t_{c,d}(x) \equiv \frac{d^2\mathcal{G}_2^{-1}(\mathcal{G}_1(x))}{dx^2} = p(x)\left\{d(\log h)'(c + dH^{-1}(x)) - (\log h)'(H^{-1}(x))\right\}$$

where $p(x) = dh(c + dH^{-1}(x))/h^2(H^{-1}(x)) > 0 \;\forall x$. For fixed δ, i.e. $d = 1$, consider the case $c > 0$ i.e. $\epsilon_1 > \epsilon_2$; then $t_{c,1}(x) > 0$, the requirement for ϵ to act as a skewness parameter, corresponds precisely to $(\log h)'(x) > 0$ for all x. Likewise, $c < 0$ requires $(\log h)'(x) < 0$ for all x. Now fix $c = 0$ for the symmetric case $\epsilon_1 = \epsilon_2 = 0$. For δ to be a kurtosis parameter we need $t_{0,d}(x) > 0$ for $x > 0$ and for this it is certainly also sufficient that $\log h$ is increasing if $d > 1$ or that $\log h$ is decreasing if $d < 1$. ■

Another requirement that potentially further narrows the field of potential H's is unimodality of all members of the resulting family of distributions. We are keen on this since we believe that we are in the business of providing 'component' unimodal distributions which can be combined, interpretably, by mixture modelling if multimodality is present in one's data. Unfortunately, unimodality seems to require verification on a case-by-case basis (though it was used to disqualify $H(x) = \sinh^{-1}(x)$ for normal G in

Section 3.4). That said, it reinforces the requirement that $h(x) > 0 \; \forall x$ else, if $h(x_0) = 0$ for some x_0, the density associated with distribution (9) will be zero at $x = H((x_0 - \epsilon)/\delta)$ and nonzero to either side; this removes candidates of the form $H(x) = |x|^\gamma$, $\gamma > 0$.

Other considerations include explicit invertibility of H, differentiability (perhaps), and the type and breadth of effect on tails. We have not been able to come up with any viable alternative to $H(x) = \sinh(x)$.

8.2. Which central density?

The normal distribution is, of course, but one particular choice for the 'central' simple symmetric distribution mentioned in Section 8.1; let this distribution have density g. Then the transformed family of distributins has densities of the form

$$g_{\epsilon,\delta}(x) = \frac{\delta C_{\epsilon,\delta}(x)}{\sqrt{1+x^2}} g\left\{S_{\epsilon,\delta}(x)\right\}. \tag{10}$$

Several of the properties developed for the normal distribution hold immediately for other g too: examples include its distribution and quantile function (in terms of G and G^{-1}), skewness and kurtosis ordering properties, etc; some properties need to be investigated on a case-by-case basis. A sufficient condition for unimodality is that

$$1 + x(1+x^2)(\log g)'(x) + (1+x^2)^2(\log g)''(x) < 0 \; \forall x$$

which has been satisfied for all the g we have considered.

A major reason for choosing a different g would be if testing for some other simple symmetric distribution, such as the logistic, were of interest. We would expect likelihood ratio testing within a g-based family to perform as well as it does for the normality case in Section 5.

A second consideration might be the tailweight properties of g-based families. For small δ, and ignoring all constants,

$$g_{\epsilon,\delta}(|x|) \sim |x|^{\delta-1} g(|x|^\delta) \quad \text{as} \quad |x| \to \infty;$$

for example, simple exponential tails like those of the logistic lead to 'Weibull-type' tails, $|x|^{\delta-1}\exp(-|x|^\delta)$, while power tails, of the form $g(|x|) \sim |x|^{-(\alpha+1)}$, $\alpha > 0$, lead to 't-type' power tails for $g_{\epsilon,\delta}$ of the limiting form $|x|^{-(\nu+1)}$

where $\nu = \alpha\delta$. The Cauchy distribution as g leads to the particularly simple expression

$$g_{\epsilon,\delta}(x) = \frac{\delta}{\pi} \frac{1}{\sqrt{1+x^2}\; C_{\epsilon,\delta}(x)}.$$

But centring the family of distributions at such a heavy-tailed case has consequences for the lightness of tails as $\delta \to \infty$. In the symmetric case, $\epsilon = 0$, the Cauchy-based family tends to the hyperbolic secant density $\{\pi \cosh(x)\}^{-1}$, which is both intriguing and indicative of relatively heavy 'light' tails; they are of simple exponential form.

Again, aside from distributional testing requirements, it is difficult to see beyond the normal-based family as the most useful general tool.

8.3. Which method of introducing skewness?

Formulae (3) and (4) suggest an alternative method of introducing skewness into the symmetric sinh-arcsinh transformation. Instead of $S_{\epsilon,\delta}(X)$ as defined there, consider

$$S_{\delta,\gamma}(X) \equiv \frac{1}{2}\left\{\exp(\delta \sinh^{-1}(X)) - \exp(-\gamma \sinh^{-1}(X))\right\}, \qquad (11)$$

where $\delta, \gamma > 0$. Then define $X_{\delta,\gamma}$ by $Z = S_{\delta,\gamma}(X_{\delta,\gamma})$, $X_{\delta,\gamma}$ having density $f_{\delta,\gamma}$, not shown to save space. Note that (the same) symmetric cases now arise from setting $\gamma = \delta$. In fact, δ now controls the weight of the right-hand tail of the distribution, while γ controls the left-hand tail in the same way. Skewness arises implicitly from the imbalance between the tails when $\delta \neq \gamma$: if $\delta < \gamma$, the left-hand tail is lighter than the right and the resulting skewness is positive, if $\delta > \gamma$, negative skewness ensues. This can be contrasted with the way in which skewness in (2) is introduced and controlled by differential scaling of tails. It is clear that $f_{\gamma,\delta}(x) = f_{\delta,\gamma}(-x)$.

Many properties of these skew sinh-arcsinh distributions can also be determined although the family is a little less tractable than is that based on $S_{\epsilon,\delta}$. Briefly, the distribution function and quantile functions associated with (11) — in the normal case — are $F_{\delta,\gamma}(x) = \Phi(S_{\delta,\gamma}(x))$ and $Q_{\delta,\gamma}(u) = S^{-1}_{\delta,\gamma}(\Phi^{-1}(u))$; the latter is not explicitly invertible in general although its inverse is easy to compute. A nice property of the family based on (11) is that its median is always zero. We have much numerical evidence that $f_{\gamma,\delta}$ remains unimodal for all values of $\delta, \gamma > 0$, but have been unable to prove it. Plots of the Bowley skewness (not shown) indicate that the entire range of Bowley skewness values, from -1 to $+1$, is achieved within this family.

Tests of normality and symmetry can, of course, be based on fitting $f_{\gamma,\delta}$ in the same way as they were in Sections 5 and 6 for $f_{\epsilon,\delta}$. We repeated all the simulations reported there for the alternative skewness family too. Much the most striking feature of the results is their extreme similarity; in almost all cases, results like those shown in Figs 5 to 10 provide excellent approximations to the equivalent results for the alternative family. In terms of testing normality, and from the viewpoint of the alternative family, we might claim a slightly better holding of size for $n \geq 100$ but slightly worse for $n \leq 50$, with slightly lower power for heavy-tailed distributions and slightly higher power for lighter-tailed distributions. But emphasis here is on the word 'slightly'. Something similar was observed in the context of testing symmetry save for one exceptional case: the test based on $f_{\gamma,\delta}$ was rather less able to maintain size than was the test based on $f_{\epsilon,\delta}$ for the (heavy-tailed) t_2 distribution. (A nominal 5% test exhibited significance levels around 10% for the LRT based on $f_{\epsilon,\delta}$, rising to 20% and more for the test based on $f_{\gamma,\delta}$.)

All told, however, there is relatively little to choose between $f_{\epsilon,\delta}$ and $f_{\gamma,\delta}$ in many respects. We have focussed on the former in this paper primarily because of its greater tractability and secondarily because of minor practical advantages.

9. Discussion

We would like to argue that, far from being 'just another' four-parameter family of distributions on the real line with rather similar properties, the distributions of this paper fill a niche that is currently very sparsely populated. On the one hand, many if not most families of distributions on $\mathbb{R}$ concentrate on providing tailweights heavier than those of the normal (often with the normal distribution as their lightest tailed limit). Examples include stable laws and various 'skew-t' distributions which include Student's t distributions as their symmetric special cases; see, for example, Jones and Faddy (2003) and Azzalini and Genton (2008). On the other hand, few families of distributions on $\mathbb{R}$ have much in the way of light-tailed membership. An exception is the exponential power distributions (Box and Tiao, 1973, Tadikamalla, 1980) and their natural two-piece skew counterparts. The new distributions fill something of a gap between these two sorts of distributions. Like skew-t distributions, they allow tails considerably heavier than the normal, although their tails are not quite as heavy as the t's power tails can be, but unlike skew-t distributions they allow lighter than normal tails also. Like exponential power distributions, the new distributions allow much lighter tails than normal (though not as light as the uniform limit of the exponential power) and heavier tails than the normal, but in the latter case escape the purely exponential nature of the exponential power tails. We reiterate that the sinh-arcsinh distributions achieve these properties in a manner something like an amalgamation of Johnson S_U and sinh-normal distributions. Indeed, the sinh-arcsinh distribution can be seen as a generalised Johnson distribution where the sinh transformation (as in Johnson, 1949) is applied not to the normal distribution but to the sinh-normal distribution!

It is also especially appealing, in our view, to have such a family of distributions 'centred' on the normal distribution in order, as exemplified in Section 5, to allow standard likelihood ratio testing for normality against skew and light- and heavy-tailed distributions within the sinh-arcsinh family. This is in contrast to families in which the normal distribution is a limiting case. Moreover, the resulting tests are widely applicable: they turn out to compete with, and essentially outperform, existing omnibus tests of normality against alternatives not in the sinh-arcsinh family. (Essentially, of course, the tests work by approximating the distribution of the data by a member of the sinh-arcsinh family, which proves to be an adequate approximation at least for most unimodal densities.) Similar remarks apply to testing for symmetry via LRTs within this class.

Finally, this paper has been rather long in gestation and the first author has talked on the topic a number of times, including Jones (2005). It is therefore the case that the sinh-arcsinh distribution has already been implemented (under the acronym 'shash') in the GAMLSS software package (Stasinopoulos and Rigby, 2007).

Acknowledgements

We are very grateful to Professor Mohsen Pourahmadi for drawing our attention to the Gebelein inequality, to Dr Chris Banks and Professor Paul Garthwaite for the snow depth data, and to various colleagues for the enthusiasm with which they have greeted versions of these ideas. The work of the second author was partially supported by project MTM2007–61470 of the Spanish Ministry of Science and Education.

References

Anderson, T.W. and Darling, D.A. (1952) Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Ann. Math. Statist.*, **23**, 193–212.

Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12**, 171–178.

Azzalini, A. and Genton, M.G. (2008) Robust likelihood methods based on the skew-t and related distributions. *Internat. Statist. Rev.*, **76**, 106–129.

Baghdachi, J. and Balakrishnan, N., eds. (2008) *Handbook of the Logistic Distribution*, Second Edition, to appear.

Banks, C.J. (2006) Sea ice thickness and iceberg distribution in the Southern Ocean. Ph.D. thesis, Department of Earth Sciences, The Open University, Milton Keynes.

Barndorff-Nielsen, O.E. (1978) Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Statist.*, **5**, 151–157.

Boos, D.D. (1982) A test for asymmetry associated with the Hodges-Lehmann estimator. *J. Am. Statist. Assoc.*, **77**, 647–651.

Bowley, A.L. (1937) *Elements of Statistics*, Sixth Edition. New York: Scribner.

Box, G.E.P. and Tiao, G.C. (1973) *Bayesian Inference in Statistical Analysis.* Reading, MA: Addison-Wesley.

Cabilio, P. and Masaro, J. (1996) A simple test of symmetry about an unknown median. *Can. J. Statist.*, **24**, 349–361.

D'Agostino, R.B. (1971) An omnibus test of normality for moderate and large sample size. *Biometrika*, **58**, 341–348.

D'Agostino, R.B. (1972) Small sample probability points for the D test of normality. *Biometrika*, **59**, 219–221.

Downton, F. (1966) Linear estimates with polynomial coefficients. *Biometrika*, **53**, 129–141.

Fechner, G.T. (1897) *Kollectivmasslehre*. Leipzig: Engleman.

Fernandez, C. and Steel, M.F.J. (1998) On Bayesian modeling of fat tails and skewness. *J. Amer. Statist. Assoc.*, **93**, 359–371.

Gebelein, H. (1941) Das statistische Problem der Korrelation als Variations – und Eigenwertproblem und sein Zusammenhang mit der Ausgleichungsrechnung. *Z. Angew. Math. Mech.*, **21**, 364–379.

Genton, M.G., ed. (2004) *Skew-Elliptical Distributions and Their Applications; a Journey Beyond Normality*. Boca Raton: Chapman and Hall/CRC.

Gradshteyn, I.S. and Ryzhik, I.M. (1994) *Table of Integrals, Series, and Products*, Fifth Edition. San Diego: Academic Press.

Groeneveld, R.A. and Meeden, G.L. (1984) Measuring skewness and kurtosis. *Statistician*, **33**, 391–399.

Holland, P.W. and Wang, Y.J. (1987) Dependence function for continuous bivariate densities. *Commun. Statist. Theory Meth.*, **16**, 863–876.

Jarque, C. and Bera, A. (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economet. Lett.*, **6**, 255–259.

Joe, H. (1997) *Multivariate Models and Dependence Concepts*. London: Chapman and Hall.

Johnson, N.L. (1949) Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 149–76.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994) *Continuous Univariate Distributions*, Volume 1, Second Edition. New York: Wiley.

Jones, M.C. (1996) The local dependence function. *Biometrika*, **83**, 899–904.

Jones, M.C. (2004) Families of distributions arising from distributions of order statistics (with discussion). *Test*, **13**, 1–43.

Jones, M.C. (2005) Contribution to the discussion of "Generalized additive models for location, scale and shape", by R.A. Rigby and D.M. Stasinopoulos. *Appl. Statist.*, **54**, 546–547.

Jones, M.C. and Faddy, M.J. (2003) A skew extension of the t distribution, with applications. *J. R. Statist. Soc. B*, **65**, 159–174.

Koyak, R.A. (1987) On measuring internal dependence in a set of random variables. *Ann. Statist.*, **15**, 1215–1228.

Mudholkar, G.S. and Hutson, A. (2000) The epsilon-skew-normal distribution for analyzing near-normal data. *J. Statist. Planning Inference*, **83**, 291–309.

Nelder, J.A. and Mead, R. (1965) A simplex-method for function minimization. *Comput. J.*, **7**, 308–313.

Rieck, J.R. and Nedelman, J.R. (1991) A log-linear model for the Birnbaum–Saunders distribution. *Technometrics*, **33**, 51–60.

Royston, P. (1965) AS R94 A remark on algorithm AS 181: the W-test for normality. *Appl. Statist.*, **44**, 547–551.

Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non-Gaussian Random Processes; Stochastic Models with Infinite Variance.* Boca Raton: Chapman and Hall/CRC.

Shapiro, S. and Wilk, M. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611.

Stasinopoulos, D.M. and Rigby, R.A. (2007) Generalized additive models for location scale and shape (GAMLSS) in R. *J. Statist. Software*, **23**, Issue 7.

Tadikamalla, P.R. (1980) Random sampling from the exponential power distribution. *J. Am. Statist. Assoc.*, **75**, 683–686.

Thadewald, T. and Büning, H. (2007) Jarque-Bera test and its competitors for testing normality – a power comparison. *J. Appl. Statist.*, **34**, 87–105.

Tiku, M.L., Islam, M.Q. and Selcuk, A.S. (2001) Non-normal regression II: symmetric distributions. *Commun. Statist. Theory Meth.*, **30**, 1021–1045.

Zhang, J. and Wu, Y. (2005) Likelihood-ratio tests for normality. *Comput. Statist. Data Anal.*, **49**, 709–721.

van Zwet, W.R. (1964) *Convex Transformations of Random Variables.* Amsterdam: Mathematisch Centrum.