

Error propagation and attribution in simulation-based capital models

Daniel J. Crispin 

Head of Risk Strategists, Rothesay Life Plc, London, UK
Email: dan.crispin@rothesay.com

(Received 31 March 2023; revised 10 July 2023; accepted 22 September 2023; first published online 28 November 2023)

Abstract

Calculation of loss scenarios is a fundamental requirement of simulation-based capital models and these are commonly approximated. Within a life insurance setting, a loss scenario may involve an asset-liability optimization. When cashflows and asset values are dependent on only a small number of risk factor components, low-dimensional approximations may be used as inputs into the optimization and resulting in loss approximation. By considering these loss approximations as perturbations of linear optimization problems, approximation errors in loss scenarios can be bounded to first order and attributed to specific proxies. This attribution creates a mechanism for approximation improvements and for the eventual elimination of approximation errors in capital estimates through targeted exact computation. The results are demonstrated through a stylized worked example and corresponding numerical study. Advances in error analysis of proxy models enhance confidence in capital estimates. Beyond error analysis, the presented methods can be applied to general sensitivity analysis and the calculation of risk.

Keywords: Proxy modeling; variational analysis; optimization; error analysis; simulation-based capital models

1. Approximation in Capital Models

Pension insurers are required by regulation to demonstrate resilience and solvency by holding sufficient capital to meet their insurance obligations under a range of events. The capital requirement is calculated from a distribution of possible losses formed from identifying important risk factors and determining their joint distributions under prudent assumptions.

Androschuck *et al.* (2017) describe how internal model firms often use Monte Carlo techniques to overcome difficulties in determining the loss distribution analytically. Monte Carlo methods sample the risk factor space and evaluate the associated losses to form an approximate loss distribution. Even then, this evaluation is often computationally demanding and infeasible to perform for the number of simulations required to achieve acceptable convergence. These limitations have led to the adoption of approximations of the loss calculation that are faster to calculate, commonly referred to as proxy models or simply *proxies*. Capital requirements calculated through Monte Carlo simulation with proxies are estimates that contain both statistical and approximation errors.

1.1 The use of proxies represents a potential single point of failure of internal models

While the use of proxies makes the estimation of capital in simulation-based capital models tractable, their use introduces a possibility of model failure: the capital approximation may not

be close to the capital requirement implied by the financial assumptions, stress calibrations, and model design. Proxies therefore represent a possible single point of failure of internal models.

Careful justification and validation of the use of proxies allow internal model firms to demonstrate their use is appropriate for estimating capital requirements, or for other use cases. However, there is interest in improving proxy modeling and their validation. In the public letter to internal model firms issued by the UK's Prudential Regulation Authority (PRA),¹ they recognize that proxy modeling is an area where thinking and techniques continue to evolve.

It is apparent then that the justification and validation of proxies have significant importance to both the estimation of capital and to widening possible use cases into areas such as new business pricing, liquidity management, and stress testing, among others.

1.2 Using proxies to calculate capital requirements does not make sense without error analysis

Measuring or estimating the possible error in the capital estimate, introduced by proxies, removes proxies from being a potential single point of failure. Practitioners having a measurement of potential error can assess whether the accuracy is appropriate for the use case at hand. For example, the use of proxies may be acceptable if the error analysis indicated they introduce appropriate prudence or accuracy.

Here we consider an approach to measuring proxy errors that may assist firms in remedying proxies from being considered a potential single point of failure within their internal model.

In our setting, a loss scenario is calculated as the change in own funds, defined as the change in value of assets less liabilities. We assume that the liability valuation is given by the market value of a *matching* asset portfolio, defined by a minimum cost optimization, where eligibility conditions that define a possible portfolio are posed as constraints to the optimization.

Krah *et al.* (2018) report that the main computational burden in simulation-based capital models arises from the calculation of projected liability cashflows. We say that the calculations of asset values and projected cashflows are computationally expensive, or “heavy”. Once asset values and cashflows have been calculated for a given risk scenario, the related optimization calculation to establish the liability value is computationally inexpensive, or “light”.

When proxies are used to approximate the optimization data, the approximate loss scenario is formed from the optimal value function of optimization problem perturbed by approximation error. This motivates the consideration of variational analysis of optimization problems as a tool to estimate how error propagates through the loss calculation.

When the optimal value function defining the liabilities is differentiable, explicit expressions for the derivatives can be stated, indicating that a calculus-based approach to error analysis within simulation-based capital models may be applicable in some circumstances.

1.3 Proxies represent a source of systematic error

Consideration of potential errors arising from approximations used to estimate capital requirements is of fundamental importance. Establishing whether an estimate is *prudent* or *appropriate* is not possible without consideration of potential error. To further establish whether an approximation is *good* requires quantification of the potential error.

Once potential sources of error have been recognized, and their potential size quantified, error analysis describes the process of quantifying their potential impact as they spread into subsequent calculations. Hughes & Hase (2010) provide a thorough and elementary introduction to error analysis in calculations. They introduce a taxonomy of three sources of errors that have natural parallels to the setting of proxy modeling within simulation-based capital models:

¹Letter from Sid Malik, “Proxy Modelling Survey: Best Observed Practice”, Head of Division – Life Insurance and Pensions Risk, Prudential Regulation Authority, 14 June 2019.

- Random errors – e.g., values being approximated by averages from Monte Carlo sampling.
- Systematic errors – e.g., values approximated by a proxy function.
- Mistakes – e.g., bad data points, coding errors.

They observe that error analysis is often an exercise that requires different approaches for different situations. Kirkup & Frenkel (2006) note that the terms “error” and “uncertainty” are often used interchangeably. Here we use the term *error* to refer to the difference between the calculated value and the true value. This usage follows the guidelines of International Organization for Standardization (ISO) (1993).

In our application, approximation errors arising from the use of proxies are systematic and repeatable and are not necessarily random when considered as functions of the risk factor space. We assume that data are accurate, and no errors are due to mistakes. Where an underlying model itself uses sampling or Monte Carlo simulation, the model’s output can be considered random. In our study, we assume in such cases that the model error is bounded by known values, and at the discretion of the practitioner, can be reduced to an arbitrary small amount – for example, by increasing the number of simulations used, or by refinement of approximation schemes.

We use the calculus-based approach to error analysis, since we wish to assert candidates for loss error bounds. Alternatively, a distributional approach could be taken in the analysis. This approach does not form part of this study, since we are interested in establishing analytic bounds on approximated loss scenarios. Crispin & Kinsley (2022) show that once analytic bounds on approximated loss scenarios are known, the potential error in the corresponding capital estimates may be calculated. The calculation of error bounds on capital estimates allows the practitioner to pursue error reduction and, possibly, error elimination.

1.4 Error analysis and attribution creates a mechanism for error reduction and removal

The loss error bound can be attributed to individual proxy functions, leading to a mechanism for improvements to the approximations, and when combined with the proxy error elimination technique of Crispin & Kinsley (2022), gives the possibility of completely removing proxy errors from capital estimates.

Christiansen (2008) describes examples of uses of first order sensitivity analysis within life insurance applications, referred to as the sensitivity concept. In our application, we consider the full derivative and assert the tangent plane as approximating the function at nearby points.

Here, a mathematical framework for the analysis of errors in simulation-based capital models is developed. The techniques are demonstrated with a stylized example and associated numerical study. The purpose is to demonstrate that formal error analysis may be feasible for some internal model designs. Formal methods of error analysis can have a role in the validation of capital estimates and in the communication of capital uncertainty.

1.5 Wider applications beyond error analysis

Beyond error analysis within simulation-based capital modeling, the presented methods can be applied to general sensitivity analysis and the calculation of risk within the setting of classical asset-liability-management. Many studies have presented the notion of matching as a technique to reduce exposure to interest rate changes through the formulation of a portfolio optimization problem where matching conditions are represented as optimization constraints. Examples of such studies include: Tilley (1980), Wise (1984a, 1984b), Kocherlakota *et al.* (1988), Daykin & Hey (1990), and Conwill (1991).

Consider, for example, the optimization problem formulated by Kocherlakota *et al.* (1988). They consider a minimal cost portfolio of bonds chosen to cover liability cashflows. They allow for cash deposits and for borrowing, and define matching by the requirement that the terminal

balance of payments is non-negative. Positive cash balances are assumed to accrue at a cash rate, and borrowing, represented by negative cash balances, is assumed to accrue at a higher borrowing rate. The trick used in their formulation to represent the matching portfolio as a solution to a linear optimization problem is to note that matching constraints can be written as linear expressions of separate positive and negative cash balances, with these placed as decision variables within the optimization.

The novelty of our presentation within this general application lies within the consideration of Lipschitz continuity (Lemma 4) to motivate the likely existence of the full derivative with respect to all data components of the linear optimization problem (Lemma 5).

2. Basic Computational and Approximation Concepts

In simulation-based capital models, the risk factor space is typically high dimensional, and the computational burden to calculate loss scenarios is sufficiently heavy to make full Monte Carlo simulations with exact loss calculations infeasible. We distinguish between computationally feasible and infeasible calculations based on whether it is feasible to repeatedly evaluate them within a Monte Carlo context.

Definition 1. (Heavy and light models) *A function $x : E \mapsto \mathbb{R}$ is said to be computationally inexpensive (or “light”) if given $\{\mathbf{r}_i\}_{i=1}^N \subset E$ with $N \gg 1$ it is feasible to compute $\{x(\mathbf{r}_i)\}_{i=1}^N$. Conversely, if the computation of $\{x(\mathbf{r}_i)\}_{i=1}^N$ is infeasible, we say that the function is computationally expensive (or “heavy”). In our application $\{\mathbf{r}_i\}_{i=1}^N \subset E$ represent risk factor scenarios sampled from the random variable $\mathbf{R} : \Omega \mapsto E$ where $\Omega, E \subset \mathbb{R}^m$ and $m > 0$.*

The expression $N \gg 1$ is used to mean that N is orders of magnitude bigger than 1, for example, $N = 1,000,000$, as can occur within the Monte Carlo setting of simulation-based capital models. In applications, whether a particular function is considered to be heavy or light is a matter of judgment and practical experience with the underlying models. In our application, the terminology of heavy or light is used to highlight where there is significant computational burden.

The existence of calculations, or function evaluations, which are too expensive to be used in the Monte Carlo context of simulation-based capital models, has motivated practitioners to introduce faster to calculate approximations called *proxy models*, or simply proxies.

Hursey *et al.* (2014)² define a proxy model as a model that approximates a more complex model. They further distinguish between proxy models that aim to approximate the complex model, and those that aim to emulate it, for example, by agreeing exactly with the complex model at some risk factor scenarios. For this study, we define a proxy function as follows:

Definition 2. (Proxy) *A proxy for a function is an approximation to it that is computationally inexpensive: given a function $x : E \mapsto \mathbb{R}$, a proxy function for x , written $x^* : E \mapsto \mathbb{R}$, is a function posed as an approximation to x that has the property that computations of the form $\{x^*(\mathbf{r}_i)\}_{i=1}^N$ are feasible, as well as any computations relating to its construction.*

Terminology for describing proxies is varied. The term proxy is common when discussing simulation-based capital models in actuarial literature, for example: Murphy & Radun (2021), Lazzari & Bentley (2017), Androschuck *et al.* (2017), Robinson & Elliott (2014), Hursey *et al.* (2014). Two examples of alternative terminology to proxies, describing a “model-of-a-model” are “meta-model” (Blanning, 1975) and “surrogate model” (Lin & Yang, 2020).

Alongside being computationally inexpensive, the usefulness of proxy functions depends on their quality as approximations. In Definition 2, it is not implied that the accuracy of a given proxy is necessarily adequate for any particular use case, since this assessment is a matter of modeling

²Paragraph 2.3.7.

objective and is subject to actuarial judgment. Also, during the design and implementation of proxies, their accuracy may need to be improved until sufficient quality is achieved. Indeed, a proxy may be sufficiently accurate in one setting and not in another. However, the usefulness of a proxy clearly relies upon its accuracy as an approximation, and whether its accuracy is quantified.

Estimating error across the loss distribution is crucial in determining whether the introduced approximations are appropriate for use cases, in particular the error in the 99.5% loss estimate corresponding to the capital requirements. Where the potential error in the capital estimate has been estimated and expressed as an error bound, two natural cases arise:

- The error bound implies the proxies are appropriate for their use case.
- The error bound is used to further refine the design of proxies to reduce error bounds to a level acceptable for the use case.

The error bound plays the role of a validation input that actuarial practitioners can use as part of their assessment of whether the proxies are appropriate for the use case, and in particular for the assessment of their appropriateness for estimating capital requirements.

Definition 3. (Proxy error bounds) *A proxy function $x^* : E \mapsto \mathbb{R}$ for $x : E \mapsto \mathbb{R}$ is said to have error bounds if there exists a function $\varepsilon(\cdot; x^*) : E \mapsto \mathbb{R}$ satisfying*

$$|x^*(\mathbf{r}) - x(\mathbf{r})| \leq \varepsilon(\mathbf{r}; x^*) \quad \text{for all } \mathbf{r} \in E. \quad (1)$$

The value $x^*(\mathbf{r}) - x(\mathbf{r})$ for a given $\mathbf{r} \in E$ is called the approximation (also referred to as proxy, or residual) error, and its exact value is typically unknown unless explicitly calculated. The values $x^*(\mathbf{r}) \pm \varepsilon(\mathbf{r}; x^*)$ represent upper and lower bounds on the value $x(\mathbf{r})$.

The notation for proxy error bounds introduced in Definition 3 only refers to functions that map into \mathbb{R} . This restriction avoids potential notational complexity involved with dependencies across components when working with the concept of error in higher dimensions and is sufficient for our exposition.

For practical use in error analysis within simulation-based capital modeling, a proxy error bound would also normally be required to be computationally light (Definition 1). In our application, this computational requirement is fulfilled by Assumption 2.

2.1 Direct loss approximation

The setting of this paper is the analysis of error propagation arising from the use of proxies in simulation-based capital models. One approach to using proxies is to pose a proxy for the loss calculation directly as a function of the risk factor space. Let \mathbf{r} be a risk factor scenario and let $x(\mathbf{r})$ denote the associated exact loss, and let $x^*(\mathbf{r})$ denote the associated proxy value. In this study, positive values of x represent gains, and negative values represent losses. The difference in computational burden between the heavy and light models can be shown schematically as

$$\mathbf{r} \xrightarrow{\text{heavy}} x(\mathbf{r}), \quad (2a)$$

$$\mathbf{r} \xrightarrow{\text{light}} x^*(\mathbf{r}). \quad (2b)$$

However, directly approximating loss as a function of the risk factor space may be problematic. The potentially high dimensionality of the risk factor space means it can be computationally difficult, or even impossible, to sample the risk factor sufficiently to ensure goodness of fit, and to perform validation. Problems associated with performing data analysis in high dimensions is often referred to as the curse of dimensionality. Liu & Özsü (2009) describe the curse of dimensionality as the phenomena whereby the number of samples needed to approximate an arbitrary function with a given level of accuracy grows exponentially with respect to the number of input variables

(i.e., dimensionality) of the function. They note the expression was introduced by Bellman (1961). Debie & Shafi (2019) describe the curse of dimensionality as a phenomenon that arises when analyzing data in high-dimensional spaces, challenging algorithm performance and accuracy.

The curse of dimensionality is associated with the direct approximation approach when the dimensionality of the risk factor space is high. Therefore, in such a circumstance, it can be advantageous to investigate alternative approaches involving functions that are, somehow, low-dimensional. There are many ways that a function of high dimensions could be considered low dimensional. For simplicity, we consider a function as being low dimensional if it is dependent only on a small number of risk factor components.

Hejazi & Jackson (2017) report that the primary computational difficulty in computing capital requirements often lies within the computation of liability cashflow scenarios. In alignment with this observation, here we consider the situation where asset values and projected cashflows, considered as functions of the risk factor space, are low dimensional and can be easily approximated with proxies. We further assume that once these have been approximated, the onward computation of the approximate loss scenario is light. In this setting, we may consider an indirect loss approximation.

2.2 Indirect loss approximation

Suppose, as before, that \mathbf{r} represents the risk-factor scenario. Denote by $\boldsymbol{\pi}$ a computationally expensive (heavy) function of the risk factor that determines all required information to calculate the loss scenario. This includes asset values (possibly aggregated), asset and liability cashflows and all information required to specify the matching portfolio within the loss calculation. The loss function, considered as function of the data $\boldsymbol{\pi}$, is denoted $\mathcal{X}(\boldsymbol{\pi})$. The composition of computational burdens, in terms of heavy and light models, can be shown schematically as

$$\mathbf{r} \xrightarrow{\text{heavy}} \boldsymbol{\pi}(\mathbf{r}) \xrightarrow{\text{light}} \mathcal{X}(\boldsymbol{\pi}(\mathbf{r})). \quad (3a)$$

The expression $x(\mathbf{r})$ in (2a) evaluates identically to the expression $\mathcal{X}(\boldsymbol{\pi}(\mathbf{r}))$ in (3a), that is, $x(\mathbf{r}) = \mathcal{X}(\boldsymbol{\pi}(\mathbf{r}))$. The notational difference between (2a) and (3a) is used to emphasize that the loss scenario can be calculated directly as a function of the risk factor, or indirectly via intermediate calculations such as cashflows and asset values.

The computational burden from risk factor to loss scenario is heavy due to the processing of data under the risk factor scenario. Once these have been calculated, the computational burden of calculating the corresponding loss scenario is assumed light. In order to make the computation tractable under Monte Carlo simulation, a proxy function for the data $\boldsymbol{\pi}$, denoted $\boldsymbol{\pi}^*$ is introduced, with corresponding schematics of the computational burden given by:

$$\mathbf{r} \xrightarrow{\text{light}} \boldsymbol{\pi}^*(\mathbf{r}) \xrightarrow{\text{light}} \mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r})). \quad (3b)$$

If the data proxy $\boldsymbol{\pi}^*$ approximates the exact data $\boldsymbol{\pi}$ sufficiently well, a practitioner may hope that the approximate loss $\mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r}))$ closely approximates the exact loss $\mathcal{X}(\boldsymbol{\pi}(\mathbf{r}))$ sufficiently well for practical applications. However, this approach does not address the important question of how closely the approximation matches the (possibly unknown) exact value.

A key feature of the indirect approach is the requirement to develop proxies for each data element of $\boldsymbol{\pi}$. The number of proxies that may need to be developed could be significant due to the potentially large number of data elements. However, for a given data element, the number of risk factors that it is sensitive to is likely to be low, and therefore the development of proxies is more easily achieved due to the issues of dimensionality not being present.

Here, we investigate estimation of the approximation error $|\mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r})) - \mathcal{X}(\boldsymbol{\pi}(\mathbf{r}))|$ given knowledge of the proxy errors $|\pi_i^*(\mathbf{r}) - \pi_i(\mathbf{r})|$ at each component of the data $\boldsymbol{\pi}^*(\mathbf{r}) - \boldsymbol{\pi}(\mathbf{r})$. First,

we clarify the definition of proxy functions and their assumed computational properties in the context of their role in simulation-based capital models.

The use of the full optimization for liability valuation does propagate fitting errors from the proxies to the loss scenario, but does not itself introduce any new sources of uncertainty. The use of a potentially high-dimensional loss proxy is avoided. Through a variational analysis of the optimization problem used to define total liability, we show in a realistic abstract setting that the potential size of errors propagating from data proxies through to the loss proxy may be estimated analytically.

Looking forward, we find two categories of expressions required to form the total approximate error bounds:

- the derivative of the (indirect) loss function with respect to each data component $\frac{\partial \mathcal{L}}{\partial \pi_i^*}(\boldsymbol{\pi}^*(\mathbf{r}))$,
- the component-wise error bounds $\varepsilon(\mathbf{r}, \pi_i^*)$ of the data proxy components $\pi_i^*(\mathbf{r})$.

In our study and analysis, an important assumption is made that component-wise data proxies can be considered as low-dimensional functions of the risk factor space and can be formed with known error bounds.

Assumption 1. (Data proxies) *The component-wise data proxies π_i^* of the data proxy $\boldsymbol{\pi}^*$, as shown schematically in (3b), can be considered as low-dimensional functions of the risk factor space E , and have known proxy error bounds $\varepsilon(\mathbf{r}, \pi_i^*)$, as defined in Definition 3.*

Androschuck *et al.* (2017) connect the activity of proxy model validation with proxy model error bounds. For example, they observe the common approach of measuring the maximum absolute error observed on points used in the construction of the proxy and at other points throughout the function's domain. They also highlight the usefulness of visual inspection of a proxy's fit, not only for evidence used to support validation but also as part of the communication of the validation work. They note that one of the drawbacks of visualization techniques is that they naturally tend to be useful for functions of one or two dimensions. When a function is high dimensional, effective exploration of the function's domain for out-of-sample testing also becomes increasingly impractical. Therefore, common approaches to proxy validation may only be valid in low dimensions, motivating our assumption within this analysis.

While the activity of estimating error bounds of data proxies is necessarily model dependent, it is important to note that the effectiveness of error propagation analysis may be affected by the quality of the error bounds of the data proxies. The relationship between the quality of the (approximate) loss error bounds and the (approximate) error bounds of the data proxies is explored in a sensitivity analysis in Table 1, demonstrating possible failure of approximate error bounds on losses may occur even when error bounds for data proxies are chosen prudently.

2.3 Posing error bounds through linear approximations

It is commonly known that a differentiable function can be approximated by its tangent plane in a neighborhood of the point that the derivative is taken. Here, we use the tangent plane as an approximating function and consider the potential approximation error that may arise as errors propagate through onward calculations. Once derivatives have been established, and the tangent plane proposed as an approximating function, the triangle inequality can be used to propose approximate bounds on the potential error. We show that the approximate error bound may have potential application, although we also demonstrate that the bound is still approximate and can fail. Therefore, careful validation and prudence is required before this approach is used for capital requirements modeling.

Here, we use the mean value theorem to create an analytical bound between the values of a function between two points, in order to compare the bound to an approximate bound

Table 1. The sensitivity of the success of approximate error analysis to the estimation of error bounds of data proxies within the stylized example of a simulation-based capital model.*

Error bound Scaling factor (γ)	Successful Unordered bounds (%)	Successful Ordered bounds (%)
0.85	32.1	31.2
1	95.4	100.0
1.15	97.7	100.0
1.3	99.3	100.0
1.45	100.0	100.0
1.6	100.0	100.0

The table shows the sensitivity of the success of the approximated loss error bound with respect to a scaling factor γ applied to the estimate of the data proxy error bounds. Loss and approximate bound data are from a Monte Carlo of 1,000,000 simulations from the model described in Section 5.1 with example data from Example 5.1. The proportion of successful unordered bounds is measured by the proportion of times within the Monte Carlo simulation the posed bounds in (32a) with $l_i = x^(r_i) - \gamma \varepsilon(r_i; x^*)$ and $u_i = x^*(r_i) + \gamma \varepsilon(r_i; x^*)$ hold with mathematical (nonstrict) inequality. Similarly, the analogous sensitivity to γ for ordered approximate bounds to hold with mathematical (nonstrict) inequality is shown. The success of the approximate bounds is more successful in the setting of the ordered bounds, suggesting that sorting provides additional prudence. It is important to note that, in this example, the unordered bounds fail (in the mathematical sense of inequality) in about 4.6% of cases in the base case of $\gamma = 1$ and is only fully successful when prudence to the data proxy error bounds is applied, represented by $\gamma = 1.45$. Numbers are shown rounded to 1d.p. with the underlying data for the entries marked 100.0% being exactly 100% showing for these entries all proxy losses lie within the bounds.

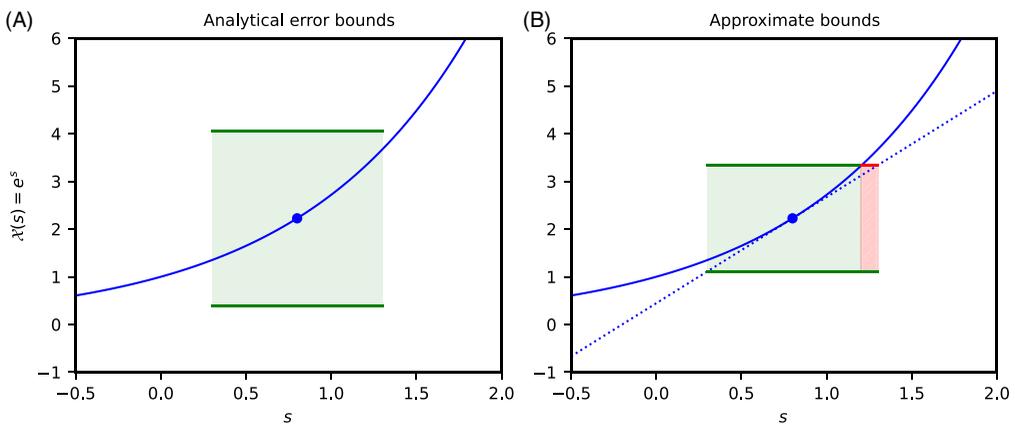


Figure 1. An illustration of analytical error bounds (Panel A) and approximate error bounds (Panel B) based on Example 2.1. The function $\mathcal{X} : \mathbb{R} \mapsto \mathbb{R}$ is defined to be $\mathcal{X}(s) = \exp(s)$, $s^* = 0.8$ and $\varepsilon = 0.5$ with s satisfying $|s - s^*| \leq \varepsilon$. The function $s \mapsto \mathcal{X}(s)$ (blue line) and the point $(s^*, \mathcal{X}(s^*))$ (blue dot) are shown identically in Panels A and B. The left and right boundaries of the (green) rectangles of Panels A and B are identical and given by $s^* \pm \varepsilon$. Panel A: The (green) rectangle depicts the feasible region for $(s, \mathcal{X}(s))$ defined by analytical error bounds in (12a) depicted as horizontal lines (green). The use of green indicates that the bounds are effective. Panel B: The (green and hatched-red) rectangle depicts the approximated feasible region for $(s, \mathcal{X}(s))$ defined by approximate error bounds in (12b). Regions where the approximated upper and lower approximate bounds fail to hold are shown in hatched-red. Note the approximate lower bound is effective while the approximate upper bound fails for values of s near to its maximum value $s^* + \varepsilon$.

developed though using the tangent as an approximation. Fig. 1 illustrates that approximate bounds developed this way can be effective, but importantly, may also fail.

In order to compare a concrete example of exact and approximate error bounds, recall the mean value theorem from elementary calculus. Here, we state the theorem in its high-dimensional form applicable to our application, where there are many data components forming input to the loss calculation. In the following, the integer d denotes the number of components of the data π .

Lemma 1. (Mean value theorem for functions of several variables) *Suppose that $\mathcal{X} : \mathbb{R}^d \mapsto \mathbb{R}$ is differentiable at each point on an open set $S \subset \mathbb{R}^d$. If $\mathbf{a}, \mathbf{b} \in S$ and $L(\mathbf{a}, \mathbf{b}) := \{\mathbf{a}t + (1 - t)\mathbf{b} : t \in (0, 1)\} \subset S$, then there exists $\mathbf{z} \in L(\mathbf{a}, \mathbf{b})$ such that*

$$\mathcal{X}(\mathbf{b}) = \mathcal{X}(\mathbf{a}) + \sum_{i=1}^d \frac{\partial \mathcal{X}(\mathbf{z})}{\partial x_i} (\mathbf{b}_i - \mathbf{a}_i). \tag{4}$$

Proof. This is a standard result of calculus, see for instance, Theorems 6–17 of Apostol (1957). □

Under the additional assumption that the partial derivatives are continuous on the closure of the interval $L(\mathbf{a}, \mathbf{b})$, written $L[\mathbf{a}, \mathbf{b}]$, there holds:

$$|\mathcal{X}(\mathbf{b}) - \mathcal{X}(\mathbf{a})| \leq \max_{\mathbf{z} \in L[\mathbf{a}, \mathbf{b}]} \sum_{i=1}^d \left| \frac{\partial \mathcal{X}(\mathbf{z})}{\partial x_i} \right| |\mathbf{b} - \mathbf{a}|. \tag{5}$$

The analytic bound (5) is considered for illustrative and comparative purposes only. It has limited use in practical applications since the evaluation point \mathbf{z} is unknown. To understand the potential size of the analytical error bound, all derivatives must be known along the line $L[\mathbf{a}, \mathbf{b}]$. Apart from offering no computational advantage to simply using the heavy model directly, even estimation of the bound may be infeasible due to the computational limitations of using the heavy model in a Monte Carlo setting. These factors motivate the consideration of approximate methods of error analysis.

Consider again the indirect loss calculation (3a) with the loss function, $\mathcal{X} : \mathbb{R}^d \mapsto \mathbb{R}$, defined on a high-dimensional domain representing the large number of data components used in the loss calculation. The data components are dependent of the risk factor scenario so can be considered functions of the risk factor space: $\pi_i : E \mapsto \mathbb{R}$ with $\boldsymbol{\pi}(\mathbf{r}) = (\pi_1(\mathbf{r}), \pi_2(\mathbf{r}), \dots, \pi_d(\mathbf{r}))$. The loss scenario can then be written $\mathcal{X}(\pi_1(\mathbf{r}), \pi_2(\mathbf{r}), \dots, \pi_d(\mathbf{r}))$ or simply $\mathcal{X}(\boldsymbol{\pi}(\mathbf{r}))$.

Suppose now that proxies π_i^* have been developed for π_i with corresponding errors bounds as known functions (Definitions 2 and 3). That is, suppose that there are functions $\varepsilon(\cdot; \pi_i^*)$ of the risk factor space satisfying:

$$|\pi_i^*(\mathbf{r}) - \pi_i(\mathbf{r})| \leq \varepsilon(\mathbf{r}; \pi_i^*). \tag{6}$$

The reasonableness of this assumption relies on two aspects. First, the data-components depend, when considered individually, only on a small number of risk factors. That is, the curse of dimensionality is not encountered. Second, the computational resources are available to design, calibrate, and validate the proxies and their error bounds.

We are interested in understanding how well $\mathcal{X}(\pi_1(\mathbf{r}), \pi_2(\mathbf{r}), \dots, \pi_d(\mathbf{r}))$ is approximated by $\mathcal{X}(\pi_1^*(\mathbf{r}), \pi_2^*(\mathbf{r}), \dots, \pi_d^*(\mathbf{r}))$. Since a differentiable function is approximated by its tangent plane at nearby points, we pose the tangent plane as an approximating function:

$$\mathcal{X}(\boldsymbol{\pi}(\mathbf{r})) - \mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r})) \approx \sum_{i=1}^d \frac{\partial \mathcal{X}}{\partial x_i}(\boldsymbol{\pi}^*(\mathbf{r})) (\pi_i(\mathbf{r}) - \pi_i^*(\mathbf{r})). \tag{7}$$

Here, the symbol $\alpha \approx \beta$ is used to denote that an expression β has been posed as an approximation to α , and importantly, without any claim to its accuracy. Through a basic application of the triangle inequality, and applying expression (6), there holds:

$$\left| \sum_{i=1}^d \frac{\partial \mathcal{X}}{\partial x_i}(\boldsymbol{\pi}^*(\mathbf{r})) (\pi_i(\mathbf{r}) - \pi_i^*(\mathbf{r})) \right| \leq \sum_{i=1}^d \left| \frac{\partial \mathcal{X}}{\partial x_i}(\boldsymbol{\pi}^*(\mathbf{r})) \right| \varepsilon(\mathbf{r}; \pi_i^*). \tag{8}$$

Following the idea that a differentiable function is approximated by its tangent plane, the value $\sum_{i=1}^d \left| \frac{\partial \mathcal{X}}{\partial x_i}(\boldsymbol{\pi}^*(\mathbf{r})) \right| \varepsilon(\mathbf{r}; \pi_i^*)$ is posed as a potential bound for $|\mathcal{X}(\boldsymbol{\pi}(\mathbf{r})) - \mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r}))|$.

Notation 1. (Posing approximate bounds) For $A, B \in \mathbb{R}$, the notation $A \approx B$ means that B is posed as an approximation to A , and the notation $A \lesssim B$ means that B is posed as a potential upper bound for A . In either usage, any of $A \leq B$ or $A > B$ may hold in practice. For example, we write

$$|\mathcal{X}(\boldsymbol{\pi}(\mathbf{r})) - \mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r}))| \lesssim \sum_{i=1}^d \left| \frac{\partial \mathcal{X}}{\partial x_i}(\boldsymbol{\pi}^*(\mathbf{r})) \right| \varepsilon(\mathbf{r}; \pi_i^*) \tag{9}$$

where, in steps (6), (7) and (8), the expression $\sum_{i=1}^d \left| \frac{\partial \mathcal{X}}{\partial x_i}(\boldsymbol{\pi}^*(\mathbf{r})) \right| \varepsilon(\mathbf{r}; \pi_i^*)$ is posed as an approximate upper bound for $|\mathcal{X}(\boldsymbol{\pi}(\mathbf{r})) - \mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r}))|$.

It is important to note that the use of the symbol \lesssim highlights that the bound (9) has not been mathematically proven and should therefore be subject to further validation before its use in capital requirements modeling.

The approximate bound (9) is posed as having potential practical applications. The approximate bound may be used as a starting point in developing prudent bounds that can be defended through validation activities. It is important to note that approximate error bounds may fail in practical conditions, for example, when the function exhibits significant curvature or when the valuation point differs significantly from the point used to establish the tangent plane, see Fig. 1. Therefore, it is important to validate approximate error bounds, or ones posed from these as prudent, before their use within capital requirement modeling. Next we consider an elementary example that illustrates how approximate error bounds may fail.

2.4 An elementary example illustrating success and failure of approximate error bounds

Here, we compare the analytic bound (5) with the approximate bound (9) in a simple setting in order to illustrate how approximate error bounds developed in this way may fail.

For this purpose, consider a one-dimensional loss function, $\mathcal{X} : \mathbb{R} \mapsto \mathbb{R}$, and its value at two points $s, s^* \in \mathbb{R}$. The value s^* represents a known value derived through a proxy, and s represents an exact value that is unknown unless computational resources are used to establish it. Suppose, mimicking (6), that there exists a proxy bound $\varepsilon > 0$, so that $|s - s^*| \leq \varepsilon$. Assuming sufficient continuity and differentiability, by (5), there holds:

Analytic: $|\mathcal{X}(s) - \mathcal{X}(s^*)| \leq \max_{s \in [s^* - \varepsilon, s^* + \varepsilon]} |\mathcal{X}'(s)| \varepsilon.$ (10)

The approximate bound, proposed in (9), becomes:

Approximate: $|\mathcal{X}(s) - \mathcal{X}(s^*)| \lesssim |\mathcal{X}'(s^*)| \varepsilon.$ (11)

The following example illustrates a comparison between the analytic expression (10) and the quantity in (11) representing an approximate error bound.

Example 2.1. Let $\mathcal{X} : \mathbb{R} \mapsto \mathbb{R}$ be given by $\mathcal{X}(s) = \exp(s)$ for all $s \in \mathbb{R}$. Let $s^* = 0.8$, and $\varepsilon = 0.5$, so that $|s - s^*| \leq \varepsilon$. In this example, the analytic bound (10) and approximate bound (11) are given for, $s \in [s^* - \varepsilon, s^* + \varepsilon]$, by

Analytic: $|\mathcal{X}(s) - \mathcal{X}(s^*)| \leq \exp(s^* + \varepsilon)\varepsilon,$ (12a)

Approximate: $|\mathcal{X}(s) - \mathcal{X}(s^*)| \lesssim |\exp(s^*)| \varepsilon.$ (12b)

The approximate error bound fails for large $s \in [s^* - \varepsilon, s^* + \varepsilon]$, but is effective for other values, see Fig. 1.

The success and failure of approximate bounds in Example 2.1 are illustrated in Fig. 1. Regions of feasibility are differentiated by colors. Green indicates success, as is guaranteed in the analytical case, and hatched-red indicated failure. The use of the triangle inequality (8) has meant, in

this example, that the approximate bound is successful over the majority of the range of s , though fails when s is near its upper boundary. In higher dimensions, where errors arising from different components may be offsetting, the triangle inequality can be expected to yield a more prudent bound. However, this example illustrates that failure can still be observed. In these instances, the approximate upper bound was close but not strict. While the bounds are posed as approximate, there may be circumstances where suitable additional prudence ensures that the bounds are effective for their use case. In these circumstances, bounds posed as prudent would still require suitable validation before their use in capital estimation.

In order to use a calculus-based approach to error analysis, we must consider the differentiability of the underlying model.

2.5 Differentiability phenomena

Classical calculus-based error analysis requires differentiability. In the setting of the indirect loss approximation, where data are proxied, the derivative of a loss scenarios with respect to its model data is a key input into a calculus-based error analysis. The question of whether a model is differentiable, considered as a function of its input data, naturally arises. In simulation-based capital models, calculation steps are complex, and analytical expressions for loss scenarios are not necessarily available. Also, it may not be straightforward to prove differentiability at any given risk factor scenario. Related to this, it may be easy to identify possible situations where differentiability may not be expected. For example, in cases where contractual cashflows depend discontinuously on risk factor values. However, when the risk factor is chosen randomly, as in our setting of simulation-based capital models, a general differentiability phenomena may arise. A striking and emblematic example of a differentiability phenomena is Lebesgue's Theorem on monotone functions.³

Theorem (Lebesgue)⁴ *Every monotone function $f: \mathbb{R} \mapsto \mathbb{R}$ possesses a finite derivative at every point $x \in \mathbb{R}$ with the possible exception of a set of points with zero Lebesgue measure, i.e., f is differentiable almost everywhere.*

Consider Lebesgue's theorem in a Monte Carlo setting. Suppose the function $f: \mathbb{R} \mapsto \mathbb{R}$ is monotone, and suppose R is a random variable with distribution μ_R that is absolutely continuous with respect to Lebesgue measure. Then, by Lebesgue's Theorem, there is zero probability that R takes a value where f is nondifferentiable. Similarly, suppose $\{r_i\}_{i=1}^N$ is an independent sample from R . Then Lebesgue's theorem implies f is differentiable at each point r_i with probability one.

Rademacher (1919) established differentiability almost everywhere for a large class of functions, applicable to our study, having a *locally Lipschitz* property.

Definition (Locally Lipschitz) *A function f from \mathbb{R}^n (or a subset of \mathbb{R}^n) into \mathbb{R} is called locally Lipschitz if for any bounded set B from the interior of the domain of f there exists $K > 0$ such that*

$$|f(x) - f(y)| \leq K\|x - y\|, \text{ for all } x, y \in B. \quad (13)$$

Theorem (Rademacher)⁵ *Let $f: \mathbb{R}^d \mapsto \mathbb{R}$ be locally Lipschitz. Then f is differentiable almost everywhere. That is, f is differentiable everywhere, except on a set of (Lebesgue) measure zero.*

³A function $f: \mathbb{R} \mapsto \mathbb{R}$ is said to be monotone if it is either entirely nondecreasing or entirely nonincreasing function.

⁴See, for example, Riesz & Nagy (1990) for a proof of Lebesgue's theorem. Lebesgue's original work appeared in 1904 – see Lebesgue (2003) for a modern and corrected printing (French).

⁵See, for example, Evans & Garzepy (2018) for a proof of Rademacher's theorem.

In the one-dimensional setting, it can be readily observed that Lebesgue’s theorem implies locally Lipschitz functions are differentiable almost everywhere. Consider, for example, the function $f : \mathbb{R} \mapsto \mathbb{R}$ assumed locally Lipschitz. Then for each bounded open interval $B \subset \mathbb{R}$, there exists $K \geq 0$ such that $|f(x) - f(y)| \leq K|x - y|$ for all $x, y \in B$. Then $x \mapsto f(x) + Kx$ is nondecreasing on B , and so, by an application of Lebesgue’s Theorem, it is differentiable almost everywhere on B . Therefore $x \mapsto f(x)$ is differentiable almost everywhere on \mathbb{R} .

In what follows, we establish that our loss function \mathcal{X} may exhibit the locally Lipschitz property when considered as a function of the data forming the loss calculation. Whenever the locally Lipschitz property holds, Rademacher’s theorem implies the loss model is differentiable almost everywhere with respect to the data. Differentiability then allows us to consider the calculus-based approach to proposing proxy error bounds.

In summary, computationally heavy models may be approximated by tangent planes in Monte Carlo settings, when it can be shown that functions exhibit the differentiability phenomenon. Looking ahead, this approach can be used to form approximate loss calculations with computationally tractable approximate error bounds.

3. A Simulation-Based Capital Model

Here, we consider a basic simulation-based capital model where loss, or equivalently, the change in own funds, is considered as a function of a risk factor space. We consider the setting of a life insurer whose liabilities are typically both long dated and nontradable where by regulation the liabilities are valued through a matching portfolio of assets held by the insurer. The concept of matching is implemented within the model as a feasible set of portfolios defined through optimization constraints. The liability value is defined to be equal to the value of a least-cost matching portfolio. Within this setting, the cost minimization of the matching portfolio ensures the insurer’s own funds are maximized, while the matching criteria ensure only a suitable portfolio can be chosen.

Definition 4. (Loss model) *Let $\mathbf{R} : \Omega \mapsto E$ be a random variable representing risk factors, with $\Omega, E \subset \mathbb{R}^m$ and $m > 0$. The loss, being the change in own funds for scenario \mathbf{r} , is denoted $x(\mathbf{r})$ and defined by*

$$\text{Scenario loss: } x(\mathbf{r}) := \mathcal{X}(\boldsymbol{\pi}(\mathbf{r})) \tag{14}$$

where: $\boldsymbol{\pi} : E \mapsto \mathbb{R}^d, d = n + nm + m$, is a model specific data function with $\boldsymbol{\pi}(\mathbf{r}) = [\mathbf{A}(\mathbf{r}), \mathbf{b}(\mathbf{r}), \mathbf{c}(\mathbf{r})]$, $\mathbf{A}(\mathbf{r}) \in \mathbb{R}^{m,n}, \mathbf{b}(\mathbf{r}) \in \mathbb{R}^m$ and $\mathbf{c}(\mathbf{r}) \in \mathbb{R}^n$; and with $\mathcal{X}, \mathcal{A}, \mathcal{L} : \mathbb{R}^d \mapsto \mathbb{R}$ defined by

$$\text{Data: } \boldsymbol{\pi} := [\mathbf{A}, \mathbf{b}, \mathbf{c}], \tag{15a}$$

$$\text{Loss: } \mathcal{X}(\boldsymbol{\pi}) := (\mathcal{A}(\boldsymbol{\pi}) - \mathcal{L}(\boldsymbol{\pi})) - (\mathcal{A}_0 - \mathcal{L}_0), \tag{15b}$$

$$\text{Assets: } \mathcal{A}(\boldsymbol{\pi}) := \mathbf{c}^T \mathbf{1}, \tag{15c}$$

$$\text{Liabilities: } \mathcal{L}(\boldsymbol{\pi}) := \inf \left\{ \mathbf{c}^T \boldsymbol{\alpha} \mid \mathbf{A}\boldsymbol{\alpha} \geq \mathbf{b}, \boldsymbol{\alpha} \geq \mathbf{0} \right\}, \tag{15d}$$

where $\mathcal{A}_0 = \mathcal{A}(\boldsymbol{\pi}(\mathbf{0}))$ and $\mathcal{L}_0 = \mathcal{L}(\boldsymbol{\pi}(\mathbf{0}))$ are constants representing the unstressed asset and liability valuations. Positive values of x represent gains, and negative values represent losses. The asset valuation (15c) is simply defined by the sum of asset values represented by the product of the transpose of the vector \mathbf{c} with the unit vector $\mathbf{1} \in \mathbb{R}^m$. The liability valuation is given by the minimum of an optimization problem (15d), where the decision variable $\boldsymbol{\alpha}$ represents portfolio allocation weights, and the data \mathbf{A} and \mathbf{b} define feasible matching portfolios through optimization constraints.

Under this definition of the loss model, without further assumptions, a minimizer of (15d) may not be obtained. And further, even if the problem does have a minimizer, it may not be unique.

It is important to note that the loss model of Definition 4 is posed as plausible and simplified representation of a life insurance loss model to demonstrate formalized methods of error analysis. Approved internal models of life insurance firms must meet all applicable regulatory requirements, and therefore loss models in practical use are likely to be more complex than the simple representation above. For example, the PRA outlined that firms must not treat the value of liabilities under stress in a purely mechanistic way.⁶ They emphasize, for example, the buy-and-hold nature of the matching portfolio, and the importance of carefully reflecting available management actions in stress-scenarios. In this setting, the matching portfolio is only to be reoptimized if matching criteria fail under stress. They further emphasize that the full range of matching criteria must be satisfied in stress conditions. These not only include quantitative criteria, such as cashflow matching tests,⁷ but also additional modeling criteria arising from risk-management use cases, such as those that promote stability of the model and matching portfolio. Firms with approved internal models may consider whether the methods of error analysis presented here in a simplified setting can be adapted to their circumstances.

We now consider the optimization of the matching portfolio within the simplified loss model (Definition 4). In the analysis of linear optimization problems, a related optimization problem, called the dual problem, plays an important role. In accordance to custom, the linear optimization problem of initial interest is referred to as the primal problem.⁸ In our setting, the primal and dual problems are related as follows:

$$\text{Primal: } \mathcal{L}(\boldsymbol{\pi}) = \inf \left\{ \mathbf{c}^T \boldsymbol{\alpha} \mid \mathbf{A}\boldsymbol{\alpha} \geq \mathbf{b}, \boldsymbol{\alpha} \geq \mathbf{0} \right\}, \tag{16}$$

$$\text{Dual: } \mathcal{D}(\boldsymbol{\pi}) = \sup \left\{ \mathbf{b}^T \boldsymbol{\lambda} \mid \mathbf{A}^T \boldsymbol{\lambda} = \mathbf{c}, \boldsymbol{\lambda} \geq \mathbf{0} \right\}, \tag{17}$$

where $\boldsymbol{\pi} = [\mathbf{A}, \mathbf{b}, \mathbf{c}]$. In the setting of linear optimization, whenever a finite optimal value exists to the primal problem, the primal and dual values are equal, and the supremum of the dual problem is attained.⁹ That is,

$$\mathcal{D}(\boldsymbol{\pi}) = \mathcal{L}(\boldsymbol{\pi}) \quad \text{whenever there exists } \boldsymbol{\alpha} \text{ such that } \mathcal{L}(\boldsymbol{\pi}) = \mathbf{c}^T \boldsymbol{\alpha}, \mathbf{A}\boldsymbol{\alpha} \geq \mathbf{b}, \boldsymbol{\alpha} \geq \mathbf{0}. \tag{18}$$

Outside of the set-builder notation in (16), we also write $\boldsymbol{\alpha}$ for the minimizer of the primal problem (16), whenever it exists and is unique. Similarly, we will denote by $\boldsymbol{\lambda}$ the maximizer of the dual problem (17), whenever it exists and is unique.

Since the value of liabilities is defined as the optimal value of an optimization problem, and that we wish to understand error propagation through calculus-based methods, we are motivated to understand the differentiability properties of the optimal value function. Freund (1985) established conditions guaranteeing differentiability of the loss function as a function of the constraint matrix \mathbf{A} . This means that partial derivatives of the loss function exist with respect to some of the components of the input data, and thereby establishing some of the information required to specify the approximating tangent plane.

Lemma 2. *(Derivatives with respect to components of A) Suppose that solutions to the primal problem (16) and dual problem (17) are attained and unique. That is, there exist unique $\boldsymbol{\alpha}$ and unique $\boldsymbol{\lambda}$ satisfying:*

⁶Paragraph 2.5 of Supervisory Statement SS8/18, Prudential Regulation Authority (2018a)
⁷Appendix 1 of Supervisory Statement SS7/18, Prudential Regulation Authority (2018b) define three cashflow matching tests called Tests 1, 2, and 3.
⁸An introduction to duality theory can be found in Section 5.2 of Boyd & Vandenberghe (2004)
⁹See Section 5.3 of Borwein & Lewis (2000).

$$c^T \alpha = \mathcal{L}(\pi), A\alpha \geq b, \alpha \geq 0, \text{ and} \tag{19}$$

$$b^T \lambda = \mathcal{D}(\pi), A^T \lambda = c, \lambda \geq 0, \tag{20}$$

where $\pi = [A, b, c]$. Then the function $A \mapsto \mathcal{L}(\pi)$ is differentiable at π with

$$\frac{\partial \mathcal{L}(\pi)}{\partial A_{i,j}} = -\alpha_i \lambda_j, \tag{21}$$

where $A_{i,j}$ denotes the elements of A .

Proof. This result was established in Freund (1985) under equality constraints. The result in our stated form follows from duality. \square

Analytical expressions for the other partial derivatives relating to components of b and c , when they exist, are also well-known and stated in Lemma 3, below.¹⁰

Lemma 3. (Derivatives with respect to components of b and c) *Suppose at $\pi = [A, b, c]$ the minimization problem in (16) is finite and attained. Suppose further that the functions $b_j \mapsto \mathcal{L}(\pi)$ and $c_i \mapsto \mathcal{L}(\pi)$ are differentiable, then:*

$$\frac{\partial \mathcal{L}(\pi)}{\partial b_j} = \lambda_j \quad \text{and} \quad \frac{\partial \mathcal{L}(\pi)}{\partial c_i} = \alpha_i. \tag{22}$$

Proof. See, for example, Section 5.6.3 of Boyd & Vandenberghe (2004) for the derivation of the derivative with respect to b . The result for c follows trivially by duality. \square

From Lemmas 2 and 3, it is clear that under the assumption of attainment and uniqueness of the primal and dual problems, and under the assumption that the partial derivatives with respect to components of b and c exist, all partial derivatives of \mathcal{L} can be calculated. However, differentiability of \mathcal{L} with respect to π , in the sense of the full derivative, must be established in order to show the tangent plane exists and can be proposed as an approximating function.

We proceed by first establishing conditions under which the optimal value function is Lipschitz continuous. Then, Radamacher’s theorem is used to establish differentiability almost everywhere. Using Lemmas 3 and 2, we then state our main result concerning the derivative of the loss function with respect to its data. To first establish the Lipschitz property, we appeal to a special case of Theorem 1 of Klatte & Kummer (1985), simplified to our linear optimization setting (16).

Lemma 4. (Lipschitz regularity of the optimal value function) *Let \mathcal{L} be the optimal value function of a linear optimization problem defined by*

$$\mathcal{L}(\pi) = \inf \left\{ c^T \alpha \mid A\alpha \geq b, \alpha \geq 0 \right\}, \quad \pi = [A, b, c], \tag{23}$$

where $A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. Let $\Psi(r)$ be the (so far, possibly empty) set of feasible points where the optimal value function is attained:

$$\Psi(\pi) = \left\{ \alpha \mid A\alpha \geq b, \alpha \geq 0, c^T \alpha = \mathcal{L}(\pi) \right\}. \tag{24}$$

Suppose for some $\pi' = [A', b', c']$ there holds:

- (i) the set $\Psi(\pi')$ is bounded and nonempty, and
- (ii) there is a point α' satisfying $A'\alpha' > b'$ and $\alpha' > 0$.

Then $\mathcal{L}(\pi)$ is locally Lipschitz (with respect to π) around π' .

¹⁰Freund (1985) cites Dinkelbach (1969) and Gal (1979) as establishing early results regarding the behavior of the optimal value function $\mathcal{L}(\pi)$ under perturbations of b and c .

Proof. This is a special case of Theorem 1 of Klatte & Kummer (1985). □

Requirement (ii) of Lemma 4 is attributed to, and referred to, as Slater’s condition. Conditions establishing regularity properties of the optimal value function, and related, are often referred to as constraint qualifications. Slater’s condition assumes the existence of an interior point, in the topological sense, of the feasible set. A related theorem establishing the Lipschitz property, but under different constraint qualification assumptions, is given in Theorem 5 of Rockafellar (1984). Lipschitz properties of optimal value functions in infinite-dimensional settings has been studied in Dempe & Mehliitz (2015) indicating the wide range of situations where the Lipschitz property of the optimal value function can be established.

Lemma 4 establishes conditions for which the optimal-value function \mathcal{L} is locally Lipschitz in a neighborhood of the data π' . Rademacher’s theorem then implies that the optimal value function is differentiable almost everywhere in that neighborhood. Importantly it does not imply that the function is known to be differentiable at the specific point π' . If it is known that conditions of Lemma 4 hold at π' , then the optimal value function is certainly differentiable at points within a small neighborhood of π' . If differentiability does hold then analytic derivatives of the loss function can be established.

Lemma 5. *(Derivatives of the loss function) Consider the loss model (Definition 4). Suppose that the primal problem (16) and dual problem (17) have finite optimal values, attained at unique points α and λ , respectively. Suppose further that the function \mathcal{L} of (15d) is differentiable at π . Then the loss value $\mathcal{X}(\pi(\mathbf{r}))$, given by (15b), is differentiable with respect to the problem data $\pi(\mathbf{r})$ with:*

$$\frac{\partial \mathcal{X}(\pi(\mathbf{r}))}{\partial A_{i,j}} = \alpha_i(\mathbf{r})\lambda_j(\mathbf{r}), \quad \frac{\partial \mathcal{X}(\pi(\mathbf{r}))}{\partial b_j} = -\lambda_j(\mathbf{r}), \quad \frac{\partial \mathcal{X}(\pi(\mathbf{r}))}{\partial c_i} = 1 - \alpha_i(\mathbf{r}). \tag{25}$$

Proof. Lemmas 2 and 3 give expressions for the partial derivatives of \mathcal{L} with respect to components of the data $\pi = [A, \mathbf{b}, \mathbf{c}]$. The partial derivatives of \mathcal{X} with respect to components of π follow trivially. □

Note that the expressions for the derivatives in (25) only involve basic algebra of the optimal portfolio vector α and optimal dual vector λ . Therefore, when the conditions of Lemma 5 hold, the derivatives can be readily calculated and used in practical settings, such as calculating approximate error bounds.

In what follows, a proxy for the loss model is defined and assumed differentiable within the Monte Carlo setting. Under the assumption of differentiability, analytical expressions for derivatives of the loss value with respect to data are found, allowing approximate error bounds on the loss approximation to be posed.

4. Proxy Model and Approximate Error Bounds

Computational limitations motivate the use of proxies in capital requirements modeling. In order to introduce such restrictions into our framework, we first formalize the notation of feasible and infeasible computations. Under the assumption of differentiability, tangent planes are then used as approximating functions to the loss function, and approximate error bounds on the proxy are posed.

In practice, whether a particular computation is feasible within a simulation setting will be dependent on the scale of the problem and the availability of computational resources. These considerations will also be shaped by the timings of associated business processes.

4.1 Computational assumptions

Androschuck *et al.* (2017) and Crispin & Kinsley (2022)¹¹ describe that practical computational limitations motivate the use of approximations in capital requirements modeling. To reflect the role computational limitations play in motivating our study, they are introduced within our framework through the following assumption.

Assumption 2. (Computational assumptions) *For $n \ll N$, we make the following computational assumptions. It is feasible to compute:*

- ✓ A large number of independent risk factor scenarios $\{\mathbf{r}_i\}_{i=1}^N$.
- ✓ Exact losses $\{x(\mathbf{r}_i)\}_{i=1}^n$ for a small number of risk factor scenarios.
- ✓ Exact data-scenarios $\{\boldsymbol{\pi}(\mathbf{r}_i)\}_{i=1}^n$, also for a small number of risk factor scenarios.
- ✓ Construction of data proxies $\boldsymbol{\pi}^*(\mathbf{r}) = [\mathbf{A}^*(\mathbf{r}), \mathbf{b}^*(\mathbf{r}), \mathbf{c}^*(\mathbf{r})]$ across all data components.

However, it is infeasible to evaluate:

- ✗ Exact losses for all the risk factor scenarios $\{x(\mathbf{r}_i)\}_{i=1}^N$.
- ✗ All exact data scenarios $\{\boldsymbol{\pi}(\mathbf{r}_i)\}_{i=1}^N$.

The infeasibility of using the heavy model in large-scale Monte Carlo simulations motivates use of approximations. It is feasible to evaluate:

- ✓ Approximate data scenarios $\{\boldsymbol{\pi}^*(\mathbf{r}_i)\}_{i=1}^N$.
- ✓ Approximate losses $\{\mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r}_i))\}_{i=1}^N$, optimal portfolios $\{\boldsymbol{\alpha}(\boldsymbol{\pi}^*(\mathbf{r}_i))\}_{i=1}^N$ defined in (16), optimal dual values $\{\lambda(\boldsymbol{\pi}^*(\mathbf{r}_i))\}_{i=1}^N$ defined in (17).

A consequence of Assumption 2 and Lemma 5 is that the derivatives of the loss function with respect to data, shown in (25), are feasible to compute for a large number of risk factor scenarios, whenever the scenario data are available.

4.2 Definition of the proxy loss function

Here, we define the data proxy $\boldsymbol{\pi}^*$ and proxy loss function x^* of the indirect approach (3b). When considering proxies for data elements of the loss calculation, we recall that we are motivated to consider this approach, since when taken individually the components of the data are only dependent on a small number of components of the risk factors. The data components can therefore be considered as low-dimensional functions, and therefore development of proxies for these are not exposed to the curse of dimensionality. This motivates our assumption that proxies for data components can be developed with appropriately validated proxy error bounds.

Assumption 3. (Data proxy) *Consider the loss model (Definition 4) and suppose there exists a proxy $\boldsymbol{\pi}^*$ (Definition 2) with error bounds (Definition 3) for the scenario-data function $\boldsymbol{\pi}$ defined in (15a). That is, suppose there exists computationally light function (Definition 1) $\boldsymbol{\pi}^* : E \mapsto \mathbb{R}^{n+mn+m}$ satisfying:*

$$\text{Data proxy: } \boldsymbol{\pi}^*(\mathbf{r}) = [\mathbf{A}^*(\mathbf{r}), \mathbf{b}^*(\mathbf{r}), \mathbf{c}^*(\mathbf{r})] \tag{26}$$

for all $\mathbf{r} \in E$, where components of the data proxies are written: $A_{i,j}^(\mathbf{r})$, $b_j^*(\mathbf{r})$ and $c_i^*(\mathbf{r})$. Suppose further that the components of the data proxy have computationally light error bounds. That is, there exist computationally light functions $\varepsilon(\cdot; A_{i,j}^*)$, $\varepsilon(\cdot; b_j^*)$, $\varepsilon(\cdot; c_i^*) : E \mapsto \mathbb{R}$, satisfying:*

¹¹ Assumption 2 and discussion within Section 2.6

$$|A_{ij}(\mathbf{r}) - A_{ij}^*(\mathbf{r})| \leq \varepsilon(\mathbf{r}; A_{ij}^*), \tag{27a}$$

$$|b_j(\mathbf{r}) - b_j^*(\mathbf{r})| \leq \varepsilon(\mathbf{r}; b_j^*), \tag{27b}$$

$$|c_i(\mathbf{r}) - c_i^*(\mathbf{r})| \leq \varepsilon(\mathbf{r}; c_i^*), \tag{27c}$$

for all $\mathbf{r} \in E$.

The practicality of Assumption 3 will depend on whether proxies and their bounds can be asserted and validated for all the required scenario-data inputs. In our exposition, it is assumed that the computational demand to fit and validate the data proxies are within practical computational limitations. Assumption 3 may not be reasonably expected to hold in situations where data components are essentially high dimensional and depend on many dimensions of the risk-factor space.

Definition 5. (Proxy loss model) Consider the loss model (Definition 4) and recall the (exact) loss function $x : E \mapsto \mathbb{R}$ defined in (14). The function $x^* : E \mapsto \mathbb{R}$ defined by

$$x^*(\mathbf{r}) := \mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r})), \tag{28}$$

is posed a proxy function for x , where \mathcal{X} is defined in (15b) and $\boldsymbol{\pi}^*$ is defined in (26).

The fact that the proxy x^* is computationally light follows from Assumptions 2 and 3. The utility of x^* as a proxy function will depend on whether error bounds can be established. In forming error bounds, we note that the differentiability phenomena may also occur when using proxied data. In such situations, analytical derivative expressions can be established in a similar manner to Lemma 5.

Lemma 6. (Derivatives of the proxy loss function) Consider the loss model (Definition 6). Suppose that the primal problem (16) and dual problem (17) have finite optimal values, attained at unique points $\boldsymbol{\alpha}^*$ and $\boldsymbol{\lambda}^*$, respectively. Suppose further that the function \mathcal{L} of (15d) is differentiable at $\boldsymbol{\pi}^*$. Then the approximate loss value $\mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r}))$, given by (15b), is differentiable with respect to the problem data $\boldsymbol{\pi}^*(\mathbf{r})$ with:

$$\frac{\partial \mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r}))}{\partial A_{ij}^*} = \alpha_i^*(\mathbf{r}) \lambda_j^*(\mathbf{r}), \quad \frac{\partial \mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r}))}{\partial b_j^*} = -\lambda_j^*(\mathbf{r}), \quad \frac{\partial \mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r}))}{\partial c_i^*} = 1 - \alpha_i^*(\mathbf{r}). \tag{29}$$

Proof. This follows exactly from Lemma 5 with the symbols replaced with their proxied values: $\boldsymbol{\pi}^*(\mathbf{r}), A^*(\mathbf{r}), \mathbf{b}^*(\mathbf{r}), \mathbf{c}^*(\mathbf{r}), \boldsymbol{\alpha}^*(\mathbf{r}), \boldsymbol{\lambda}^*(\mathbf{r})$ replace $\boldsymbol{\pi}(\mathbf{r}), \mathbf{A}(\mathbf{r}), \mathbf{b}(\mathbf{r}), \mathbf{c}(\mathbf{r}), \boldsymbol{\alpha}(\mathbf{r}), \boldsymbol{\lambda}(\mathbf{r})$. □

Before moving onto proposing approximate error bounds for the proxy loss function, it is useful to consider some limitations of the approach so far discussed. We have appealed to Rademacher’s theorem in order to justify a calculus-based approach to error analysis. This required establishing the Lipschitz property for the optimal value function defining the liabilities. However, Rademacher’s theorem only establishes differentiability almost everywhere in a neighborhood. Specifically, for a given model $\mathbf{r} \mapsto \boldsymbol{\pi}(\mathbf{r})$, it is not directly implied that the loss function $\mathcal{X}(\boldsymbol{\pi}(\mathbf{r}))$ is differentiable with respect to $\boldsymbol{\pi}(\mathbf{r})$ for almost all \mathbf{r} . For future research, it may be interesting to investigate whether an extension of Stepanov’s theorem¹² to functions of the extended reals could be used to infer differentiability for almost all \mathbf{r} under the conditions of Lemma 4 applied to $\boldsymbol{\pi}(\mathbf{r})$.

In the following, we appeal to the locally Lipschitz property to motivate the use of derivative expressions for posing of error bounds. The expression for the partial derivatives of \mathcal{X} with respect to components A_{ij} of \mathbf{A} required attainment and uniqueness of the primal problem (16) and dual

¹²See, for example, Maly (1999)

problem (17). When primal optimality is not attained, it may indicate that improvements to the model design could be made, as nonattainment indicates no optimal matching portfolio exists. However, uniqueness may not hold in practical applications, even when a linear optimization has been used in the model specification. Practitioners faced with nonuniqueness may consider the following options:

- In certain circumstances, it may be possible to pose an economically equivalent formulation of the model that exhibits uniqueness.
- Consider adding a small positive-definite quadratic penalization term of the decision variables to the objective function. This achieves uniqueness and the Lipschitz property of the optimal value function, but at the cost of having to reformulate equations for the derivatives used within the error analysis.
- Motivated by Theorem 5.1 of De Wolf & Smeers (2021), concerning the generalized derivative of the optimal value function with respect to A , practitioners may explore accepting nonuniqueness and treat derivative calculations as elements of a generalized *subgradient* derivative. In this setting, feasible solutions can be explored to find maximum approximate bounds.

Next, we consider the specific form of the loss model, with its associated derivatives, to pose approximate error bounds on the proxy loss.

4.3 Posing approximate error bounds for the proxy loss

Suppose the loss function \mathcal{X} is differentiable with respect to the approximate data $\boldsymbol{\pi}^*(\mathbf{r})$ for almost all $\mathbf{r} \in E$. In the following development of approximate error bounds, we recall that the symbols \approx and \lesssim , defined in Notation 1, mean that the approximations are being posed as having potential practical application, and that strict mathematical equality or inequality is neither claimed nor proven. By definition (28), the exact, but unknown, approximation error is given by

$$|x^*(\mathbf{r}) - x(\mathbf{r})| = |\mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r})) - \mathcal{X}(\boldsymbol{\pi}(\mathbf{r}))|. \tag{30a}$$

By posing the tangent plane as an approximating function, we may equivalently write,

$$|x^*(\mathbf{r}) - x(\mathbf{r})| \approx \left| \sum_{i,j} \frac{\partial \mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r}))}{\partial A_{i,j}^*} (A_{i,j}^* - A_{i,j}) + \sum_j \frac{\partial \mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r}))}{\partial b_j^*} (b_j^* - b_j) + \sum_i \frac{\partial \mathcal{X}(\boldsymbol{\pi}^*(\mathbf{r}))}{\partial c_i^*} (c_i^* - c_i) \right|. \tag{30b}$$

By bounding the approximation through the triangle inequality, using the expressions (29) for the partial derivatives, and applying known bounds (27a), (27b) and (27c) on the data proxies from Assumption 3, we may pose an approximate error bound for the loss function proxy:

$$|x^*(\mathbf{r}) - x(\mathbf{r})| \lesssim \sum_{i,j} \alpha_i^*(\mathbf{r}) \lambda_j^*(\mathbf{r}) \varepsilon(\mathbf{r}; A_{i,j}^*) + \sum_j \lambda_j^* \varepsilon(\mathbf{r}; b_j^*) + \sum_i |1 - \alpha_i^*| \varepsilon(\mathbf{r}; c_i^*). \tag{31}$$

Observe also that the error bound (31) is local – that is, it changes across the risk factor space E . The local refinement of the error estimate is a desirable feature of the method. However, in some circumstances, the approximation may not be effective. These may include: cases where data error bounds have not been chosen prudently, cases where the actual errors in the data approximation are significant, and in cases where the matching portfolio has been misidentified.

4.4 Approximate errors bounds for percentile estimators

We now consider the approximate error bound in the Monte Carlo context of a simulation-based capital model with the use case of estimating capital requirements. Take, for example, a simulation-based capital model that has been calibrated to form the one-year horizon loss distribution. Under Solvency II, the capital requirement is then given by the 99.5% percentile loss. Given an independent sample of the risk factor space $\{\mathbf{r}_i\}_{i=1}^N$, we may consider forming approximate error bounds on capital estimates as follows. For each risk factor scenario, calculate the loss proxy and the estimated error bound and use (31) to pose approximate lower and upper bounds on the (unknown and exact) loss scenarios:

$$l_i \lesssim x_i \lesssim u_i \tag{32a}$$

where the approximate lower and upper bounds on the exact loss are given by

$$x_i := x(\mathbf{r}_i) \tag{32b}$$

$$l_i := x^*(\mathbf{r}_i) - \varepsilon(\mathbf{r}_i; \mathbf{x}^*) \tag{32c}$$

$$u_i := x^*(\mathbf{r}_i) + \varepsilon(\mathbf{r}_i; \mathbf{x}^*) \tag{32d}$$

with $\varepsilon(\mathbf{r}; \mathbf{x}^*)$, for $\mathbf{r} \in E$, denoting the upper bound in (31) so that:

$$\varepsilon(\mathbf{r}; \mathbf{x}^*) := \sum_{ij} \alpha_i^*(\mathbf{r}) \lambda_j^*(\mathbf{r}) \varepsilon(\mathbf{r}; A_{ij}^*) + \sum_j \lambda_j^* \varepsilon(\mathbf{r}; b_j^*) + \sum_i |1 - \alpha_i^*| \varepsilon(\mathbf{r}; c_i^*). \tag{32e}$$

Consider the lists $\{l_i\}_{i=1}^N$, $\{x_i\}_{i=1}^N$ and $\{u_i\}_{i=1}^N$ individually and sort them each into increasing order, written $\{l_{(i)}\}_{i=1}^N$, $\{x_{(i)}\}_{i=1}^N$ and $\{u_{(i)}\}_{i=1}^N$ where $l_{(i)} \leq l_{(i+1)}$, $x_{(i)} \leq x_{(i+1)}$ and $u_{(i)} \leq u_{(i+1)}$ for all $i = 1, \dots, N - 1$.

Following the approach of Lemma 2 of Crispin & Kinsley (2022), we pose the ordered approximate bounds as having potential practical application in estimating bounds on the unknown ordered losses:

$$l_{(i)} \lesssim x_{(i)} \lesssim u_{(i)}, \tag{33}$$

for $i = 1, \dots, N$. In particular, consider the basic percentile L-estimator $\xi = x_{(k)}$, where $k = \lceil 0.005 \times N \rceil$ is the index associated with the 99.5% loss, due to our convention that positive values of x are gains in own funds. Under our sign convention, an estimate of the capital requirements is given by $-\xi$. The value of the estimator is unknown, since $\{x_i\}_{i=1}^N$ is unknown and is infeasible to calculate (Assumption 2). However, under the assumption that the approximate error bounds are mostly effective, or represent bounds to first order, we may pose approximate bounds on the percentile estimator:

$$l_{(k)} \lesssim \xi \lesssim u_{(k)}. \tag{34}$$

Note, as before, that the symbol \lesssim does not imply that the bounds are mathematically true, only that the expressions are being posed as potential bounds.

4.5 Error attribution creates a mechanism for proxy model refinement and error reduction

Consider, as an example, the situation where a practitioner has identified that the error bounds calculated in (32a) and (33) and aims to further improve the estimate of the 99.5% loss through refinement of the data proxies.

For single loss scenario i , the summation form of the error term in (32e) can be used in a natural manner to attribute the approximate error $\varepsilon(\mathbf{r}_i; \mathbf{x}^*)$ across each of the data proxies A_{ij}^* , b_j^* , and c_i^* : the proxy A_{ij}^* can be attributed the error $\alpha_i^*(\mathbf{r}_i) \lambda_j^*(\mathbf{r}_i) \varepsilon(\mathbf{r}_i; A_{ij}^*)$, and similarly for the proxies b_j^*

and c_i^* . This attribution can be used to identify proxies with disproportionate attributed error for improvement.

When considering improvements to error estimates of ordered losses $x_{(k)}$ both upper and lower bounds must be considered separately, since the index that sorts lower and upper approximate bounds may not be equal. One approach to refinement is to consider attribution and refinement to the approximate error interval formed from the approximate lower and upper bounds as follows. Suppose $u_{(k)} = u_{i_k}$ for some i_k and $l_{(k)} = l_{j_k}$ for some j_k . Then from (32a) there holds:

$$u_{(k)} - l_{(k)} = \underbrace{x^*(r_{i_k}) - x^*(r_{j_k})}_{\text{Non-attributable}} + \underbrace{\varepsilon(r_{i_k}; x^*)}_{\text{Attributable}} - \underbrace{\varepsilon(r_{j_k}; x^*)}_{\text{Attributable}}. \tag{35}$$

The first term, $x^*(r_{i_k}) - x^*(r_{j_k})$, contains no direct proxy error, and so cannot be improved or attributed to individual proxies by direct methods; proxy error manifests in this expression indirectly through the possible misidentification of the true risk factor index corresponding to $x_{(k)}$. The second and third expressions, given by $\varepsilon(r_{i_k}; x^*)$ and $\varepsilon(r_{j_k}; x^*)$ respectively, may however be attributed to individual proxies using their summation-form as above. With a (partial) attribution of approximate error to each data proxy in place, proxies with outsized error attribution can be chosen for improvement. The approximate error bound on the proxy loss function in (31) exhibits some intuitive properties that assist refinement by attributing zero error to proxies in many circumstances:

- Assets assigned to the matching portfolio satisfy $\alpha_i^* = 1$. For such assets, errors in their valuation data c_i^* are not propagated. The approximation errors in asset values of assets assigned to the matching portfolio offset since they occur on both asset and liability sides of the balance sheet.
- Assets not assigned to the matching portfolio satisfy $\alpha_i^* = 0$. For such assets, the approximation errors $A_{i,j}^*$ are not propagated.
- Nonbinding constraints are identified by having a zero dual value: $\lambda_j^* = 0$. Therefore, whenever the j th constraint is not binding, errors associated with b_j^* and $A_{i,j}^*$ are not propagated.

Improvements to proxies may be made in different ways. Analysis of an individual data proxy may identify that a different proxy technique is desirable. For example, a proxy may be chosen to exhibit a certain discontinuity to accommodate discontinuous relationships between data and risk factors. Alternatively, the family of functions the proxy is drawn from may be suitable, and the proxy improved with the use of extra fitting points, lowering the associated proxy error estimate. Whether refinement of the chosen data proxies increases computational burden will depend on the form of improvement chosen.

4.6 Use of approximate error bounds in practice

Practitioners may face several situations where approximate bounds on capital estimates could be of assistance:

- Approximate bounds may give evidence that existing capital estimates are effective by strengthening the body of evidence supporting their effectiveness in their use cases.
- Prudence can be applied to the approximate bounds. If these prudent bounds are suitably validated, exact methods of error analysis can be performed. Lemma 1 and Theorem 1 of Crispin & Kinsley (2022) imply that the proxy error in the capital estimate can be bounded, and the method of targeted exact computation may apply resulting in the elimination of proxy error from the estimate.
- Approximate bounds may be attributed across the individual data proxies to identify where improvements to data proxies may be most useful, and then to use improved proxies to further refine the approximate capital bounds (Section 4.5).

- The proxy loss model and approximate error bounds could be used to cross-validate separately developed proxies, including those aimed at directly approximating the loss function whose development and validation may have encountered the curse-of-dimensionality.

If approximate error bounds are not in agreement with existing capital estimates, this may justify further validation work on production proxy models. Next we consider a stylized example, chosen for its simplicity to illustrate posing approximate error bounds.

5. Stylized Example

In this section, we consider the stylized example of matching optimization from Berry & Sharpe (2021), slightly modified from a penalized quadratic problem into a linear programming setting, and extended to a stylized simulation-based capital model by allowing the input data to vary according to a vector of risk factors. The novelty of our presentation is the consideration of approximate data, and the analysis of the subsequent error propagation.

As noted in Section 3, any example of a loss model following Definition 4 is likely not fully reflective of regulatory requirements applicable to any given internal model firm. While the example presented is a simplification, it provides evidence that the methods used to analyze error propagation may have potential application.

5.1 Model definition

Let \mathbf{r} denote a risk factor scenario, a realization from the random variable \mathbf{R} . Let T denote the year-end corresponding to the maximum cashflow date arising from assets or liabilities, and let $\{t_j\}_{j=0}^m$ be the year-end dates such that $t_0 < t_1 < \dots < t_m = T$. Period j refers to the time interval $[t_{j-1}, t_j)$. Assets are indexed by $i = 1, \dots, n$, with market value $v_i(\mathbf{r})$, and have risk-adjusted expected cashflows in period j denoted $C_{i,j}(\mathbf{r})$. Let l_j be the expected liability cashflow for period j , where positive values represent monies that must be paid (negative liabilities represent monies to be received). Denote the discount curve for tenor t by $DF(0, t)(\mathbf{r})$.

A matching portfolio is formed from assigning it a portion of each asset. Denote by $\alpha_i \in [0, 1]$ the proportion of asset i assigned to the matching portfolio. Then the matching portfolio can be represented by the vector $[\alpha_1, \alpha_2, \dots, \alpha_n]^T$ and has market value given by $\sum_{i=1}^n \alpha_i v_i(\mathbf{r})$.

The notion of matching is captured by placing constraints on the allowable portfolio. Here we follow the example of Berry & Sharpe (2021), where the PRA’s Test 1 and Test 3 matching constraints are formulated as follows.

First, denote the present value of the accumulated asset and liability cashflows, up to period j , by $C_{i,j}^{\rightarrow}(\mathbf{r})$ and $l_j^{\rightarrow}(\mathbf{r})$, respectively:

$$C_{i,j}^{\rightarrow}(\mathbf{r}) = \sum_{k=1}^j C_{i,k} DF(0, t_k)(\mathbf{r}), \tag{36}$$

$$l_j^{\rightarrow}(\mathbf{r}) = \sum_{k=1}^j l_k DF(0, t_k)(\mathbf{r}), \tag{37}$$

where $DF(0, t_k)(\mathbf{r})$ is a (stochastic) discount factor, specified later in (47), and the arrow notation denotes the cumulative present values. Then the Test 1 and Test 3 matching constraints are represented by

$$\text{Test 1: } \sum_{i=1}^n \alpha_i C_{i,j}^{\rightarrow}(\mathbf{r}) - l_j^{\rightarrow}(\mathbf{r}) \geq -0.03 l_m^{\rightarrow}(\mathbf{r}), \quad j \in \{1, 2, \dots, m-1\}, \tag{38}$$

$$\text{Test 3: } \sum_{i=1}^n \alpha_i C_{i,m}^{\rightarrow}(\mathbf{r}) - l_m^{\rightarrow}(\mathbf{r}) \geq 0. \tag{39}$$

Note that an m -th Test 1 constraint would be redundant due to the Test 3 constraint, so is omitted. Note also that this example is further stylized by the omission of the PRA’s Test 2, and by its simplification of the Test 1 to one-sided inequality constraints. The insurer minimizes the cost of the matching portfolio by solving the corresponding linear optimization problem:

$$\text{minimise } \sum_{i=1}^n \alpha_i v_i(\mathbf{r}) \tag{40a}$$

$$\text{subject to } \begin{cases} \sum_{i=1}^n \alpha_i C_{i,j}^{\rightarrow}(\mathbf{r}) - l_j^{\rightarrow}(\mathbf{r}) \geq -0.03 l_m^{\rightarrow}(\mathbf{r}), j \in \{1, 2, \dots, m-1\}, \\ \sum_{i=1}^n \alpha_i C_{i,m}^{\rightarrow}(\mathbf{r}) - l_m^{\rightarrow}(\mathbf{r}) \geq 0, \\ 0 \leq \alpha_i \leq 1, i \in \{1, 2, \dots, n\}. \end{cases} \tag{40b}$$

This may be written in standard form as

$$\text{minimise } \mathbf{c}^T(\mathbf{r})\boldsymbol{\alpha} \tag{41a}$$

$$\text{subject to } \mathbf{A}(\mathbf{r})\boldsymbol{\alpha} \geq \mathbf{b}(\mathbf{r}), \boldsymbol{\alpha} \geq \mathbf{0} \tag{41b}$$

with $\mathbf{A}(\mathbf{r}) \in \mathbb{R}^{m+n,n}$, $\mathbf{b}(\mathbf{r}) \in \mathbb{R}^{m+n}$, $\mathbf{c}(\mathbf{r}) \in \mathbb{R}^n$ and $\boldsymbol{\alpha} \in \mathbb{R}^n$ where

$$\mathbf{A}(\mathbf{r}) = \begin{bmatrix} C_{1,1}^{\rightarrow}(\mathbf{r}) & C_{2,1}^{\rightarrow}(\mathbf{r}) & \cdots & C_{n,1}^{\rightarrow}(\mathbf{r}) \\ C_{1,2}^{\rightarrow}(\mathbf{r}) & C_{2,2}^{\rightarrow}(\mathbf{r}) & \cdots & C_{n,2}^{\rightarrow}(\mathbf{r}) \\ \vdots & \vdots & & \vdots \\ C_{1,m-1}^{\rightarrow}(\mathbf{r}) & C_{2,m-1}^{\rightarrow}(\mathbf{r}) & \cdots & C_{n,m-1}^{\rightarrow}(\mathbf{r}) \\ C_{1,m}^{\rightarrow}(\mathbf{r}) & C_{2,m}^{\rightarrow}(\mathbf{r}) & \cdots & C_{n,m}^{\rightarrow}(\mathbf{r}) \\ -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ & & \ddots & \\ & & & -1 \end{bmatrix}, \mathbf{b}(\mathbf{r}) = \begin{bmatrix} l_1^{\rightarrow}(\mathbf{r}) - 0.03 l_m^{\rightarrow}(\mathbf{r}) \\ l_2^{\rightarrow}(\mathbf{r}) - 0.03 l_m^{\rightarrow}(\mathbf{r}) \\ \vdots \\ l_{m-1}^{\rightarrow}(\mathbf{r}) - 0.03 l_m^{\rightarrow}(\mathbf{r}) \\ l_m^{\rightarrow}(\mathbf{r}) \\ -1 \\ -1 \\ \vdots \\ -1 \end{bmatrix}$$

$$\mathbf{c}(\mathbf{r}) = [v_1(\mathbf{r}), v_2(\mathbf{r}), \dots, v_n(\mathbf{r})]^T \text{ and } \boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_n]^T. \tag{41c}$$

Up to now, the model closely follows that given by Berry & Sharpe (2021). We now extend the example by specifying how the asset values, cashflows, and liabilities may depend on risk factor components. The exact forms have been chosen for simplicity, rather than realism, to aid the explanation of the techniques.

Denote the components of the risk factor \mathbf{r} by $[r_0, s_1, \dots, s_n]^T$. The first component r_0 represents an interest rate risk factor, and the subsequent components represent spread risk factors, in the following sense. Let the asset prices $v_i(\mathbf{r})$ be given by

$$v_i(\mathbf{r}) = \sum_{j=1}^m C_{i,j} \exp(-(\bar{s}_i + s_i)j) \tag{42}$$

where the initial spread \bar{s}_i is implicitly defined by

$$v_i = \sum_{j=1}^m C_{i,j} \exp(-\bar{s}_i j) \tag{43}$$

where v_i is the current value of asset i . The currently observed discount curve $t \mapsto DF(0, t)$ can be expressed through implicitly defined continuously compounded annualized rates \bar{r}_j through the following relationship:

$$DF(0, t) = \exp\left(-\int_0^t r(s) ds\right) \quad \text{where } r(s) = \bar{r}_j \quad \text{whenever } s \in (t_{j-1}, t_j]. \tag{44}$$

For year-end dates t_j there holds $t_j - t_{j-1} = 1$, so the expression simplifies to give:

$$DF(0, t_j) = \exp\left(-\sum_{k=1}^j \bar{r}_k\right). \tag{45}$$

Therefore, the rates \bar{r}_j are given by

$$\bar{r}_j = \log \frac{DF(0, t_{j-1})}{DF(0, t_j)}. \tag{46}$$

The risk factor $r_0 \in \mathbb{R}$ is defined to be an additive spread to the current risk-free annualized rates \bar{r}_k so that the discount curve, considered as a function of the risk factor, is given by

$$DF(0, t_j)(\mathbf{r}) = \exp\left(-\sum_{k=1}^j (\bar{r}_k + r_0)\right). \tag{47}$$

Under a risk factor scenario the present values of the cumulated cashflows are assumed to be given by

$$C_{i,j}^{\rightarrow}(\mathbf{r}) = \sum_{k=1}^j C_{i,k} \exp\left(-\sum_{k=1}^j (\bar{r}_k + r_0)\right), \tag{48}$$

$$I_j^{\rightarrow}(\mathbf{r}) = \sum_{k=1}^j I_k \exp\left(-\sum_{k=1}^j (\bar{r}_k + r_0)\right). \tag{49}$$

Finally, we assume the risk factor \mathbf{r} is distributed as a multivariate normal $N(0, \Sigma)$, where the covariance matrix $\Sigma \in \mathbb{R}^{n+1, n+1}$ is given by

$$\Sigma_{i,j} = \begin{cases} \sigma_i^2, & 0 \leq i, j \leq n, i = j, \\ \sigma_0 \sigma_j \rho_a, & i = 0, 1 \leq j \leq n, \\ \sigma_i \sigma_0 \rho_a, & 1 \leq i \leq n, j = 0, \\ \sigma_i \sigma_j \rho_b, & 1 \leq i, j \leq n, i \neq j, \end{cases} \tag{50}$$

with $\rho_a, \rho_b \in [-1, 1]$ specified such that Σ is positive definite. With the stylized example now defined, we next calculate a numerical example demonstrating that data proxies and associated error analysis may have potential application.

5.2 Conclusions from the numerical example

Perhaps surprisingly, the approximate error bound on ordered losses (34) is found to hold with equality for all ordered losses in the numerical example. In the results of the numerical example, depicted in Fig. 2, we observe that:

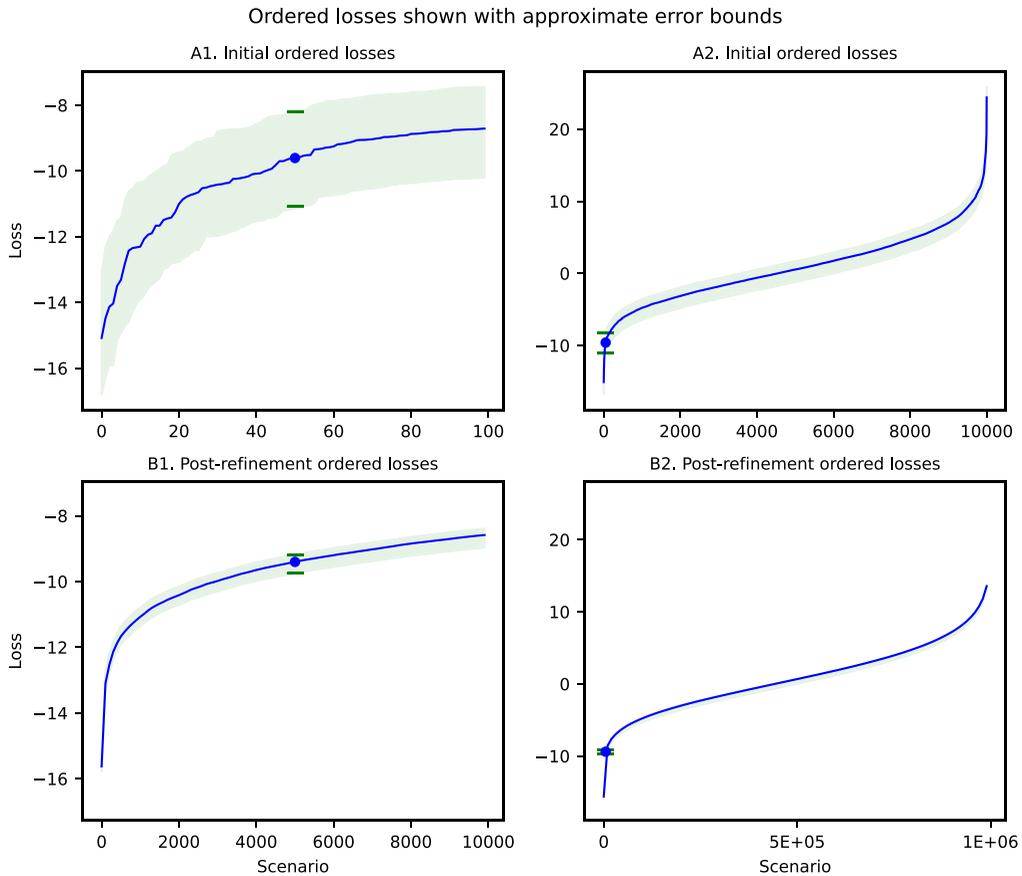


Figure 2. The panels show the result of two Monte Carlo loss simulations of the stylized model (Section 5.1) with example data (Example 5.1). Each row of panels depicts the same data. The right-hand column of panels shows the full ordered loss data with the left-hand column scaled to show detail of the ordered loss data around the 0.5th loss percentile. On the panels, blue lines show the ordered exact loss, and the green shaded region indicates ordered loss values between the approximate lower and upper error bounds. The blue dot is the exact loss at the $0.005 \times N$ ordered loss, representing the 0.5th loss percentile. The green horizontal lines below and above the dot show the associated approximate lower and upper bounds. That the blue line always lies within the shaded region shows that, in this example, the approximated bounds hold mathematically. Panels A1 and A2 show a Monte Carlo simulation with $N = 10,000$ with data proxies chosen to be Chebyshev interpolation with 4 points. The error attribution process is applied (Section 4.5) and all proxies attributed with more than 1% of the upper or lower approximate error bounds are refined – they are rebuilt with 7 Chebyshev points. Panels B1 and B2 show the Monte Carlo loss simulation with the refined proxies and $N = 1,000,000$. The reduced approximate error bounds show that error measurement and attribution can be used as a mechanism for error reduction in percentile estimates.

- Approximate error bounds can be calculated at percentiles relevant for use cases of the proxy model. Using error attribution, practitioners may target improvements to individual proxies to achieve acceptable levels of approximation error across use cases.
- The approximate bound for unordered losses (32a), denoted \lesssim , may be observed to hold with mathematical inequality in practical settings. Sensitivity analysis shown in Table 1 indicates the importance of choosing data proxies prudently.
- The approximate ordered bounds for ordered losses (33) are wider than the unordered bounds, therefore, these may also hold with mathematical inequality. While sensitivity analysis indicates these bounds are less sensitive to prudence, it remains important that error bounds for data proxies are chosen carefully.

- The approach to attribution identified in Section 4.5 and (35) results in the identification of proxies whose associated error contributes significantly to the error bounds. When these proxies are improved, a reduction in the error bound is observed. This can be seen as a systematic mechanism for proxy design improvements and for the reduction in approximation errors propagating into percentile estimators and capital estimates.
- When designing data proxies, practitioners need to consider the computational cost of calculating interpolation and validation points. The error attributed to an individual proxy may be used to decide whether further computational resources should be applied to improvements in the proxy.
- It may be feasible to improve proxy error bounds through making a different choice of approximation or interpolation method, without further increasing the computational cost of the proxy's construction or validation.

In the numerical example, two distinct Monte Carlo loss simulations are performed. The loss calculations are based on the stylized model (Section 5.1) with example data (Example 5.1). In the initial simulation, data proxies are constructed with four interpolation points. The choice to use so few interpolation points is to create proxy errors that will propagate into percentile and capital estimates, and reflects that in practical settings, computational considerations may impose a choice of which proxies most interpolation points are assigned. In the first simulation, only 10,000 scenarios are used, representing a setting where a practitioner may be interested in testing and refining the initial design of data proxies without the use of large Monte Carlo runs. The initial simulation is depicted graphically in Panels A1 and A2 of Fig. 2. Panel A2 shows the exact loss calculation lies within the approximate proxy error bounds at all ordered loss scenarios. The specific 0.5th percentile loss is highlighted with the approximate bounds. Panel A1 depicts a subset of the same scenario data around this 0.5th percentile loss to aid inspection of the approximate error bounds around this percentile.

The approach to partial error attribution to individual proxies, as outlined in Section 4.5, is applied with the basic (and somewhat arbitrary) choice to use seven interpolation points whenever the attribution of the lower or upper approximate error bounds to an individual proxy exceeds 1%. The exact choice of improvement to the proxy is arbitrary for our example, except to illustrate how improvements to individual proxies, and their associated error bounds, can be shown to propagate into improved percentile error bounds.

Panels B1 and B2 of Fig. 2 show a loss simulation with $N = 1,000,000$ and confirm improvements to the approximate error bounds of ordered losses arising from the refinements to certain proxies as an outcome of the error attribution above. Both panels confirm a notable improvement to the approximate error bounds on ordered losses, and further, show that the approximate ordered error bounds still hold mathematically when compared against the exact loss calculations. A sensitivity analysis of the mathematical success of approximate ordered and unordered bounds with respect to scaling of the initial data proxy error bounds is shown in Table 1. It demonstrates increased success as estimates of the data proxy error bounds are made more prudent with the approximate ordered bounds holding in all cases without the need in this example for extra prudence.

As a computational assumption, we assume that full Monte Carlo with exact loss calculations are not feasible. Therefore, practitioners may choose to validate the performance of approximate error bounds on computationally feasible subset of scenarios representative of the loss distribution. In practical settings, the attribution of approximate error bounds to proxies may be performed iteratively to achieve the desired computational performance and error bounds across percentiles suitable for the various use cases at hand. In the numerical example, no extrapolation points are encountered due to the significant interpolation domain chosen.

Example 5.1. (Market data) Market and liability data is given as follows. Table 2 gives example data, with $n = 5$ and $m = 5$, for: asset cashflows $C_{i,j}$ of (37), liability cashflows l_j of (37), discount

Table 2. Data for the stylized example of the analysis of error propagation and attribution in simulation-based capital models.*

Index	j	1	2	3	4	5
Year	t_j	1	2	3	4	5
Discount factor	$DF(0, t_j)$	0.99	0.978	0.962	0.942	0.91
Annualized interest rate (%)	\bar{r}_j	1.0050	1.2195	1.6495	2.1009	3.4561
Liability cashflows	l_j	67.2	57.6	48	32	9.6
Asset 1 cashflows	C_{1j}	8	8	8	32	20
Asset 2 cashflows	C_{2j}	2	2	2	24	-
Asset 3 cashflows	C_{3j}	6	60	-	-	-
Asset 4 cashflows	C_{4j}	71	-	-	-	-
Asset 5 cashflows	C_{5j}	3	3	30	-	-

*Data for the stylized example are from Berry & Sharpe (2021) except for the annualized rates \bar{r}_j , shown rounded to 4 decimal places, that are derived from the discount factors $DF(0, t_j)$ through the expression (46). Positive liability cashflow values represent (expected) monies to be paid, and positive asset cashflows represent (risk adjusted) monies to be received.

rates $DF(0, t_j)$, and annualized rates \bar{r}_j of (46). Table 2 gives example data for asset values v_i and spreads \bar{s}_i of (43). The covariance matrix $\Sigma \in \mathbb{R}^{6,6}$ is defined by (50) with $\rho_a = 0.9$, $\rho_b = 0.8$ and $\sigma_i = 0.05$ for $i = 0, 1, \dots, 6$.

(Computational assumption) The number of Monte Carlo simulations representing the size of calculations that can be performed with light models is assumed to be 1,000,000, and with the heavy model 100. That is $N = 1,000,000$ and $n = 100$ in Assumption 2.

(Heavy model) Scenario data for the heavy model is given by $\pi(\mathbf{r}) = [\mathbf{A}(\mathbf{r}), \mathbf{b}(\mathbf{r}), \mathbf{c}(\mathbf{r})]$, where $\mathbf{A}(\mathbf{r})$, $\mathbf{b}(\mathbf{r})$ and $\mathbf{c}(\mathbf{r})$ are defined in (41c).

(Light model) Observe that each nonconstant component of $\pi(\mathbf{r})$ is only dependent on one of the components of the risk factor vector \mathbf{r} . We consider the data proxies as one-dimensional functions of their respective risk factor component. Data proxies were formed through basic Chebyshev interpolation with the number of Chebyshev points as an input variable.¹³ The interpolation bounds were chosen to be ± 0.2806 across all risk factors representing domain where extrapolation is not expected to occur under Monte Carlo simulation with $N = 1,000,000$. Initial (approximate) error bounds for the data proxies were formed through measuring the maximum observed error at 100 points across the domain. The choice of 100 points is a somewhat arbitrary starting point for a further sensitivity analysis of the estimated data proxy error bounds within the error analysis shown in Table 1. In practical settings, the number of validation points may be limited by computational considerations. Therefore, an appreciation of the sensitivity of the error analysis to this estimate is important within applications. Uniqueness, expected by inspection, was confirmed in the base scenario through numerical exploration of the feasible set, constrained to the optimal value function.

6. Conclusions and Perspectives

Error analysis removes proxies from being a potential single point of failure of internal models. In the present work, a setting is presented that is characteristic of problems faced by life insurance firms with internal models utilizing proxies. The stylized example indicates that approximate error bounds may perform in practical settings.

¹³See, for example, Press et al. (2007) for an introduction to Chebyshev interpolation and to a selection of other common interpolation techniques. The exact choice of interpolation here is somewhat arbitrary since we are primarily concerned with measuring error propagation.

Table 3. Asset price data for the stylized example of the analysis of error propagation and attribution in simulation-based capital models.*

Index	i	1	2	3	4	5
Market value	v_i	40	18	63	69	29
Asset spread (%)	\bar{s}_i	7.6716	6.0824	1.6518	2.8560	4.0934

*The stylized example market data is from Berry & Sharpe (2021). Asset spreads \bar{s}_i are derived from (43) and are shown rounded to 4 decimal places.

The development of error bounds on loss scenarios, valid over the entire loss distribution, may assist the validation of a firm's proxy model across a wide range of use cases beyond just capital requirement modeling. For example, balance sheet estimation, business forecasting and pricing. Overall, formal error analysis may be used to strengthen existing validation evidence, and may have wider applications to the calculation of sensitivities. In certain circumstances, approximate loss bounds can be created for any loss scenario, implying that this data can be used as part of the design and validation of proxies, directly targeting the whole risk factor space.

The concepts presented, which could be characterized as a variational approach to error calculus, applicable to calculations involving an optimal value function, encounter some limitations. It was assumed that the loss model was differentiable with respect its data within a Monte Carlo setting. Uniqueness of the primal and dual problems was assumed in order to justify a local-Lipschitz property and to assert values for partial derivatives. In practice, a firm's loss model may encounter nonuniqueness and conditions that do not imply differentiability. In such circumstances, the posed error bound is not a unique expression, and therefore its potential use is less well-motivated. It was assumed that when the loss function was differentiable with respect to its data, a high dimensional tangent plane approximates the loss function. This assumption may fail when the underlying function exhibits significant deviations or when the approximation is used far from the tangent point. In practice, the error bounds for data proxies should be chosen prudently since failure can be observed when bounds are chosen sharp. In contrast, the approximate ordered bounds appear prudent, but are also sensitive to the accuracy of the data proxy error bounds.

It is envisioned that formal error analysis could feature more prominently in internal model validations across the life insurance industry. Advances in error analysis offers a route to enhanced confidence in capital estimates.

Further applications of the presented methods, beyond those outlined for error analysis, include general sensitivity analysis and the calculation of risk. For example, the methods enable analytical expressions to be formulated for capital sensitivities arising from changes to market data, such as asset prices and interest rates. Analytical sensitivity expressions utilizing primal and dual values may also have use in enhancing the performance of Automatic Differentiation frameworks, such as Paszke *et al.* (2017), when applied to life insurance settings whenever optimization is integral to balance sheet and capital calculations. Overall, beyond error analysis, the presented methods may find a wide range of applications where sensitivity analysis and the calculation of risk are required.

Acknowledgments. This work was funded by Rothesay Life Plc where DJC is a full-time employee. Opinions expressed here are those of the author and do not necessarily represent those of Rothesay. I would like to thank Shayanthan Pathmanathan for introducing me to this topic. I would also like to thank Oliver Dixon and Lucinda Parlett for providing thoughtful comments on an early draft. Sam Kinsley and Tasos Stylianou provided helpful comments and assisted the research for mathematical references. I am also grateful to Professor Stephan Dempe (Technische Universität Bergakademie Freiberg) for his correspondence and for kindly highlighting the work of Klatte & Kummer (1985). Finally, I would like to thank Simon Johnson, Baptiste Grassion, and Yann Samuelides for supporting the public dissemination of this work. Data supporting this study are included in Tables 2 and 3.

References

- Androschuck, T., Gibbs, S., Katrakis, N., Lau, J., Oram, S., Raddall, P., Semchyshyn, L., Stevenson, D. & Waters, J. (2017). Simulation-based capital models: testing, justifying and communicating choices. a report from the life aggregation and simulation techniques working party. *British Actuarial Journal*, **22**(2), 257–335.
- Apostol, T. (1957). *Mathematical Analysis*. Reading, USA: Addison-Wesley.
- Bellman, R. (1961). *Adaptive Control Processes: A Guided Tour*. Princeton University Press.
- Berry, T. & Sharpe, J. (2021). Asset–liability modelling in the quantum era. *British Actuarial Journal*, **26**, E7.
- Blanning, R.W. (1975). The construction and implementation of metamodels. *Simulation*, **24**(6), 177–184.
- Borwein, J. & Lewis, A. (2000) *Convex Analysis and Nonlinear Optimization*, CMS Books in Mathematics, vol. 3. New York: Springer.
- Boyd, S. & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.
- Christiansen, M.C. (2008). A sensitivity analysis concept for life insurance with respect to a valuation basis of infinite dimension. *Insurance: Mathematics and Economics*, **42**(2), 680–690.
- Conwill, M.F. (1991). A linear program approach to maximizing policy holder value. *Actuarial Research Clearing House (Arch)*, **3**.
- Crispin, D.J. & Kinsley, S.M. (2022). Eliminating proxy errors from capital estimates by targeted exact computation. *Annals of Actuarial Science*, **17**(2) 219–242.
- Daykin, C.D. & Hey, G.B. (1990). Managing uncertainty in a general insurance company. *Journal of the Institute of Actuaries*, **117**(2), 173–277.
- De Wolf, D. & Smeers, Y. (2021). Generalized derivatives of the optimal value of a linear program with respect to matrix coefficients. *European Journal of Operational Research*, **291**(2), 491–496.
- Debie, E. & Shafi, K. (2019). Implications of the curse of dimensionality for supervised learning classifier systems: theoretical and empirical analyses. *Pattern Analysis and Applications*, **22**(2), 519–536.
- Dempe, S. & Mehrlitz, P. (2015). Lipschitz continuity of the optimal value function in parametric optimization. *Journal of Global Optimization*, **61**(2), 363–377.
- Dinkelbach, W. (1969). *Sensitivitätsanalysen und Parametrische Programmierung*. Berlin: Springer-Verlag.
- Evans, L.C. & Garzepy, R.F. (2018). *Measure Theory and Fine Properties of Functions*. Boca Raton: Routledge.
- Freund, R.M. (1985) Postoptimal analysis of a linear program under simultaneous changes in matrix coefficients. In *Mathematical Programming Essays in Honor of George B. Dantzig Part I*, pp. 1–13. Springer.
- Gal, T. (1979). *Postoptimal Analyses, Parametric Programming, and Related Topics*. McGraw-Hill.
- Hejazi, S.A. & Jackson, K.R. (2017). Efficient valuation of SCR via a neural network approach. *Journal of Computational and Applied Mathematics*, **313**, 427–439.
- Hughes, I. & Hase, T. (2010). *Measurements and their Uncertainties: A Practical Guide to Modern Error Analysis*. Oxford: Oxford University Press.
- Hursey, C., Cocke, M., Hannibal, C., Jakhria, P., MacIntyre, I. & Modisett, M. (2014). Heavy models, light models and proxy models, a working paper. The Proxy Model Working Party, Institute and Faculty of Actuaries.
- International Organization for Standardization (ISO) (1993). Guide to the Expression of Uncertainty in Measurement. ISO (Geneva).
- Kirkup, L. & Frenkel, R.B. (2006). *An Introduction to Uncertainty in Measurement: Using the GUM (Guide to the Expression of Uncertainty in Measurement)*. Cambridge: Cambridge University Press.
- Klatte, D. & Kummer, B. (1985). Stability properties of infima and optimal solutions of parametric optimization problems. In *Nondifferentiable Optimization: Motivations and Applications*, pp. 215–229. Springer.
- Kocherlakota, R., Rosenbloom, E. & Shiu, E.S. (1988). Algorithms for cash-flow matching. *Transactions of the Society of Actuaries*, **40**(1), 477–484.
- Krah, A.-S., Nikolić, Z. & Korn, R. (2018). A least-squares Monte Carlo framework in proxy modeling of life insurance companies. *Risks*, **6**(2), 62.
- Lazzari, S. & Bentley, O. (2017). Perfecting proxy models. *The Actuary*.
- Lebesgue, H. (2003). *Leçons sur l'intégration et la recherche des fonctions primitives*, vol. 267. American Mathematical Society. Unabridged reprint of 2nd edition, 1928 Paris, with minor changes and corrections.
- Lin, X.S. & Yang, S. (2020). Fast and efficient nested simulation for large variable annuity portfolios: a surrogate modeling approach. *Insurance: Mathematics and Economics*, **91**, 85–103.
- Liu, L. & Özsu, M.T. (2009). *Encyclopedia of Database Systems*, vol. 6. Springer.
- Maly, J. (1999). A simple proof of the Stepanov theorem on differentiability almost everywhere. *Expositiones Mathematicae*, **17**, 059–062.
- Murphy, P. & Radun, M. (2021). Proxy models: uncertain terms. *The Actuary*.
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L. & Lerer, A. (2017). Automatic differentiation in pytorch. In *NIPS 2017 Autodiff Workshop: The Future of Gradient-based Machine Learning Software and Techniques*.

- Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P.** (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press.
- Prudential Regulation Authority** (2018a). Solvency II: internal models – modelling of the matching adjustment. Supervisory Statement SS8/18.
- Prudential Regulation Authority** (2018b). Solvency II: matching adjustment. Supervisory Statement SS7/18.
- Rademacher, H.** (1919). Über partielle und totale differenzierbarkeit von funktionen mehrerer variablen und über die transformation der doppelintegrale. *Mathematische Annalen*, **79**(4), 340–359.
- Riesz, F. & Nagy, B.S.** (1990). *Functional Analysis*. New York: Dover Publications.
- Robinson, B. & Elliott, M.** (2014). Proxy models: the way of the future? *The Actuary*.
- Rockafellar, R.** (1984) Directional differentiability of the optimal value function in a nonlinear programming problem. In *Sensitivity, Stability and Parametric Analysis*, pp. 213–226. Springer.
- Tilley, J.A.** (1980). The matching of assets and liabilities. *Transactions of the Society of Actuaries*, **32**, 263–300.
- Wise, A.** (1984a). The matching of assets to liabilities. *Journal of the Institute of Actuaries*, **111**(3), 445–501.
- Wise, A.** (1984b). A theoretical analysis of the matching of assets to liabilities. *Journal of the Institute of Actuaries*, **111**(2), 375–402.