TWEEDIE DOMINANCE FOR AUTOCALIBRATED PREDICTORS AND LAPLACE TRANSFORM ORDER

Michel Denuit, Julien Trufin

LIDAM Discussion Paper ISBA
2022 / 40

UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

TWEEDIE DOMINANCE FOR AUTOCALIBRATED PREDICTORS AND LAPLACE TRANSFORM ORDER

Michel Denuit
Institute of Statistics, Biostatistics and Actuarial Science - ISBA
Louvain Institute of Data Analysis and Modeling - LIDAM
UCLouvain
Louvain-la-Neuve, Belgium

Julien Trufin
Department of Mathematics
Université Libre de Bruxelles (ULB)
Brussels, Belgium

November 26, 2022

Abstract

While Krüger and Ziegel (2021) defined forecast dominance, or Bregman dominance as dominance for every Bregman loss function, this short note explores Tweedie dominance proposed by Denuit et al. (2021) to compare competing models. A necessary and sufficient condition is established under autocalibration. Moreover, Laplace transform order turns out to be a sufficient condition for Tweedie dominance between autocalibrated predictors. This shows that Tweedie dominance is a rather weak concept compared to Bregman dominance that reduces to the well-known convex order among autocalibrated predictors.

Keywords: Tweedie distribution family, Autocalibrated models, Tweedie dominance, Laplace transform order.

1 Introduction and motivation

1.1 Regression for the mean

Consider a response Y and a set of features $X_1, \ldots, X_p$ gathered in the vector $\boldsymbol{X} \in \mathcal{X}$ (classically, $\mathcal{X} \subset \mathbb{R}^p$). In a regression setting, the dependence structure inside the random vector $(Y, X_1, \ldots, X_p)$ is exploited to extract the information contained in $\boldsymbol{X}$ about Y. Here, we assume that the target is the conditional expectation $\mu(\boldsymbol{X}) = \mathrm{E}[Y|\boldsymbol{X}]$ of the response Y given the available information $\boldsymbol{X}$.

The function $\boldsymbol{x} \mapsto \mu(\boldsymbol{x}) = \mathrm{E}[Y|\boldsymbol{X} = \boldsymbol{x}]$ is approximated by a predictor $\boldsymbol{x} \mapsto \pi(\boldsymbol{x})$ with a simpler structure. Once fitted by minimizing an empirical loss function on the training data set, this produces estimates $\widehat{\pi}(\boldsymbol{x})$ for $\mu(\boldsymbol{x})$, or fitted values.

1.2 Bregman loss functions

Conditional expectations can be consistently estimated only if the expected loss is minimum for the mean response. This corresponds to the concept of consistent loss function. Precisely, a loss function is consistent for the mean if no other quantity leads to a lower expected loss than the mean.

Bregman loss functions are of the form

$$L(y, m) = \ell(y) - \ell(m) - \ell'(m)(y - m) \text{ for a convex function } \ell. \tag{1.1}$$

Savage (1971) showed, subject to weak regularity conditions, that for any loss function (1.1) and response Y,

$$\mathrm{E}[L(Y, \mathrm{E}[Y])] \leq \mathrm{E}[L(Y, m)] \text{ for any } m \text{ in the support of } Y.$$

The class of Bregman loss functions is thus consistent for the (conditional) mean functional.

1.3 Tweedie loss functions

Generalized Linear Models (GLMs) with power variance function have been successfully used in many applications. Poisson, Gamma and Inverse Gaussian distributions belong to this family, as well as compound Poisson distributions with Gamma-distributed summands. Within the Tweedie subclass of the Exponential Dispersion family, the logarithm of the probability mass function for a discrete response, or of the probability density function for a continuous response, is of the form $(y\theta - a(\theta))/\phi$ up to a constant term, for some known dispersion parameter ϕ and non-decreasing and convex cumulant function $a(\cdot)$ corresponding to a variance function of the form $V(\mu) = \mu^\xi$ for some power parameter $\xi \geq 1$.

For an appropriate choice of ℓ in (1.1), we recover Tweedie deviance. Precisely, the

function $\ell(m) = 2\big(m\theta_m - a(\theta_m)\big)$ with $a'(\theta_m) = m$ results in the loss function

$$\begin{aligned} L(y,m) &= 2\big(y(\theta_y - \theta_m) - a(\theta_y) + a(\theta_m)\big) \text{ where } a'(\theta_y) = y \\ &= \begin{cases} 2\left(y \ln \frac{y}{m} - (y-m)\right) & \text{for } \xi = 1 \\ 2\left(-\ln \frac{y}{m} + \frac{y}{m} - 1\right) & \text{for } \xi = 2 \\ 2\left(\frac{y^{2-\xi}}{(1-\xi)(2-\xi)} - \frac{y m^{1-\xi}}{1-\xi} + \frac{m^{2-\xi}}{2-\xi}\right) & \text{for } \xi > 1 \text{ and } \xi \neq 2. \end{cases} \end{aligned} \tag{1.2}$$

The deviance loss function (1.2) is thus a Bregman loss function

1.4 Performance evaluation

The developments in this paper hold the training set as fixed. Precisely, $\widehat{\pi}$ is assumed to have been obtained by minimizing the empirical loss on some training set and all formulas in the text are meant given this training set. The respective performances of competing models can then be assessed on the basis of a validation set $\{(Y_i, \boldsymbol{X}_i),\ i = 1, 2, \ldots, n\}$, that has not been used to obtain $\widehat{\pi}$. Provided the validation set comprises a large number of independent and identically distributed pairs $(Y_i, \boldsymbol{X}_i)$, this allows us to perform calculations with a generic random vector $(Y, \boldsymbol{X})$, independent of, and distributed as the observations contained in the training set.

We adopt this approach in this note and we assess performances of $\widehat{\pi}$ with the help of the predictive Tweedie deviance corresponding to (1.2) given by

$$D(\xi, \widehat{\pi}) = \begin{cases} \mathrm{E}\left[\widehat{\pi}(\boldsymbol{X}) - Y \ln \widehat{\pi}(\boldsymbol{X})\right] & \text{for } \xi = 1 \\ \mathrm{E}\left[\ln \widehat{\pi}(\boldsymbol{X}) + \frac{Y}{\widehat{\pi}(\boldsymbol{X})}\right] & \text{for } \xi = 2 \\ \mathrm{E}\left[\frac{\widehat{\pi}(\boldsymbol{X})^{2-\xi}}{2-\xi} - \frac{Y \widehat{\pi}(\boldsymbol{X})^{1-\xi}}{1-\xi}\right] & \text{for } \xi > 1 \text{ and } \xi \neq 2 \end{cases} \tag{1.3}$$

where the terms not involving $\widehat{\pi}(\boldsymbol{X})$ have been discarded.

2 Tweedie dominance and autocalibration

Henceforth, we consider $Y \geq 0$ as well as $\widehat{\pi}(\boldsymbol{X}) \geq 0$. We assume that $\widehat{\pi}(\boldsymbol{X})$ as well as the conditional expectation $\mu(\boldsymbol{X})$ are continuous random variables admitting probability density functions. This is generally the case when there is at least one continuous feature contained in the available information $\boldsymbol{X}$ and the function $\widehat{\pi}$ is a continuously increasing function of a real score built from $\boldsymbol{X}$.

Bregman dominance, also called forecast dominance is defined as dominance for every Bregman loss function (1.1). Precisely, $\widehat{\pi}_2$ outperforms $\widehat{\pi}_1$ in terms of Bregman dominance if the inequality $\mathrm{E}[L(Y, \widehat{\pi}_2(\boldsymbol{X}))] \leq \mathrm{E}[L(Y, \widehat{\pi}_1(\boldsymbol{X}))]$ holds true for every loss function L of

the form given in (1.1). We refer the interested reader to Krüger and Ziegel (2021) and the references therein for an extensive presentation of this concept.

Tweedie dominance introduced in Denuit et al. (2021) restricts Bregman loss functions to Tweedie deviances (1.3). Precisely, $\widehat{\pi}_2$ outperforms $\widehat{\pi}_1$ in terms of Tweedie dominance if the inequality $D(\xi, \widehat{\pi}_2) \leq D(\xi, \widehat{\pi}_1)$ holds true for every power parameter $\xi \geq 1$. In their Proposition 4.1, Denuit et al. (2021) established sufficient conditions for Tweedie dominance in terms of simple test functions.

Recall that a predictor $\widehat{\pi}$ is said to be autocalibrated if $\widehat{\pi}(\boldsymbol{X}) = \mathrm{E}[Y|\widehat{\pi}(\boldsymbol{X})]$. Autocalibration thus ensures that the average response in a neighborhood defined with the help of the autocalibrated predictor under consideration matches the latter. See, e.g., Krüger and Ziegel, (2021).

The next result shows that Tweedie dominance can be characterized by the comparison of simple test functions for autocalibrated predictors.

Proposition 2.1. *Consider two autocalibrated predictors $\widehat{\pi}_1$ and $\widehat{\pi}_2$. Then, $\widehat{\pi}_2$ outperforms $\widehat{\pi}_1$ in terms of Tweedie dominance if, and only if, the inequality*

$$\mathrm{E}\left[\psi_\xi(\widehat{\pi}_1(\boldsymbol{X}))\right] \leq \mathrm{E}\left[\psi_\xi(\widehat{\pi}_2(\boldsymbol{X}))\right] \tag{2.1}$$

holds true for every power parameter $\xi \geq 1$, where

$$\psi_\xi(\pi) = \begin{cases} \pi \ln \pi \text{ for } \xi = 1 \\ -\ln \pi \text{ for } \xi = 2 \\ \frac{\pi^{2-\xi}}{\xi - 2} \text{ for } \xi > 1 \text{ and } \xi \neq 2. \end{cases} \tag{2.2}$$

Proof. For any autocalibrated predictor $\widehat{\pi}$ and measurable function $g : \mathbb{R} \to \mathbb{R}$, we have

$$\begin{aligned} \mathrm{E}\big[Y g\big(\widehat{\pi}(\boldsymbol{X})\big)\big] &= \mathrm{E}\Big[\mathrm{E}\big[Y g\left(\widehat{\pi}(\boldsymbol{X})\right) \big| \widehat{\pi}(\boldsymbol{X})\big]\Big] \\ &= \mathrm{E}\Big[\mathrm{E}\big[Y|\widehat{\pi}(\boldsymbol{X})\big] g\left(\widehat{\pi}(\boldsymbol{X})\right)\Big] \\ &= \mathrm{E}\left[\widehat{\pi}(\boldsymbol{X}) g\left(\widehat{\pi}(\boldsymbol{X})\right)\right]. \end{aligned}$$

Therefore, the predictive deviance $D(\xi, \widehat{\pi})$ defined in (1.3) simplifies to

$$D(\xi, \widehat{\pi}) = \begin{cases} \mathrm{E}\left[\widehat{\pi}(\boldsymbol{X}) - \widehat{\pi}(\boldsymbol{X}) \ln \widehat{\pi}(\boldsymbol{X})\right] \text{ for } \xi = 1 \\ \mathrm{E}\left[\ln \widehat{\pi}(\boldsymbol{X}) + 1\right] \text{ for } \xi = 2 \\ \mathrm{E}\left[\frac{\left(\widehat{\pi}(\boldsymbol{X})\right)^{2-\xi}}{2-\xi} - \frac{\left(\widehat{\pi}(\boldsymbol{X})\right)^{2-\xi}}{1-\xi}\right] \text{ for } \xi > 1 \text{ and } \xi \neq 2 \end{cases}. \tag{2.3}$$

Since $\widehat{\pi}_1$ and $\widehat{\pi}_2$ are autocalibrated, we have $\mathrm{E}\left[\widehat{\pi}_1(\boldsymbol{X})\right] = \mathrm{E}\left[\widehat{\pi}_2(\boldsymbol{X})\right]$. Hence,

$$D(1, \widehat{\pi}_2) \leq D(1, \widehat{\pi}_1) \Leftrightarrow \mathrm{E}\left[\widehat{\pi}_1(\boldsymbol{X}) \ln \widehat{\pi}_1(\boldsymbol{X})\right] \leq \mathrm{E}\left[\widehat{\pi}_2(\boldsymbol{X}) \ln \widehat{\pi}_2(\boldsymbol{X})\right]$$

which corresponds to the stated inequality (2.1) for ψ_1 defined according to (2.2). Similarly,

$$D(2,\widehat{\pi}_2) \leq D(2,\widehat{\pi}_1) \Leftrightarrow \mathrm{E}\left[\ln \widehat{\pi}_2(\boldsymbol{X})\right] \leq \mathrm{E}\left[\ln \widehat{\pi}_1(\boldsymbol{X})\right]$$

which corresponds to (2.1) with ψ_2 as in (2.2). Finally, for $\xi > 1$ and $\xi \neq 2$,

$$D(\xi,\widehat{\pi}_2) \leq D(\xi,\widehat{\pi}_1) \Leftrightarrow \mathrm{E}\left[\frac{\left(\widehat{\pi}_2(\boldsymbol{X})\right)^{2-\xi}}{(\xi-2)(1-\xi)}\right] \leq \mathrm{E}\left[\frac{\left(\widehat{\pi}_1(\boldsymbol{X})\right)^{2-\xi}}{(\xi-2)(1-\xi)}\right]$$

which is indeed (2.1) with ψ_ξ in (2.2), noticing that $1-\xi < 0$. This ends the proof. □

For any $\xi \geq 1$, the function ψ_ξ defined in (2.2) is convex. Therefore, if $\widehat{\pi}_1(\boldsymbol{X})$ and $\widehat{\pi}_2(\boldsymbol{X})$ satisfy $\widehat{\pi}_1(\boldsymbol{X}) \preceq_{\mathrm{CX}} \widehat{\pi}_2(\boldsymbol{X})$ where $\preceq_{\mathrm{CX}}$ is the convex order defined as $\mathrm{E}[g(\widehat{\pi}_1(\boldsymbol{X}))] = \mathrm{E}[g(\widehat{\pi}_2(\boldsymbol{X}))]$ for every convex function g such that the expectations exist, then $\widehat{\pi}_2$ outperforms $\widehat{\pi}_1$ in Tweedie dominance. This was expected since we know from Theorem 3.1 in Krüger and Ziegel (2021) that for two autocalibrated predictors $\widehat{\pi}_1$ and $\widehat{\pi}_2$, $\widehat{\pi}_2$ outperforms $\widehat{\pi}_1$ in terms of Bregman dominance if, and only if, $\widehat{\pi}_2(\boldsymbol{X})$ dominates $\widehat{\pi}_1(\boldsymbol{X})$ in the convex order. The convex order is thus a sufficient condition for Tweedie dominance in this case.

The test functions ψ_ξ in (2.2) are actually not only convex but have second, third, fourth, etc. derivatives alternating in sign. Precisely, the mth derivative of ψ_ξ is positive when m is even and negative when m is odd for any $m \geq 2$. This echoes the characterization of the Laplace transform order with the help of functions with a completely monotone first derivative (Theorem 5.A.4 in Shaked and Shanthikumar, 2007). It turns out that Laplace transform order is a sufficient condition for Tweedie dominance under autocalibration, as shown next.

Proposition 2.2. *Let $\mathrm{L}_{\widehat{\pi}}(\cdot)$ denote the Laplace transform of $\widehat{\pi}$, that is, $\mathrm{L}_{\widehat{\pi}}(s) = \mathrm{E}[\exp(-s\widehat{\pi})]$. For two autocalibrated predictors $\widehat{\pi}_1$ and $\widehat{\pi}_2$, if the inequality $\mathrm{L}_{\widehat{\pi}_1}(s) \leq \mathrm{L}_{\widehat{\pi}_2}(s)$ holds true for all $s \geq 0$, then $\widehat{\pi}_2$ outperforms $\widehat{\pi}_1$ in terms of Tweedie dominance.*

Proof. The second derivative of the test functions ψ_ξ in (2.2) is given by $\psi''_\xi(t) = t^{-\xi}$ for all $\xi \geq 1$. We thus see that ψ''_ξ is non-negative and possesses negative odd derivatives and positive even derivatives. With the representation of completely monotone functions given e.g. in Lemma 2.1 of Merkle (2014), we obtain

$$\psi''_\xi(t) = \int_0^\infty \exp(-ut)\mathrm{d}\vartheta(u)$$

for some measure ϑ on $(0,\infty)$. This gives

$$\begin{aligned}
\psi'_\xi(t) &= c_1 - \int_0^\infty \frac{\exp(-ut)}{u}\mathrm{d}\vartheta(u) \\
\psi_\xi(t) &= c_2 + c_1 t + \int_0^\infty \frac{\exp(-ut)}{u^2}\mathrm{d}\vartheta(u)
\end{aligned}$$

for some real constants c_1 and c_2. We then obtain the following representation for the expectations of the test functions ψ_ξ:

$$\mathrm{E}[\psi_\xi(\widehat{\pi}(\boldsymbol{X}))] = c_2 + c_1\mathrm{E}[\widehat{\pi}(\boldsymbol{X})] + \int_0^\infty \frac{1}{u^2}\mathrm{L}_{\widehat{\pi}}(u)\mathrm{d}\vartheta(u),$$

which ends the proof. □

Proposition 2.2 shows that Tweedie dominance for autocalibrated predictors is a rather weak dominance concept compared to Bregman dominance that reduces to the convex order. The testing procedure proposed by Bhattacharyya et al. (2021) can be applied to check whether two autocalibrated predictors are ranked according to Laplace order. Contrary to Bregman dominance that is equivalent to the ordering of the Lorenz curves associated to the autocalibrated predictors under consideration, Tweedie dominance thus allows multiple crossings between the Lorenz curves. We refer the interested reader to Chiu (2007, 2021) for more information about intersecting Lorenz curves.

It is tempting to conclude that Laplace transform order is in fact equivalent to Tweedie dominance. This is indeed the case if $\ln \widehat{\pi}_1 \geq 0$ and $\ln \widehat{\pi}_2 \geq 0$, that is, if $\widehat{\pi}_1 \geq 1$ and $\widehat{\pi}_2 \geq 1$, as shown in the next result.

Proposition 2.3. *Consider two autocalibrated predictors $\widehat{\pi}_1$ and $\widehat{\pi}_2$ such that $\widehat{\pi}_1 \geq 1$ and $\widehat{\pi}_2 \geq 1$. Then, $\widehat{\pi}_2$ outperforms $\widehat{\pi}_1$ in terms of Tweedie dominance if, and only if, $\mathrm{L}_{\widehat{\pi}_1}(s) \leq \mathrm{L}_{\widehat{\pi}_2}(s)$ for all $s \geq 0$.*

Proof. Tweedie dominance ensures that

$$\frac{\mathrm{E}[\widehat{\pi}_1(\boldsymbol{X})^{2-\xi}]}{\xi - 2} \leq \frac{\mathrm{E}[\widehat{\pi}_2(\boldsymbol{X})^{2-\xi}]}{\xi - 2} \quad \text{for all } \xi > 1 \text{ and } \xi \neq 2,$$

so that we have

$$\mathrm{E}[\widehat{\pi}_1(\boldsymbol{X})^{-s}] \leq \mathrm{E}[\widehat{\pi}_2(\boldsymbol{X})^{-s}] \quad \text{for all } s \geq 0,$$

which can be rewritten as

$$\mathrm{E}[\exp(-s \ln \widehat{\pi}_1(\boldsymbol{X}))] \leq \mathrm{E}[\exp(-s \ln \widehat{\pi}_2(\boldsymbol{X}))] \quad \text{for all } s \geq 0.$$

Since $\ln \widehat{\pi}_1 \geq 0$ and $\ln \widehat{\pi}_2 \geq 0$, we can conclude that Tweedie dominance ensures that $\ln \widehat{\pi}_1$ and $\ln \widehat{\pi}_2$ are ordered in the Laplace transform order, that is,

$$\ln \widehat{\pi}_2(\boldsymbol{X}) \preceq_{\mathrm{Lt}} \ln \widehat{\pi}_1(\boldsymbol{X}).$$

Now applying Theorem 5.A.7(a) in Shaked and Shanthikumar (2007) with $g(t) = \exp(-t)$ gives

$$\exp(-\ln \widehat{\pi}_2(\boldsymbol{X})) = \widehat{\pi}_2(\boldsymbol{X})^{-1} \preceq_{\mathrm{Lt}} \exp(-\ln \widehat{\pi}_1(\boldsymbol{X})) = \widehat{\pi}_1(\boldsymbol{X})^{-1}.$$

Then, applying once again the same theorem with $g(t) = 1/t$ leads to

$$\widehat{\pi}_2(\boldsymbol{X}) \preceq_{\mathrm{Lt}} \widehat{\pi}_1(\boldsymbol{X}),$$

so that $\widehat{\pi}_1$ and $\widehat{\pi}_2$ are also ordered in the Laplace transform order, which ends the proof. □

Proposition 2.3 applies when $Y \geq 1$. In the continuous case when Y obeys the Gamma distribution for instance, this essentially means that Y is at least equal to one measurement unit.

Remark 2.4. In general, i.e. when $\widehat{\pi}_1 \geq 0$ and $\widehat{\pi}_2 \geq 0$, we can only get the following inequality. Define the function $\chi(x) : x \in (0, \infty) \to x^{-s}$, $s > 0$. Since we have

$$\chi(x) = \int_0^\infty \exp(-ux) \frac{u^{s-1}}{\Gamma(s)} \mathrm{d}u,$$

then Tweedie dominance implies

$$\begin{aligned}
&\mathrm{E}[\widehat{\pi}_1(\boldsymbol{X})^{-s}] \leq \mathrm{E}[\widehat{\pi}_2(\boldsymbol{X})^{-s}] \quad \text{for all } s > 0 \\
&\Leftrightarrow \mathrm{E}\left[\int_0^\infty \exp(-u\widehat{\pi}_1(\boldsymbol{X})) \frac{u^{s-1}}{\Gamma(s)} \mathrm{d}u\right] \leq \mathrm{E}\left[\int_0^\infty \exp(-u\widehat{\pi}_2(\boldsymbol{X})) \frac{u^{s-1}}{\Gamma(s)} \mathrm{d}u\right] \quad \text{for all } s > 0 \\
&\Leftrightarrow \int_0^\infty \big(\mathrm{L}_{\widehat{\pi}_1}(u) - \mathrm{L}_{\widehat{\pi}_2}(u)\big) u^{s-1} \mathrm{d}u \leq 0 \quad \text{for all } s > 0.
\end{aligned}$$

References

- Bhattacharyya, D., Khan, R. A., Mitra, M. (2021). Tests for Laplace order dominance with applications to insurance data. Insurance: Mathematics and Economics 99, 163-173.

- Chiu, W. H. (2007). Intersecting Lorenz curves, the degree of downside inequality aversion, and tax reforms. Social Choice and Welfare 28, 375-399.

- Chiu, W. H. (2021). Intersecting Lorenz curves and aversion to inverse downside inequality. Social Choice and Welfare 56, 487-508.

- Denuit, M., Charpentier, A., Trufin, J. (2021). Autocalibration and Tweedie dominance for insurance pricing with machine learning. Insurance: Mathematics and Economics 101, 485-497.

- Krüger, F., Ziegel, J.F. (2021). Generic conditions for forecast dominance. Journal of Business & Economic Statistics 39, 972-983.

- Merkle, M. (2014). Completely monotone functions: A digest. In "Analytic Number Theory, Approximation Theory, and Special Functions" (pp. 347-364). Springer, New York.

- Savage, L.J. (1971). Elicitation of personal probabilities and expectations. Journal of the American Statistical Association 66, 783-810.

- Shaked, M., Shanthikumar, J.G. (2007). Stochastic Orders. Springer, New York.