

A Generic Method for Density Forecasts Recalibration

Jerome Collet, Michael Richard

▶ **To cite this version:**

Jerome Collet, Michael Richard. A Generic Method for Density Forecasts Recalibration. 2018. hal-01812231

HAL Id: hal-01812231
https://hal.science/hal-01812231

Preprint submitted on 11 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Generic Method for Density Forecasts Recalibration

Jérôme Collet and Michael Richard

Abstract We address the calibration constraint of probability forecasting. We propose a generic method for recalibration, which allows us to enforce this constraint. It remains to be known the impact on forecast quality, measured by predictive distributions sharpness, or specific scores. We show that the impact on the Continuous Ranked Probability Score (CRPS) is weak under some hypotheses and that it is positive under more restrictive ones. We used this method on temperature ensemble forecasts and compared the quality of the recalibrated forecasts with that of the raw ensemble and of a more specific method, that is Ensemble Model Output Statistics (EMOS). Better results are shown with our recalibration rather than with EMOS in this case study.

Key words: Density forecasting; Rosenblatt transform; PIT series; calibration; bias correction

1 Introduction

Due to the increasing need for risk management, forecasting is shifting from point forecasts to density forecasts. Density forecast is an estimate of the conditional probability distribution. Thus, it provides a complete estimate of uncertainty, in contrast to point forecast, which is not concerned whith uncertainty.
Two alternative ways to evaluate density forecast exist.

Jérôme Collet
EdF R&D, 7 Boulevard Gaspard Monge 91120 Palaiseau. e-mail: jerome.collet@edf.fr

Michael Richard
University of Orléans and EdF R&D, 7 Boulevard Gaspard Monge 91120 Palaiseau. e-mail: michael-m.richard@edf.fr

- The first one was proposed by T. Gneiting: *Probabilistic forecasting aims to maximize the sharpness of the predictive distributions, subject to calibration, on the basis of the available information set*. Calibration means predictive distributions are consistent with observations, it is more formally defined in [5]; sharpness refers to the concentration of the density forecast, and even in the survey paper of T. Gneiting, it is not formally defined. An important feature of this framework is that we face a multi-objective problem, which is difficult.
- The second way is the use of a scoring rule, which assesses simultaneously calibration and sharpness. Concerning the well-known CRPS scoring rule, Hersbach [9] showed that it can be decomposed into three parts: reliability (or calibration) part, resolution (or sharpness) part, and uncertainty, which measures the intrinsic difficulty of the forecast. Bröcker [1] generalized this result to any proper score, that is any score which is minimal if the forecasted probability distribution is the true one (w.r.t the available information). Recently, Wilks [14] proposed to add an extra miscalibration penalty, in order to enforce calibration in ensemble postprocessing. Nevertheless, even if the score we use mixes calibration and sharpness, the framework is essentially different from the first one.

Besides these two alternative ways of evaluation, probabilistic forecast is mainly used in two different contexts: finance and economics, and weather forecast. In finance and economics, calibration is the unique objective, so a recent survey on "Predictive density evaluation" [2] is in fact entirely devoted to the validation of the calibration, without any hint of sharpness. In weather forecast, both ways of evaluation are used. For a quick view on forecasting methods in atmospheric sciences, one can look at [13]. In the works of T. Gneiting [6] [7], and in the seminal work of Krzysztofowicz [10], the goal is to improve sharpness, while preserving calibration. Nevertheless, one can state that there is no formal test of calibration in these works. In [3], the only measure used is the CRPS, and [8] addresses exclusively the calibration issue.

Here, we are interested in the first method of evaluation: calibration constraint and sharpness objective. Indeed, risk management involves many stakeholders and thus, calibration is a key feature of trust between stakeholders since it impacts all of them. For example, EDF also faces a regulatory constraint: the French technical system operator imposes that the probability of employing exceptional means (e.g., load shedding) to meet the demand for electricity must be lower than 1% for each week (RTE, 2004), so EDF has to prove the calibration of its forecasts. Even inside EDF, many different business units may be involved in the management of a given risk, so calibration is compulsory to obtain confidence between risk management stakeholders.

The consequence is that we face a multi-criterion problem, the goal of our contribution is to allow us to enforce the calibration constraint, in a generic way. Furthermore, we show that, even if the evaluation framework is the proper score use, recalibrating leads in many cases to an improvement, and to a very limited loss in other cases.

The remainder of this chapter will be organized as follows. The next section explains the principle of the method. The third part provides some theoretical results while the fourth is devoted to a case study.

2 Principle of the method

The Probability Integral Transform (PIT, Rosenblatt, 1952) is usually a measure of the calibration of density forecasts. Indeed, if $Y \sim F$ and is continue, the random variable $F(Y) \sim U[0,1]$. Thus, we can find in the literature many tests based on this transformation to evaluate the correct specification of a density forecast. In our case, it is used firstly to recalibrate the forecasts.

Let's look at the following case: let E be the set of all possible states of the world; for each forecasting time j the forecaster knows the current state of the world $e(j)$, and uses it to forecast. For example, in the case of a statistical regression model, E is the set of the possible values of the regressors, in the case of the post-processing of an weather forecasting model, E is the ensemble. The conditional **estimated** distribution is G_e, whereas the **true** one is F_e. So the PIT series is:

$$\text{PIT} \equiv \left(G_{e(j)}(Y_j)\right)_j.$$

- **A.2.1**: G_e is invertible $\forall\, e \in E$.

If E is discrete, we assume that the frequency of appearance of each state of the world e is p_e. Then, under the assumption **A.2.1**, the c.d.f of the PIT is:

$$C(y) \equiv \Pr(\, G(\boldsymbol{Y}) \leq y) \equiv \sum_e p_e F_e \circ G_e^{-1}(y).$$

Note that all the results obtained under the hypothesis that E is discrete are still valid in continuous case, even if we only treat the discrete case in this article.

- **A.2.2**: F is invertible.

We propose to use C to recalibrate the forecasts. For each quantile $\tau \in [0,1]$, we use the original model to forecast the quantile τ_C, such that $\Pr(\, G(\boldsymbol{Y}) \leq \tau_c) = \tau$. We remark that this implies $\tau_c = C^{-1}(\tau)$.
This correction makes sense since under the assumptions **A.2.1** and **A.2.2**:

$$\begin{aligned}\Pr(C \circ G(\boldsymbol{Y}) \leq y) &= \Pr(\, G(\boldsymbol{Y}) \leq C^{-1}(y)) \\ &= C \circ C^{-1}(y) \\ &= y,\end{aligned}$$

which means that the recalibrated forecasts are uniformly distributed on the interval [0,1].

Note that this method is close to the quantile-quantile correction as in [11] but here, we are concerned by PIT recalibration, which allows us to consider the conditional case.

3 Impact on global score

If we evaluate our method on the basis of calibration, it ensures this constraint is enforced. But it is important to know if our method is still useful even if one of the probability forecasting users prefers to use scores, for example the Continuous Ranked Probability Score (CRPS).

The CRPS:

$$CRPS(G,x) = \int_{-\infty}^{+\infty} (G(y) - \mathbf{1}_{\{x \leq y\}})^2 \, dy,$$

with G a function and x the observation, is used to evaluate the whole distribution, since it is minimized by the true c.d.f of X.

However, since we have:

$$CRPS(G,x) = 2 \int_0^1 L_\tau(x, G^{-1}(\tau)) \mathrm{d}\tau, \tag{1}$$

as shown in [12], with L_τ the Pinball-Loss function :

$$L_\tau(x,y) = \tau(x-y)\mathbf{1}_{\{x \geq y\}} + (y-x)(1-\tau)\mathbf{1}_{\{x<y\}},$$

with y the forecast, x the observation and $\tau \in [0,1]$ a quantile level, and that L_τ is easier to work with, we use this scoring rule to obtain results on CRPS.

L_τ is used to evaluate quantile forecasts. Indeed, it is a proper scoring for the quantile of level τ, since its expectation is minimized by the true quantile of the distribution of X.

To begin with, we will prove that under some hypotheses, our correction improves systematically the quality of the forecasts in an infinite sample. Then we will show that under less restrictive hypotheses, our correction deteriorates only slightly—in the worst case—the quality of the forecasts in a more realistic case, e.g finite sample.

3.1 Impact on score: conditions for improvement

To assess conditions for improvements, we need to consider:

$$\mathrm{E}_Y[L_\tau - L_{\tau_c}] \equiv \mathrm{E}_{Y,e}[L_\tau(\boldsymbol{Y}, G_e^{-1}(\tau))] - \mathrm{E}_{Y,e}[L_\tau(\boldsymbol{Y}, G_e^{-1}(\tau_c))].$$

Here, under the assumption **A.2.1**, $G_e^{-1}(\tau)$ corresponds to the estimated conditional quantile of level $\tau \in [0,1]$ and $G_e^{-1}(\tau_c)$ to the corrected conditional quantile. Denote: $\eta_e \equiv G_e - F_e$. Considering small errors of specification and regularity conditions on the estimated c.d.f G_e, the true one F_e and their derivatives g_e and f_e:

- **A.3.1.1**: G_e are $C^3 \ \ \forall\, e \in E$.
- **A.3.1.2**: F_e are C^3 and invertible $\ \forall\, e \in E$.
- **A.3.1.3**: η_e, f_e and their derivatives are bounded $\forall\, e \in E$ by a constant which doesn't depend on e.
- **A.3.1.4**: $\forall\, \tau \in [0,1]\,, \forall\, e \in E$, η_e, its first, second and third derivatives are finite in $F_e^{-1}(\tau)$,

and using functional derivatives, directional derivatives and the implicit function theorem (proof in Appendix) we can rewrite (adding the assumption **A.2.1**):

$$\begin{aligned}\mathrm{E}_Y[L_\tau - L_{\tau_c}] \sim & \left(\sum_e \frac{p_e \eta_e(F_e^{-1}(\tau))}{f_e(F_e^{-1}(\tau))}\right)\left(\sum_e p_e \eta_e(F_e^{-1}(\tau))\right) \\ & - \left(\sum_e \frac{p_e}{2 f_e(F_e^{-1}(\tau))}\right)\left(\sum_e p_e \eta_e(F_e^{-1}(\tau))\right)^2 \\ & \text{as } max\ \eta_e \to 0, \end{aligned} \tag{2}$$

with p_e the frequency of appearance of the state e.

This result allows us to find conditions for improvement of the expectation of the Pinball-Loss score, with additional following conditions.

- **A.3.1.5**: η or f^{-1} is a constant, or $max_e(\bullet)/min_e(\bullet) < 3 + 2\sqrt{2}$ for both η_e and f_e^{-1}, $\forall\, e \in E$,
- **A.3.1.6**: the correlation between η and f^{-1}, $\sigma_{f^{-1}}$ or σ_η is null. Here the correlation is used as a descriptive statistics notation, even if the series η and f^{-1} are deterministic. The null correlation means that the difference between the true probability distribution function and the model have the same magnitude in low and in high density regions.

Under the assumption **A.3.1.5** or **A.3.1.6**, if $\exists\, \nu \geq 0$ (sufficiently small) $\forall\, e \in E$ $\forall\, y \in \mathbf{R}; |\eta_e(y)| \leq \nu$, we show that (proof in Appendix):

$$0 \leq \mathrm{E}_Y[L_\tau - L_{\tau_c}] \text{ and} \tag{3}$$

$$0 \leq \mathrm{E}_Y[CRPS_{G,C\circ G}], \tag{4}$$

with $\mathrm{E}_Y[CRPS_{G,C\circ G}] \equiv \mathrm{E}_Y[CRPS(G,\mathbf{Y}) - CRPS(C\circ G,\mathbf{Y})]$. In other words, with those restrictions, our recalibration systematically improves the quality of the forecasts. Indeed, remember that the expectation of the Pinball-Loss score is minimized by the true quantile of the distribution of $\mathbf{Y}$ and negatively oriented. Thus, the lower the expectation of the Pinball-Loss score, the better.

3.2 Impact on score: bounds on degradation

In reality, we cannot obtain the corrected probability level $\tau_c \in [0,1]$, and we need to estimate it. If we want to upper bound the degradation, we can study the more realistic case of

$$\mathrm{E}_Y[L_{\widehat{\tau}_c} - L_\tau] \equiv \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n} L_\tau(y_j, G_j^{-1}(\widehat{\tau}_c)) - L_\tau(y_j, G_j^{-1}(\tau))\right], \tag{5}$$

with τ, $\widehat{\tau}_c \in [0,1]$. In our case study, $\widehat{\tau}_c$ is obtained empirically, on the basis of the available PIT values. Thus, we have a consistant estimator of τ_c and one can rewrite (5) such as $\mathrm{E}_Y[L_\tau(Y, G^{-1}(Q_\tau))] - \mathrm{E}_Y[L_\tau(Y, G^{-1}(\tau))]$, with Q_τ a random variable converging in distribution to a Normal distribution with mean τ_c and a variance decreasing at the rate $\frac{1}{n}$.

In such a case, it is still possible to obtain bounds concerning the error induced by our correction.

- **A.3.2.1**: F_e and G_e are C^2 $\forall\, e \in E$.
- **A.3.2.2**: $\forall\, y \in \mathbf{R}, \forall\, e \in E$, $|F_e(y) - G_e(y)| \leq \varepsilon$, with $\varepsilon \in [0,1]$.
- **A.3.2.3**: the derivatives of G_e are lower bounded $\forall\, e \in E$, $\forall\, \tau \in [0,1]$ by $1/\xi$, on the intervals $[G_e^{-1}(0 \vee (\tau - \varepsilon)), G_e^{-1}(1 \wedge (\tau + \varepsilon))]$, with $\xi \in]0, +\infty[$.
- **A.3.2.4**: $\forall\, e \in E, \forall\, \tau \in [0,1]$, $f_e(G_e^{-1}(\tau_c)) \leq \beta$, with $\beta \in]0, +\infty[$ and f_e the derivatives of F_e.
- **A.3.2.5**: f_e are continuous over the interval $[-\infty, G_e^{-1}(\tau_c)]$ $\forall\, e \in E$ and their derivatives are bounded, i.e $\forall y \in \mathbf{R}, \forall\, e \in E$, $|f_e'(y)| \leq M$, with $M \in]0, +\infty[$ and f_e the derivative of F_e.

- **A.3.2.6**: the derivatives of g_e are bounded, i.e $\forall y \in \mathbf{R}, \forall\, e \in E, |g_e^{'}(y)| \leq \alpha$, with $\alpha \in]0, +\infty[$ and g_e the derivative of G_e.

Under the assumptions **A.2.1**, **A.3.2.1**, **A.3.2.2**, **A.3.2.3**, **A.3.2.4**, **A.3.2.5** and **A.3.2.6**, we prove (proof in Appendix):

$$\left| \mathrm{E}_Y[L_{\widehat{\tau}_c} - L_\tau] \right| \leq 2\varepsilon^2 \xi + \frac{C\lambda}{n} \text{ and} \tag{6}$$

$$|\mathrm{E}_Y[CRPS_{G,C\circ G}]| \leq 2\left(2\varepsilon^2 \xi + \frac{C\lambda}{n}\right) \tag{7}$$

with $C = \frac{(1-\tau)\alpha\xi^3}{2} + C_{int} + C_{abs}$, $C_{int} = \frac{\xi^2\beta}{2}\left[1 + \alpha\xi^2 + \frac{\alpha^2\xi^4}{4}\right]$,
$C_{abs} = \frac{M\xi^3}{6}\left[1 + \frac{3\xi^3\alpha}{2} + \frac{3\xi^3\alpha^2}{4} + \frac{\xi^3\alpha^3}{8}\right]$ and $\frac{\lambda}{n}$, the variance of Q_τ

This inequality shows that our recalibration deteriorates only slightly the quality of the forecasts in the worst case. Obviously, it also shows that our method improves only slightly the quality, but remember that our goal is to enforce the validity constraint, which is achieved.

4 Case study

We use our method on ensemble forecasts data set from the European Centre for Medium-Range Weather Forecasts (ECMWF). One can see in [4] that the statistical post-processing of the medium range ECMWF ensemble forecast has been addressed many times. The extended range (32 days instead of 10 days) has been addressed in some studies, but with the same methods and tools. We will show here that our recalibration method, despite its genericness, is competitive with a standard post-processing method. We dispose of temperature forecasts in a 3-dimensional array. The first one represents the date of forecasts delivery. The forecasts were made every Monday and Thursday from 11/02/13 to 02/02/17. Since 3 observations are missing, we have 413 dates of forecasts delivery. The second dimension is the number of the scenario in the ensemble member, and we have 51 scenarios. The third dimension is the forecast horizon. Since we have 32 days sampled with a forecast every 3 hours, it produces 256 horizons.

We study the calibration and compare the CRPS expectation using directly the ensemble forecast, the so-called Ensemble Model Output Statistics (EMOS) method and our recalibration method with a Cauchy Kernel dressing for the ensembles.
We choose a Cauchy Kernel in order to address problems with the bounds of the ensembles. Indeed, a lot of observations were out of the bounds of the ensemble,

which produces a lot of PIT with value 0 or 1. Thus, to avoid this problem, we need to use a Kernel with heavy tail.

During the last 12 years, the ECMWF has changed its models 27 times, which means a change every 162 days on average. Thus, it is important to use a train sample significally smaller than 162 days. However, it is also important to dispose of enough observations to obtain a consistant estimator of τ_c. Our method obtains good results with 30 days used for the recalibration but the algorithm to minimize in order to find the parameters of the EMOS in the *R* package *EnsembleMOS* doesn't converge if we use less than 60 days (at least with our data set). Thus, we chose to use 60 days for the recalibration.

To recalibrate the forecasts for a particular forecasting day and a particular horizon (remember that we have 256 horizons), we use the forecasts made for the same horizon, over the 60 previous dates of forecast delivery for the two methods. However, with our method, we use a linear interpolation based on the PIT series formed by these 60 previous days to recalibrate the forecasts. The linear interpolation is also used to calculate the different quantile levels when we are not working with EMOS (in that case, for the recalibration or to calculate the quantile, we use the Normale distribution with the fitted parameters). Note that the hypotheses concerning only G_e are verified $\forall\, e \in E$. Besides, even if we cannot verifiy the other hypotheses, we show expected results

Let's start with the calibration property:

Table 1 Success rate to 5% *K-S* test

	Raw Ensemble	EMOS	Our Method
Success rate in %	14	0.39	96

We have calculated the PIT series for each horizon (256), and use 5% *K-S* test for each of them. The success rate is the percentage of horizons passing the test.

As expected, we can see in table 1 that our method allows us the test of validity to be passed while the use of the raw ensemble fails. The EMOS also failed to pass the test. Clearly, our method is useful to ensure the calibration property. But how about the quality of the density forecast? In order to evaluate the impact of our correction on the forecast quality, we are interested in the CRPS expectation.

We can see in figure 1 that EMOS as well as our method are more efficient than the raw ensemble for little horizons. However, the EMOS deteriorates clearly the quality of the forecasts when the horizon grows, contrarily to our method which deteriorates only slightly the quality of the forecasts, when it is the case.

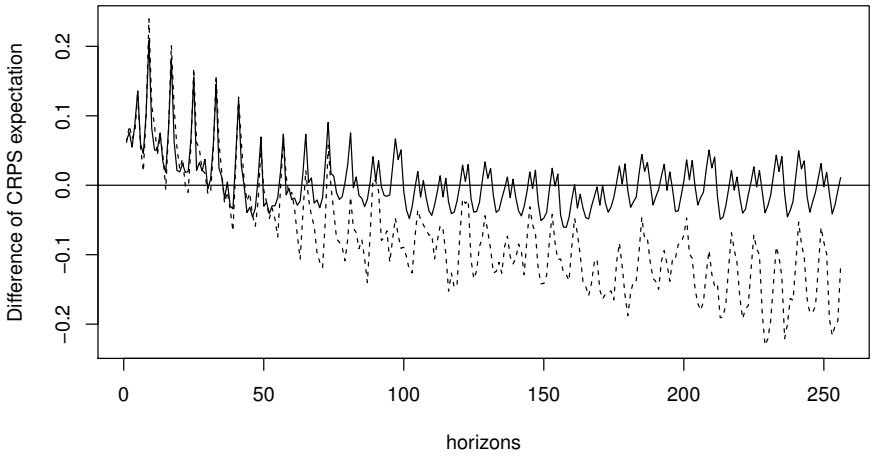


Fig. 1 Comparison of EMOS and our method CRPS expectation with that of raw ensemble The empty line corresponds to our method and the dashed one to the EMOS

Thus, this study highlights perfectly the usefulness of our method, which is very simple to use. Indeed, it shows that it allows us to ensure the validity constraint, with a limited negative impact on the quality.

References

1. Bröcker, J.: Reliability, sufficiency, and the decomposition of proper scores. Q. J. R. Meteorol. Soc. (2009) doi: 10.1002/qj.456
2. Corradi, V., and Swanson, N.R.: Predictive density evaluation. In: Elliott, G., Granger, C.W.J., and Timmermann, A. Handbook of Economic Forecasting, pp. 197-284. Elsevier, Amsterdam (2006)
3. Fortin, V., Favre, A-C., and Said, M.: Probabilistic forecasting from ensemble prediction systems: Improving upon the best-member method by using a different weight and dressing kernel for each member. Q. J. R. Meteorol. Soc. (2006) doi: 10.1256/qj.05.167
4. Gneiting, T.: Calibration of medium-range weather forecasts. In: Technical Memorandum. European Centre for Medium-Range Weather Forecasts (2014) https://www.ecmwf.int/en/elibrary/9607-calibration-medium-range-weather-forecasts
5. Gneiting, T., and Katzfuss, M.: Probabilistic forecasting. Annu. Rev. Stat. Appl. (2014) doi: 10.1146/annurev-statistics-062713-085831
6. Gneiting, T., Raftery A.E., Westveld III, A.H., and Goldman, T.: Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. Mon. Weather Rev. (2005) doi: 10.1175/MWR2904.1
7. Gneiting, T., Balabdaoui, F., and Raftery, A.E.: Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) (2007) doi: 10.1111/j.1467-9868.2007.00587.x

8. Gogonel, A., Collet, J., and Bar-Hen, A.: Improving the calibration of the best member method using quantile regression to forecast extreme temperatures. Nat. Hazards and Earth Syst. Sci. (2013) doi: 10.5194/nhess-13-1161-2013
9. Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting (2000) doi: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2
10. Krzysztofowicz, R.: Bayesian processor of output: a new technique for probabilistic weather forecasting. In: 17th Conference on Probability and Statistics in the Atmospheric Sciences. American Meteorological society (2004)
https://ams.confex.com/ams/pdfpapers/69608.pdf
11. Michelangeli, P-A., Vrac, M., and Loukos, H.: Probabilistic downscaling approaches : Application to wind cumulative distribution functions. Geophys. Res. Lett. (2009) doi: 10.1029/2009GL038401
12. Ben Taieb, S., Huser, R., Hyndman, R.J., and Genton, M.G.: Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression. IEEE Trans. on Smart Grid. **7**, 2448–2455 (2016)
13. Wilks, D.S.: Statistical methods in the atmospheric sciences. In: Wilks, D.S. International Geophysics Series Volume 100. Academic Press, Cambridge (2011)
14. Wilks, D.S.: Enforcing calibration in ensemble postprocessing. Q. J. R. Meteorol. Soc. (2017) doi: 10.1002/qj.3185

Chapter 1
A Generic Method for Density Forecasts Recalibration

5 Appendix

Here are gathered all the proofs concerning the results presented in the chapter. The first section is concerned by proofs of results in an infinite sample and the second by result in a finite sample.

Lemma 1.

$$\mathrm{E}_Y[L_\tau - L_{\tau_c}] = \sum_e p_e \int_{G_e^{-1}(\tau_c)}^{G_e^{-1}(\tau)} (F_e(y) - \tau)\mathrm{d}y\,,$$

with $\tau, \tau_c \in [0,1]$ and p_e the frequency of appearance of the state e. Under the assumption **A.2.1**, we prove Lemma 1.

Proof. We have:

$$\mathrm{E}_Y[L_\tau - L_{\tau_c}] = \sum_e p_e \left(\mathrm{E}_Y[L_\tau(\boldsymbol{Y}, G_e^{-1}(\tau))] - \mathrm{E}_Y[L_\tau(\boldsymbol{Y}, G_e^{-1}(\tau_c)]\right). \tag{8}$$

First, we only focus on a particular e. Thus, we are interested in :

$$\mathrm{E}_Y[L_\tau(\boldsymbol{Y}, G_e^{-1}(\tau))] - \mathrm{E}_Y[L_\tau(\boldsymbol{Y}, G_e^{-1}(\tau_c)] \equiv \mathrm{E}_{Y,e}[L_{\tau,\tau_c}].$$

For ease of notation and comprehension, we suppress e in the notation since there is no confusion. Moreover, we suppose, for ease of notation again (and since we

obtain the same result if we inverse the inequality) that $G^{-1}(\tau) \leq G^{-1}(\tau_c)$. So, we have:

$$\begin{aligned}
\mathrm{E}_Y[L_{\tau,\tau_c}] &= \int_{-\infty}^{+\infty} \Big([y - G^{-1}(\tau)]\,\tau + [G^{-1}(\tau) - y]\,\mathbf{1}_{\{y \leq G^{-1}(\tau)\}} \Big)\, f_Y(y)\,\mathrm{d}y \\
&\quad - \int_{-\infty}^{+\infty} \Big([y - G^{-1}(\tau_c)]\,\tau + [G^{-1}(\tau_c) - y]\,\mathbf{1}_{\{y \leq G^{-1}(\tau_c)\}} \Big)\, f_Y(y)\,\mathrm{d}y \\
&= [G^{-1}(\tau_c) - G^{-1}(\tau)]\,\tau + [G^{-1}(\tau) - G^{-1}(\tau_c)]\,F \circ G^{-1}(\tau) \\
&\quad - G^{-1}(\tau_c)\,[F \circ G^{-1}(\tau_c) - F \circ G^{-1}(\tau)] + \int_{y=G^{-1}(\tau)}^{G^{-1}(\tau_c)} \underbrace{y}_{v}\,\underbrace{f_Y(y)}_{u'}\,\mathrm{d}y\,.
\end{aligned}$$

Using integral by parts, we have:

$$\begin{aligned}
\mathrm{E}_Y[L_{\tau,\tau_c}] &= [G^{-1}(\tau_c) - G^{-1}(\tau)]\,\tau + \int_{y=G^{-1}(\tau_c)}^{G^{-1}(\tau)} F(y)\,\mathrm{d}y \\
&= \int_{y=G^{-1}(\tau_c)}^{G^{-1}(\tau)} [F(y) - \tau]\,\mathrm{d}y\,.
\end{aligned}$$

Replacing it in (8) finishes the demonstration □

5.1 Impact on score: conditions for improvement

In this section, the reader can find the proofs of results mentioned in Sect.3.1 of the chapter. We first demonstrate how to approximate the difference of L_τ expectation before showing that under some hypotheses, our correction improves systematically the quality of the forecasts.

5.1.1 Rewriting the difference of L_τ expectation

Under the assumptions **A.2.1**, **A.3.1.1**, **A.3.1.2**, **A.3.1.3** and **A.3.1.4** and using functional derivatives and the implicit function theorem, we prove (2).

Proof. Remember: Let H be a functional, h a function, α a scalar and δ an arbitrary function.
We can write the expression of the functional evaluated at $f + \delta\alpha$ as follow:

$$H[h + \delta\,\alpha] = H[h] + \frac{\mathrm{d}H[h + \delta\,\alpha]}{\mathrm{d}\alpha}|_{\alpha=0}\,\alpha + \frac{1}{2}\frac{\mathrm{d}^2H[h + \delta\,\alpha]}{\mathrm{d}\alpha^2}|_{\alpha=0}\,\alpha^2 + \cdots + \mathrm{Rem}(\alpha)\,,$$

with $\text{Rem}(\alpha)$ the remainder. Denote:

$$\begin{aligned}\Delta PL[h] &= \sum_e p_e \int_{h_e^{-1}(\tau_c)}^{h_e^{-1}(\tau)} (F_e(y) - \tau)\mathrm{d}y \\ &= \sum_e p_e \Delta PL_e[h_e]\,.\end{aligned}$$

For ease of notation, denote $\Delta PL_e[F_e + \delta_e\alpha] \equiv \Delta PL_{F,\delta,e}$. Choosing $H = \Delta PL_e$, $h = F_e$ and $\eta_e = \alpha\delta_e$ (even if we use $\alpha\delta_e$ in the developpemnt in order to use functional derivatives, directional derivatives and the implicit function theorem), we have:

$$\Delta PL_{F,\delta,e} \sim \Delta PL_e[F_e] + \frac{\mathrm{d}\Delta PL_{F,\delta,e}}{\mathrm{d}\alpha}|_{\alpha=0}\,\alpha + \frac{1}{2}\frac{\mathrm{d}^2\Delta PL_{F,\delta,e}}{\mathrm{d}\alpha^2}|_{\alpha=0}\,\alpha^2 + \text{Rem}_e(\alpha)$$

$$\begin{aligned}&= \left[\frac{\partial\Delta PL_{F,\delta,e}}{\partial\alpha}|_{\alpha=0,\,\tau_c=\tau} + \frac{\partial\Delta PL_{F,\delta,e}}{\partial\tau_c}|_{\alpha=0,\,\tau_c=\tau}\frac{\mathrm{d}\tau_c}{\mathrm{d}\alpha}\right]\alpha \\ &\quad + \left[\frac{\partial^2\Delta PL_{F,\delta,e}}{\partial\alpha^2}|_{\alpha=0,\,\tau_c=\tau} + 2\frac{\partial^2\Delta PL_{F,\delta,e}}{\partial\alpha\partial\tau_c}|_{\alpha=0,\,\tau_c=\tau}\frac{\mathrm{d}\tau_c}{\mathrm{d}\alpha}\right]\frac{\alpha^2}{2} \\ &\quad + \left[\frac{\partial^2\Delta PL_{F,\delta,e}}{\partial\tau_c^2}|_{\alpha=0,\,\tau_c=\tau}\left(\frac{\mathrm{d}\tau_c}{\mathrm{d}\alpha}\right)^2 + \frac{\partial\Delta PL_{F,\delta,e}}{\partial\tau_c}|_{\alpha=0,\,\tau_c=\tau}\frac{\mathrm{d}^2\tau_c}{\mathrm{d}\alpha^2}\right]\frac{\alpha^2}{2} \\ &\quad + \text{Rem}_e(\alpha)\,.\end{aligned}$$

To calculate $\frac{\mathrm{d}\tau_c}{\mathrm{d}\alpha}$, we will use the equation which link τ_c and α:

$$\sum_e p_e F_e \circ (F_e + \delta_e\alpha)^{-1}(\tau_c) = \tau.$$

Using the implicit function theorem, we find:

$$\frac{\mathrm{d}\tau_c}{\mathrm{d}\alpha} = \sum_e p_e \delta_e \circ F_e^{-1}(\tau)$$

Now, we need to calculate partial derivatives:

$$\frac{\partial\Delta PL_{F,\delta,e}}{\partial\alpha}|_{\alpha=0,\,\tau_c=\tau} = \frac{\partial\left(\int_{(F_e+\delta_e\alpha)^{-1}(\tau_c)}^{(F_e+\delta_e\alpha)^{-1}(\tau)}(F_e(y)-\tau)\mathrm{d}y\right)}{\partial\alpha}|_{\alpha=0,\,\tau_c=\tau} = 0;$$

$$\frac{\partial \Delta PL_{F,\delta,e}}{\partial \tau_c}|_{\alpha=0,\,\tau_c=\tau} = 0 \;;\; \frac{\partial^2 \Delta PL_{F,\delta,e}}{\partial \tau_c^2}|_{\alpha=0,\,\tau_c=\tau} = -\frac{1}{f_e \circ F_e^{-1}(\tau)};$$
$$\frac{\partial^2 \Delta PL_{F,\delta,e}}{\partial \alpha^2}|_{\alpha=0,\,\tau_c=\tau} = 0 \;;\; \frac{\partial^2 \Delta PL_{F,\delta,e}}{\partial \alpha \partial \tau_c}|_{\alpha=0,\,\tau_c=\tau} = \frac{\delta_e \circ F_e^{-1}(\tau)}{f_e \circ F_e^{-1}(\tau)}.$$

Thus, we have:

$$\begin{aligned}\Delta PL_e[F_e + \delta_e \alpha] \sim & \left[\left(\frac{\delta_e \circ F_e^{-1}(\tau)}{f_e \circ F_e^{-1}(\tau)}\right)\sum_e p_e \delta_e \circ F_e^{-1}(\tau)\right]\alpha^2 \\ & - \left[\frac{(\sum_e p_e \delta_e \circ F_e^{-1}(\tau))^2}{2 f_e \circ F_e^{-1}(\tau)}\right]\alpha^2 + \mathrm{Rem}_e(\alpha),\end{aligned}$$

and hence:

$$\begin{aligned}\Delta PL[F + \delta\alpha] \sim & \left(\sum_e \frac{p_e \delta_e(F_e^{-1}(\tau))}{f_e(F_e^{-1}(\tau))}\right)\left(\sum_e p_e \delta_e(F_e^{-1}(\tau))\right) \times \alpha^2 \\ & - \left(\sum_e \frac{p_e}{2 f_e(F_e^{-1}(\tau))}\right)\left(\sum_e p_e \delta_e(F_e^{-1}(\tau))\right)^2 \times \alpha^2 \\ & + \sum_e p_e \mathrm{Rem}_e(\alpha).\end{aligned}$$

Now, let's focus on the remainders. Following the Taylor-Lagrange inequality, if M such that $\left|\frac{\mathrm{d}^3 \Delta PL_{F,\delta,e}}{\mathrm{d}\alpha^3}\right| \leq M$ exists, we have $|\mathrm{Rem}_e(\alpha)| \leq \frac{M|\alpha^3|}{3!}$. Let' s find conditions for the existence of M. The third derivative is:

$$\begin{aligned}\frac{\mathrm{d}^3 \Delta PL_{F,\delta,e}}{\mathrm{d}\alpha^3} = & \frac{\partial \Delta PL_{F,\delta,e}}{\partial \tau_c}\frac{\mathrm{d}^3\tau_c}{\mathrm{d}\alpha^3} + 3\frac{\partial^2 \Delta PL_{F,\delta,e}}{\partial \tau_c \partial \alpha}\frac{\mathrm{d}^2\tau_c}{\mathrm{d}\alpha^2} + 3\frac{\partial^2 \Delta PL_{F,\delta,e}}{\partial \tau_c^2}\frac{\mathrm{d}^2\tau_c}{\mathrm{d}\alpha^2}\frac{\mathrm{d}\tau_c}{\mathrm{d}\alpha} \\ & + \frac{\partial^3 \Delta PL_{F,\delta,e}}{\partial \alpha^3} + 3\frac{\partial^3 \Delta PL_{F,\delta,e}}{\partial \tau_c \partial \alpha^2}\frac{\mathrm{d}\tau_c}{\mathrm{d}\alpha} + 3\frac{\partial^3 \Delta PL_{F,\delta,e}}{\partial \tau_c^2 \partial \alpha}\left(\frac{\mathrm{d}\tau_c}{\mathrm{d}\alpha}\right)^2 \\ & + \frac{\partial^3 \Delta PL_{F,\delta,e}}{\partial \tau_c^3}\left(\frac{\mathrm{d}\tau_c}{\mathrm{d}\alpha}\right)^3.\end{aligned}$$

Let's calculate the partial derivatives of order 3:

$$\frac{\partial^3 \Delta PL_{F,\delta,e}}{\partial \alpha^3}|_{\alpha=0,\,\tau_c=\tau} = 0 \; ; \; \frac{\partial^3 \Delta PL_{F,\delta,e}}{\partial \tau_c^3}|_{\alpha=0,\,\tau_c=\tau} = 2\frac{f_e^{'} \circ F_e^{-1}(\tau)}{f_e \circ F_e^{-1}(\tau)} ;$$

$$\frac{\partial^3 \Delta PL_{F,\delta,e}}{\partial \tau_c^2 \partial \alpha}|_{\alpha=0,\,\tau_c=\tau} = -2\frac{f_e^{'} \circ F_e^{-1}(\tau)}{f_e \circ F_e^{-1}(\tau)^3}(\delta_e \circ F_e^{-1}(\tau)) - 2\frac{\delta_e^{'} \circ F_e^{-1}(\tau)}{f_e \circ F_e^{-1}(\tau)^2} ;$$

$$\frac{\partial^3 \Delta PL_{F,\delta,e}}{\partial \tau_c \partial \alpha^2}|_{\alpha=0,\,\tau_c=\tau} = \frac{f_e^{'} \circ F_e^{-1}(\tau)}{f_e \circ F_e^{-1}(\tau)^3}\left(\delta_e \circ F_e^{-1}(\tau)\right)^2 - 2\frac{\delta_e^{'} \circ F_e^{-1}(\tau)}{f_e \circ F_e^{-1}(\tau)^2}\left(\delta_e \circ F_e^{-1}(\tau)\right).$$

Moreover, we have:

$$\frac{d^2\tau_c}{d\alpha^2} = \sum_e p_e \left(\frac{2\delta_e^{'} \circ F_e^{-1}(\tau) - f_e^{'} \circ F_e^{-1}(\tau)}{f_e \circ F_e^{-1}(\tau)} \right) \delta_e \circ F_e^{-1}(\tau).$$

Since η_e, its first, second and third derivatives are finite in $F_e^{-1}(\tau)$, it is also the case for δ_e and the partial derivatives are finite. Furthermore, f_e, δ_e and their derivatives are bounded (since η_e and their derivatives are bounded), which implies that the second derivatives of $\Delta PL_e[F_e + \delta_e\alpha]$ are also bounded. Thus, under these conditions, M exists. Then, we can write $\frac{\mathrm{d}^3 \Delta PL_{F,\delta,e}}{\mathrm{d}\alpha^3} = M_1\delta_e^3$ and hence $|\mathrm{Rem}_e(\alpha)| \leq \frac{|M_1||\alpha\delta_e|^3}{3!}$ which implies that $\lim \frac{\mathrm{Rem}_e(\alpha)}{(\alpha\delta_e)^2} = 0$, $\alpha\delta_e \to 0$, which shows that $\mathrm{Rem}_e(\alpha)$ is negligible compared to $\frac{\mathrm{d}^2 \Delta PL_{F,\delta,e}}{\mathrm{d}\alpha^2}$.

Moreover, since $\forall\, e \in E$ the functions F_e are C^3 and the functions f_e and their derivatives are bounded by a constant which doesn't depend on e, $\forall\, e \in E$, the development is valid for all directions and thus, since $\eta_e = G_e - F_e$, we have:

$$\mathrm{E}_Y[L_\tau - L_{\tau_c}] \sim \left(\sum_e \frac{p_e \eta_e(F_e^{-1}(\tau))}{f_e(F_e^{-1}(\tau))}\right)\left(\sum_e p_e \eta_e(F_e^{-1}(\tau))\right) - \left(\sum_e \frac{p_e}{2f_e(F_e^{-1}(\tau))}\right)\left(\sum_e p_e \eta_e(F_e^{-1}(\tau))\right)^2$$

as $max\ \eta_e \to 0$.

To finish the demonstration, remark that Lemma1 proves that:

$$\Delta PL[G] = \mathrm{E}_Y[L_\tau - L_{\tau_c}].$$

□

5.1.2 Systematic improvement of the quality

Under the assumption **A.3.1.5** or **A.3.1.6**, if $\exists\, \nu \geq 0$ (sufficiently small) $\forall\, e \in E$ $\forall\, y \in \mathbf{R}; |\eta_e(y)| \leq \nu$, we show (3) and (4):

Proof. Prove (3) is equivalent to show that $\Delta PL[G]$ is positive, and if we rewrite:

$$\Delta PL[G] \sim (2\mathrm{E}[f^{-1}\eta] - \mathrm{E}[f^{-1}]\mathrm{E}[\eta])\mathrm{E}[\eta],$$

it is clear that the assumption **A.3.1.6** ensures the positivity of $\Delta PL[G]$.

However, we need more argumentation to understand the complete utility of the assumption **A.3.1.5**. Let's look at one of the two worst cases: only two states of the world, the correlation coefficient $\rho = -1$, $\eta > 0$ (the other case is when $\rho = 1$ and $\eta < 0$) and at each bound of the support of δ and f^{-1}, there is half of the probability mass. We also consider that the ratios between max and min of the supports are equal. If we define $max_e = M$ and $min_e = \frac{M}{r}$, one has the following equation:

$$\frac{1}{2} = \frac{2(r^2+1)}{(r+1)^2} - 1.$$

Solving this equation in r produces the expected result concerning the ratio between max and min values of η and f^{-1}.

Now, let's prove (4). According to (1), we have:

$$\mathrm{E}_Y[CRPS_{G,C\circ G}] = 2\int_{-\infty}^{+\infty}\left(\int_0^1 L_\tau(y, G^{-1}(\tau)) - L_\tau(y, G^{-1}\circ C^{-1}(\tau))\mathrm{d}\tau\right) f_Y(y)\mathrm{d}y.$$

We can rewrite :

$$\begin{aligned}\mathrm{E}_Y[CRPS_{G,C\circ G}] = {} & 2\int_{-\infty}^{+\infty}\int_0^1 L_\tau(y, G^{-1}(\tau))\, f_Y(y)\,\mathrm{d}\tau\,\mathrm{d}y \\ & -2\int_{-\infty}^{+\infty}\int_0^1 L_\tau(y, G^{-1}\circ C^{-1}(\tau))\, f_Y(y)\,\mathrm{d}\,\tau\mathrm{d}y,\end{aligned}$$

and using the Fubini-Tonelli theorem, one obtains:

$$\begin{aligned} \mathrm{E}_Y[CRPS_{G,C\circ G}] &= 2\int_0^1 \mathrm{E}_Y[L_\tau - L_{\tau_c}]\mathrm{d}\tau \\ &\geq 0. \end{aligned} \tag{9}$$

□

5.2 Impact on score: bounds on degradation

Under the assumptions **A.2.1**, **A.3.2.1**, **A.3.2.2**, **A.3.2.3**, **A.3.2.4**, **A.3.2.5** and **A.3.2.6** we prove (6) and (7).

Proof. adding and substracting $\mathrm{E}_Y[L_\tau(Y,G^{-1}(\tau_c))]$ to $\mathrm{E}_Y[L_{\widehat{\tau}_c} - L_\tau]$, we obtain:

$$\begin{aligned} \mathrm{E}_Y[L_{\widehat{\tau}_c} - L_\tau] = \mathrm{E}_Y[L_\tau(Y,G^{-1}(Q_\tau))] - \mathrm{E}_Y[L_\tau(Y,G^{-1}(\tau_c))] \\ + \mathrm{E}_Y[L_\tau(Y,G^{-1}(\tau_c))] - \mathrm{E}_Y[L_\tau(Y,G^{-1}(\tau))], \end{aligned}$$

and finally:

$$\mathrm{E}_Y[L_{\widehat{\tau}_c} - L_\tau] = \mathrm{E}_{Y,e}[L_\tau(Y,G_e^{-1}(Q_\tau))] - \mathrm{E}_{Y,e}[L_\tau(Y,G_e^{-1}(\tau_c))] - \mathrm{E}_Y[L_\tau - L_{\tau_c}].$$

To begin with, we treat the third term on the right side. We have:

$$\mathrm{E}_{Y,e}[L_{\tau,\tau_c}] = \int_{y=G_e^{-1}(\tau_c)}^{G_e^{-1}(\tau)} [F_e(y) - \tau]\,\mathrm{d}y\,.$$

Using the change of variable $y = G_e^{-1}(z)$ and taking the absolute value, we find:

$$|\mathrm{E}_{Y,e}[L_{\tau,\tau_c}]| = \left| \int_{z=\tau_c}^{\tau} (F_e \circ G_e^{-1}(z) - \tau)\, \frac{1}{g_e(G_e^{-1}(z))}\,\mathrm{d}z \right|.$$

Now, one needs to distinguish two cases.

If $\tau > \tau_c$, one has:

$$|\,\mathrm{E}_{Y,e}[L_{\tau,\tau_c}]\,| = \int_{z=\tau_c}^{\tau} \left| (F_e \circ G_e^{-1}(z) - \tau)\, \frac{1}{g_e(G_e^{-1}(z))} \right| \mathrm{d}z$$

$$\leq \int_{z=\tau_c}^{\tau} \left|\, (F_e \circ G_e^{-1}(z) - \tau) \,\right| \xi \; \mathrm{d}z\,.$$

Since $|F_e(z) - G_e(z)| \leq \varepsilon,\ \forall z \in \mathbf{R}, \forall e \in E,$ one obtains $|F_e \circ G_e^{-1}(z) - z| \leq \varepsilon,$ $\forall z \in [0,1], \forall e \in E$ and then:

- if $z = \tau$, one has $\left| F_e \circ G_e^{-1}(\tau) - \tau \right| \leq \varepsilon$,
- if $z = \tau_c$, $\left| F_e \circ G_e^{-1}(\tau_c) - \tau) \right| = \left| F_e \circ G_e^{-1}(\tau_c) - \tau_c + \tau_c - \tau \right|$.

Moreover, one has:

$$\begin{aligned} |\,\tau_c - \tau\,| &= \left| \sum_e p_e \left(\tau_c - F_e \circ G_e^{-1}(\tau_c) \right) \right| \\ &\leq \sum_e p_e \left| F_e \circ G_e^{-1}(\tau_c) - \tau_c \right| \\ &\leq \varepsilon\,, \end{aligned}$$

and finally:

$$\left| F_e \circ G_e^{-1}(\tau_c - \tau) \right| \leq \left| F_e \circ G_e^{-1}(\tau_c) - \tau_c \right| + |\,\tau_c - \tau\,|$$

$$\leq 2\,\varepsilon\,.$$

One deduces, when $\tau > \tau_c$:

$$|\,\mathrm{E}_{Y,e}[L_{\tau,\tau_c}]\,| \leq 2\,(\,\tau - \tau_c\,)\,\varepsilon\,\xi\,.$$

When $\tau < \tau_c$, one obtains:

$$|\,\mathrm{E}_{Y,e}[L_{\tau,\tau_c}]\,| \leq \int_{z=\tau}^{\tau_c} \left|\, (F_e \circ G_e^{-1}(z) - \tau) \,\right| \xi \; \mathrm{d}z,$$

and using the same arguments as previously:

$$|\,\mathrm{E}_{Y,e}[L_{\tau,\tau_c}]\,| \leq 2\,(\,\tau_c - \tau\,)\,\varepsilon\,\xi.$$

Hence, one concludes that:

$$|\,\mathrm{E}_{Y,e}[L_{\tau,\tau_c}]\,| \leq 2\,|\,\tau - \tau_c\,|\,\varepsilon\,\xi.$$

To finish, replacing $\mathrm{E}_{Y,e}[L_{\tau,\tau_c}]$ in (8), we have:

$$|\,\mathrm{E}_{Y}[L_{\tau} - L_{\tau_c}]\,| \leq 2\,\varepsilon^2\,\xi.$$

Now let's focus on the remainder on the right side. First, we only focus on a particular e. Thus, we are interested in :

$$\mathrm{E}_Y[L_\tau(Y, G_e^{-1}(Q_\tau))] - \mathrm{E}_Y[L_\tau(Y, G_e^{-1}(\tau_c))] \equiv \mathrm{E}_{Y,e}[L_{\widehat{\tau}_c} - L_{\tau_c}].$$

For ease of notation and comprehension, we suppress e in the notation since there is no confusion. So, we have:

$$\begin{aligned}\mathrm{E}_Y[L_{\widehat{\tau}_c} - L_{\tau_c}] = &\left(\frac{1}{2} - \tau\right)\mathrm{E}_Y\left[G^{-1}(Q_\tau) - G^{-1}(\tau_c)\right] \\ &+\frac{1}{2}\,\mathrm{E}_Y\left[\,|Y - G^{-1}(Q_\tau)| - |Y - G^{-1}(\tau_c)|\right].\end{aligned}$$

We find:

$$\begin{aligned}\left|\,\mathrm{E}_Y[L_{\widehat{\tau}_c} - L_{\tau_c}]\,\right| \leq &\left|\frac{1}{2}\,\mathrm{E}_Y\left[\,|Y - G^{-1}(Q_\tau)| - |Y - G^{-1}(\tau_c)| - G^{-1}(Q_\tau) + G^{-1}(\tau_c)\,\right]\right| \\ &+(1-\tau)\left|\,\mathrm{E}_Y\left[G^{-1}(Q_\tau) - G^{-1}(\tau_c)\right]\,\right|.\end{aligned}$$

Let's focus on the second term on the right side. Using a Taylor series approximation around $\tau_c \in [0,1]$ and the Taylor-Lagrange formula for the remainder, one has:

$$G^{-1}(Q_\tau) = G^{-1}(\tau_c) + \frac{1}{g(G^{-1}(\tau_c))}\,(Q_\tau - \tau_c) + \frac{g^{'}(\gamma)}{g(\gamma)^3}\,\frac{(Q_\tau - \tau_c)^2}{2},$$

with $\gamma = \tau_c + (Q_\tau - \tau_c)\,\theta$, and $0 < \theta < 1$.

And so

$$(1-\tau)\left|\mathrm{E}_Y\left[G^{-1}(Q_\tau)-G^{-1}(\tau_c)\right]\right| \leq \frac{(1-\tau)\,\alpha\,\xi^3}{2}\,\frac{\lambda}{n}.$$

Now, one can study the first term on the right side. Some useful remarks before the next: one can easily see that the study of such a function can be restricted to a study on the interval $I_y :=]-\infty, G^{-1}(\tau_c)]$, since we can find results on the interval $[G^{-1}(\tau_c), \infty[$ using the same arguments.

Let's define $G^{-1}(Q_\tau) \equiv Z_\tau$, $G^{-1}_{\tau_c} \equiv G^{-1}(\tau_c)$ and $f_Y^{G^{-1}_{\tau_c}} \equiv f_Y(G^{-1}(\tau_c))$, for ease of notation.

Thus, we are interested in calculating:

$$\frac{1}{2}\int_{y=-\infty}^{G^{-1}_{\tau_c}} f_Y(y)\,\underbrace{\left(\mathrm{E}_{Z_\tau}[\,|G^{-1}_{\tau_c}-Z_\tau|+|Z_\tau-y|\,]-G^{-1}_{\tau_c}+y\right)}_{=\mathrm{E}_{Z_\tau}[\,|Z_\tau-y|-Z_\tau\,]+y}\,\mathrm{d}y. \tag{10}$$

However, the function studied in the integral is complicated to work with. So, one will prefer to use its integral version, that is,

$$\mathrm{E}_{Z_\tau}[\,|Z_\tau-y|-Z_\tau\,]+y = \int_{u=-\infty}^{y}\frac{\mathrm{d}}{\mathrm{d}u}\left(\mathrm{E}_{Z_\tau}[\,|Z_\tau-u|-Z_\tau\,]+u\right)\mathrm{d}u.$$

For the bounds of the integral, the upper one is obvious. To justify the lower one, it is important to note that $\lim \mathrm{E}_{Z_\tau}[\,|Z_\tau-y|-Z_\tau\,]+y = 0$, $y \to -\infty$.

Indeed, one has:

$$\begin{aligned}\mathrm{E}_{Z_\tau}[\,|Z_\tau-y|-Z_\tau\,]+y &= \int_{z=-\infty}^{y}(y-z)\,h(z)\,dz+\int_{z=y}^{\infty}(z-y)\,h(z)\,\mathrm{d}z\\ &\quad+\int_{z=-\infty}^{\infty}(y-z)\,h(z)\,\mathrm{d}z\\ &= \int_{z=-\infty}^{y}2(y-z)\,h(z)\,\mathrm{d}z\\ &= 2yH(y)-\int_{z=-\infty}^{y}2z\,h(z)\,\mathrm{d}z,\end{aligned}$$

with h and H the p.d.f and the c.d.f of the variable Z_t. If the variable Z_τ has a finite mean, $\lim, h(y) = 0$, $y \to -\infty$, and thus it is clear that the choice of $-\infty$ for the lower bound of the integral is the good one.

At this stage, it is not easy to see the usefulness of the transformation, but it will be after the following calculus:

$$\begin{aligned}\frac{d}{du}(\mathrm{E}_{Z_\tau}[\,|Z_\tau - u| - Z_\tau\,] + u) &= 1 + \frac{d}{du}\left(\int_{z=-\infty}^{u} (u-z)\,h(z)\,\mathrm{d}z\right) \\ &\quad + \frac{d}{du}\left(\int_{z=u}^{\infty} (z-u)\,h(z)\,\mathrm{d}z\right).\end{aligned}$$

Finally, we have:

$$\begin{aligned}\frac{d}{du}(\mathrm{E}_{Z_\tau}[\,|Z_\tau - u| - Z_\tau\,] + u) &= \int_{z=-\infty}^{u} h(z)\,dz - \int_{z=u}^{\infty} h(z)\,\mathrm{d}z + 1 \\ &= H(u) - (\,1 - H(u)\,) + 1 \\ &= 2\,H(u).\end{aligned}$$

Now, it is clear that this transformation could help us for the calculus of (10) since it is equivalent to study:

$$\int_{y=-\infty}^{G_{\tau_c}^{-1}} f_Y(y) \left(\int_{u=-\infty}^{y} H(u)\,du\right) \mathrm{d}y \equiv \text{Half Int.}$$

A difficulty remains, though. Indeed, f_Y in unknown, and in consequence, not easy to work with. That's why, at first, one will use $f_Y^{G_{\tau_c}^{-1}}$ for our calculus, and then we will study the impact of such a manipulation.

Let's start with the first task. Using an integral by part on Half Int:

$$\int_{y=-\infty}^{G_{\tau_c}^{-1}} \underbrace{f_Y^{G_{\tau_c}^{-1}}}_{u'} \underbrace{\left(\int_{u=-\infty}^{y} H(u)\,du\right)}_{v} \mathrm{d}y\,.$$

One obtains:

$$\text{Half Int} = \left[y f_Y^{G_{\tau_c}^{-1}} \left(\int_{u=-\infty}^{y} H(u)\, du \right) \right]_{y=-\infty}^{G_{\tau_c}^{-1}} - \int_{y=-\infty}^{G_{\tau_c}^{-1}} y f_Y^{G_{\tau_c}^{-1}} H(y)\, \mathrm{d}y$$

$$= \int_{u=-\infty}^{G_{\tau_c}^{-1}} f_Y^{G_{\tau_c}^{-1}} \underbrace{[G_{\tau_c}^{-1} - u]}_{u'} \underbrace{H(u)}_{v}\, \mathrm{d}u$$

$$= \left[f_Y^{G_{\tau_c}^{-1}} \left(u G_{\tau_c}^{-1} - \frac{u^2}{2} \right) H(u) \right]_{u=-\infty}^{G_{\tau_c}^{-1}} - \int_{u=-\infty}^{G_{\tau_c}^{-1}} f_Y^{G_{\tau_c}^{-1}} \left(u G_{\tau_c}^{-1} - \frac{u^2}{2} \right) h(u)\, \mathrm{d}u\,.$$

Since $\left(u G_{\tau_c}^{-1} - \frac{u^2}{2} \right) = \left(\frac{(u - G_{\tau_c}^{-1})^2}{2} - \frac{(G_{\tau_c}^{-1})^2}{2} \right)$, we have:

$$\text{Half Int} = f_Y^{G_{\tau_c}^{-1}} \left(\int_{u=-\infty}^{G_{\tau_c}^{-1}} \frac{(u - G_{\tau_c}^{-1})^2}{2} h(u)\, \mathrm{d}u \right).$$

Now, using the change of variable $G(u) = z$, a Taylor series approximation around τ_c and the Taylor-Lagrange formula, one has the following approximation for Half Int:

$$\frac{f_Y^{G_{\tau_c}^{-1}}}{2} \int_{z=0}^{\tau_c} \left[\frac{1}{g(G_{\tau_c}^{-1})^2} (z - \tau_c)^2 + \frac{g'(\gamma)}{g(G_{\tau_c}^{-1})\, g(\gamma)^3} (z - \tau_c)^3 + \frac{g'(\gamma)^2}{4 g(\gamma)^6} (z - \tau_c)^4 \right] \phi(y) \mathrm{d}y\,,$$

with ϕ the p.d.f of the random variable Q_τ. Using the Jensen inequality and since $0 \leq z \leq \tau_c$, we find:

$$|\text{Half Int}| \leq \frac{f_Y^{G_{\tau_c}^{-1}}}{2} \left[\frac{\xi^2}{2} \frac{\lambda}{n} + \frac{\alpha \xi^4}{2} \frac{\lambda}{n} + \frac{\alpha^2 \xi^6}{8} \frac{\lambda}{n} \right]$$

$$\leq \frac{1}{2} \frac{C_{int}\, \lambda}{n}.$$

Since $\frac{\lambda}{n}$, which is the variance of the random variable Q_τ, is decreasing with n, let's study:

$$\Delta_{\int f} \equiv \left| \int_{u=-\infty}^{G_{\tau_c}^{-1}} (f_Y(y) - f_Y^{G_{\tau_c}^{-1}}) \left(\int_{u=-\infty}^{y} H(u)\,\mathrm{d}u \right) \mathrm{d}y \right|.$$

Since one supports the hypothesis that $f_Y^{'}$ is bounded, using the mean value theorem, one has:

$$\Delta_{\int f} \leq \int_{y=-\infty}^{G_{\tau_c}^{-1}} | f_Y(y) - f_Y^{G_{\tau_c}^{-1}} | \left(\int_{u=-\infty}^{y} H(u)\,\mathrm{d}u \right) \mathrm{d}y$$

$$\leq \int_{y=-\infty}^{G_{\tau_c}^{-1}} M \underbrace{(G_{\tau_c}^{-1} - y)}_{u'} \underbrace{\left(\int_{u=-\infty}^{y} H(u)\,\mathrm{d}u \right)}_{v} \mathrm{d}y,$$

and thus,

$$\Delta_{\int f} \leq M \left(\left[\left(y G_{\tau_c}^{-1} - \frac{y^2}{2} \right) \int_{u=-\infty}^{y} H(u)\,\mathrm{d}u \right]_{y=-\infty}^{G_{\tau_c}^{-1}} - \int_{y=-\infty}^{G_{\tau_c}^{-1}} \left(y G_{\tau_c}^{-1} - \frac{y^2}{2} \right) H(y)\mathrm{d}y \right)$$

$$= M \left(\frac{(G_{\tau_c}^{-1})^2}{2} \int_{u=-\infty}^{G_{\tau_c}^{-1}} H(u)\,\mathrm{d}u + \int_{u=-\infty}^{G_{\tau_c}^{-1}} \left(\frac{(u - G_{\tau_c}^{-1})^2}{2} - \frac{(G_{\tau_c}^{-1})^2}{2} \right) H(u)\,\mathrm{d}u \right)$$

$$= M \int_{u=-\infty}^{G_{\tau_c}^{-1}} \underbrace{H(u)}_{v} \underbrace{\frac{(u - G_{\tau_c}^{-1})^2}{2}}_{u'} \mathrm{d}u$$

$$= M \left(\left[\frac{(u - G_{\tau_c}^{-1})^3}{6} H(u) \right]_{u=-\infty}^{G_{\tau_c}^{-1}} - \int_{u=-\infty}^{G_{\tau_c}^{-1}} \frac{(u - G_{\tau_c}^{-1})^3}{6} h(u)\,\mathrm{d}u \right).$$

Finally, we obtain with the same change of variable and Taylor approximation as previously:

$$\begin{aligned}
\Delta_{\int f} &\leq \frac{M}{6}\int_{z=0}^{\tau_c}\left[\frac{1}{g(G_{\tau_c}^{-1}))}(\tau_c - z) + \frac{g^{'}(\gamma)}{2g(\gamma)^3}(\tau_c - z)^2\right]^3 \phi(z)\,\mathrm{d}z \\
&\leq \frac{M}{6}\left[\frac{\xi^3}{2}\frac{\lambda}{n} + \frac{3\xi^3\alpha}{4}\frac{\lambda}{n} + \frac{3\xi^3\alpha^2}{8}\frac{\lambda}{n} + \frac{\xi^3\alpha^3}{16}\frac{\lambda}{n}\right] \\
&\leq \frac{1}{2}\frac{C_s\lambda}{n}.
\end{aligned}$$

Thus, one has $\left|\mathrm{E}_{Y,e}[L_{\widehat{\tau}_c} - L_{\tau_c}]\right| \leq \frac{(C_{int}+C_s)\lambda}{n}$. Since C_{int} and C_s do not depend on e, this result remains meaningful when we are interested in the conditional expectation with respect to the random variable E and so $\left|\mathrm{E}_Y[L_{\widehat{\tau}_c} - L_{\tau}]\right| \leq 2\,\varepsilon^2\,\xi + \frac{C\lambda}{n}$.

Moreover, using (9), we prove (7) □