DFit: Cloud-Based Service and Business Case

A Technical Report submitted to the Department of Systems and Information Engineering

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

John Blair Mitchell III

Spring, 2024

Technical Project Team Members

Franklin Glance

Esther Yi

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Roman Krzysztofowicz, Department of Systems and Information Engineering

DFit: Cloud-Based Service and Business Case

Franklin Glance*, Mitch Mitchell*, and Esther Yi*
* Department of Systems and Information Engineering
School of Engineering and Applied Science
University of Virginia, Charlottesville, Virginia 22904
Email: fpg2kv@virginia.edu, jbm8efn@virginia.edu, esy4nq@virginia.edu

***Abstract*—Distribution fitting is a major part of any probabilistic modeling or forecasting problem. Such problems are becoming more relevant as companies and researchers begin opting for robust stochastic models over simpler deterministic models to better interpret and use data by quantifying uncertainties. However, common probability distributions rarely fit neatly to real-world data. This motivates Distribution Fitter (DFit), a software that fits a variety of continuous parametric distributions to random samples, censored samples, or expert-assessed quantiles, and then ranks them by goodness-of-fit to find the best model.**

DFit will be provided to students and professionals as a web-based application. It is a joint venture between three parties: CapTech, a consultancy in Richmond, Virginia, represented by Wray Mills, whose team hosts, operates, and services DFit; Roman Krzysztofowicz, who authored the theory behind DFit and the book *Probabilistic Forecasts and Optimal Decisions* [1]; and Wiley & Sons, Ltd., who is to publish the book in 2024. This provides a business case for CapTech, who needs to decide whether DFit is worth supporting. Three goals arise from this business case: (i) to improve the scalability and performance of DFit, (ii) to evaluate its performance relative to competing software in the market, and (iii) to forecast the demand for DFit and the book.

First, DFit was moved to a cloud-based infrastructure to prepare for the uncertain amount of traffic it will receive once the book is published. Next, simulations and experiments were conducted to see how DFit performs on statistical and usability criteria in comparison to alternative distribution-fitting software. Finally, probabilistic forecasts of the three-year demand for the book and DFit were prepared based on the experimental results and market analyses.

The improvements to the DFit software partially dictated how it performed in experiments, which, in turn, affected the forecasts. The forecasts then serve as the primary deliverable for CapTech's business case.

***Index Terms*—distribution-fitting, web-based application, experimental design, probabilistic forecasting**

I. Introduction

When constructing any probabilistic model from data, the task of fitting parametric distributions is of utmost importance as it provides well-defined quantification of uncertainty. Unfortunately, with real-world limitations such as small sample sizes or samples drawn from very complex processes, this task is difficult, and identifying the "true" underlying distribution is unlikely, if not impossible. Instead, one might seek to fit distributions from many different families, choosing the one that best represents the sample. Chapter 3 of the book *Probabilistic Forecasts and Optimal Decisions* by Professor Krzysztofowicz [1] details a methodology for performing this task. First, an empirical distribution function is constructed from the sample using meta-Gaussian plotting positions, which are especially useful due to being well calibrated against the standard normal distribution function up to its second moment. Assessed quantiles can be used in place of the empirical distribution function. Then, the parameters for continuous distributions from a variety of families are estimated such that they minimize the maximum absolute difference (MAD) between the hypothesized distribution function and the empirical distribution function. The estimated distributions can then be compared based on MAD and visual analysis to determine which appears to best fit the sample. The Distribution Fitter (DFit) software is provided with the book and is a one-stop solution for fitting distributions based on this methodology.

The connection between the book and DFit underlies a joint venture between Professor Krzysztofowicz, Wiley & Sons, Ltd., and CapTech Consulting (Fig. 1). A student interested in probabilistic modeling may purchase the book from the publisher, Wiley. This will provide them with the theory behind the methodology, authored by Professor Krzysztofowicz. It will also give them the ability to apply the theory with DFit, which is hosted, operated, and serviced by CapTech.

At this point in time, Professor Krzysztofowicz has already authored the book and Wiley has begun publishing it. So, CapTech is the only party whose involvement is subject to decision. This creates a business case for them, as they must decide whether the DFit software is worth maintaining. To support this

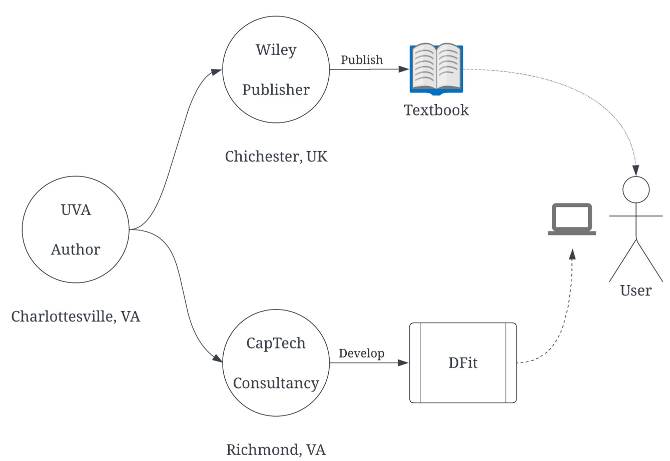

Fig. 1: Visual representation of the joint venture.

business case, CapTech requested: (i) a modernization of DFit to allow for more scalability and future improvements, (ii) an evaluation of its performance relative to competing distribution-fitting software, and (iii) a forecast of the demand for the book and DFit over the next three years.

II. DFit Cloud Migration

In the first portion of this capstone project, DFit was successfully migrated from an outdated on-premises hosting environment to a modern cloud-based infrastructure on Amazon Web Services (AWS). This transition was not simply a change of hosting. It involved a comprehensive overhaul of the software's architecture to address anticipated growth in its use and limitations in the original tech stack.

The upcoming release of Professor Krzysztofowicz's book is anticipated to result in an uptick in demand for DFit since the book references the fitting software in exercises and examples. This necessitates a scalable and efficient web hosting solution, which was not possible with the previous on-premises hosting.

Migrating DFit from on-premises hosting to the Amazon Web Services cloud required a comprehensive redesign of the existing codebase. It was not possible to upload the existing codebase to the cloud due to the outdated nature of the original tech stack. The migration process involved two main parts: the containerization and migration of the existing "fit engine" and the redevelopment of the frontend website using a modern web framework. Each part required careful planning of the new system architecture.

A. Fit Engine Migration

The fit engine is the core component of DFit. It takes in a dataset and fitting specifications and outputs the optimal parameter values for various probability distributions using a proprietary algorithm developed by Professor Krzysztofowicz. The distribution fitting algorithm is implemented in Fortran, with a Python script acting as an adapter. Despite the outdated nature of the Fortran fitting engine, the decision was made to keep the existing codebase.

Rewriting the fit engine in a more modern language was considered but ultimately decided against due to the time constraints of the project and the fact that the existing Fortran engine had been thoroughly tested over the years. Additionally, it would have required significant time and resources, which could have potentially delayed the migration process without producing noticeably better results for the end user.

To migrate the fit engine to the cloud, the Python/Fortran script was containerized using Docker. Containerization allowed the fit engine to be deployed as an AWS Lambda function. Lambda is a serverless compute platform that runs code in response to events, such as an Application Programming Interface (API) call. This results in dynamically scalable compute, which will be necessary to handle the anticipated surge in website usage. The fit engine Lambda function exposes an API that the website frontend can use to run distribution fits. Additionally, when considering future business cases, the same API can be offered as a standalone service to commercial clients looking for a high-volume distribution fitting solution.

The separation of the fit engine from the frontend also provided additional benefits in terms of the development process. For example, it allowed for simultaneous development of the fit engine alongside the frontend website. Additionally, testing could be done on each component separately. This separation significantly reduced the time spent debugging and testing, while making it easy to update and maintain the fit engine without affecting the frontend development process.

B. Frontend Migration

The original DFit frontend was created around the same time as the fit engine using Rails 3, an outdated version of the Ruby on Rails web framework. The decision to rewrite the frontend rather than update the existing codebase was motivated by several factors.

Initially, moving the existing codebase to AWS was considered. This required containerizing the repository using Docker, which was unsuccessful due to compatibility issues resulting from outdated dependencies used in the development of the original site. An additional consideration was made regarding updating the frontend to Rails 7, which would resolve the issues with Docker. However, this presented further problems due to the incompatibility of the active scaffold library with newer versions of Rails. Typically, such compatibility issues could be handled by diligent debugging, but active scaffold severely hindered this. Active scaffold was deeply ingrained in the codebase, so its incompatiblility with the more modern version of Rails made updating the codebase nearly impossible.

After considering different options, the decision was made to redevelop the frontend using the Django web framework. Django was chosen because it is widely regarded as a robust and scalable web framework. Redevelopment allowed for a clean slate coding approach for the frontend. This required strict adherence to coding best practice and architecture considerations for further development and improvements to the website.

The new frontend was designed to interact with the fit engine Lambda API to fit distributions in a similar manner to the old website. It allows for users to upload sample data and request distribution fits for the same distributions available previously. Amazon S3 was used to store uploaded data files because it provides secure and scalable data storage. A MySQL database was used to store the other relevant information, such as user data and parameter estimates. These databases connected securely to the AWS EC2 compute service which was used to host the website. Keeping the entire website hosted on a single platform like AWS made it easy to ensure that the site is secure and efficient.

C. Results

After several months of development, DFit was successfully migrated to AWS. The new frontend successfully interfaced with the fit engine Lambda API to perform distribution fitting on uploaded random samples and assessed quantiles. The cloud migration provided significant improvements in scalability, security, and performance compared to the old on-premises version. However, after discussing the website's look and feel with Professor Krzysztofowicz, it was learned that he preferred the user interface of the old DFit version. That version had been designed based on principles of cognitive psychology for human-computer interaction and for the visual display of quantitative information, which remain valid. Now, the focus has shifted towards developing another frontend version that maintains the appearance of the old version while preserving the performance and scalability improvements associated with the cloud-based model. Pivoting development direction based on feedback such as this is part of the reason why using a two-pronged approach is so helpful. Rather than having to restart from scratch, another site can be built to interface with the existing Lambda API.

Overall, the migration of DFit to the cloud has been a successful endeavor. The next phase of the project will involve refining the user interface to meet Professor Krzysztofowicz's preferences.

III. DFit Performance

Since DFit is ultimately a commercial software being maintained by CapTech, our technical venture party, it is integral to build a strong business case advocating for their continued support of it. This is especially so with its expected increase in traffic due to the publication and advertising of Professor Krzysztofowicz's book, which leads to the possibility of more monetary funds being necessary to support DFit. Consequently, a series of experiments were conducted to better understand and capitalize on the strengths of DFit. By having a firmer grasp of the ways that DFit outperforms competing software, a more effective implementation plan can be enacted to increase its market value for CapTech.

A. Methodology

First, the competitors of DFit were identified. There are numerous distribution fitting software; however, not all are necessarily competing in the same market space as DFit. The criteria for determining its competitors were: (i) the probability distributions they support, (ii) the completeness of their documentation, and (iii) the estimation methods they use. Several of the same continuous distributions must be supported by DFit and other distribution fitting software to allow for direct comparison of the performance of their estimation methods. Documentation is also a necessary component so that their estimation methods and measures of goodness-of-fit are known. DFit uses a proprietary estimation method designed by Professor Krzysztofowicz, called the Uniform Distance method, so testing it against the implementation of other estimation methods to determine the advantage of DFit is requisite. The conclusion of this analysis gave a list of three other software that fit our criteria for comparison: Statistics and Machine Learning Toolbox, stat::fit, and XLSTAT. The Statistics and Machine Learning Toolbox is a MATLAB software package; stat::fit is a stand-alone software by Geer Mountain Software that is incorporated by many simulation products; and XLSTAT is a Microsoft Excel extension. All three use the method of moments and the method of maximum likelihood, which are standard distribution fitting procedures. However, each program implements them differently, giving different results that are, in turn, also different than the results from DFit's Uniform Distance method.

The goodness-of-fit measure used for comparing software performance was MAD. The MAD ($0 \leq \text{MAD} < 1$) is the maximum absolute difference between the empirical distribution function, constructed with meta-Gaussian plotting positions, and the parametric distribution function. As such, the MAD tells how well a particular distribution represents the data.

To actually test the effectiveness of DFit, two types of experiments were conducted. The first experiment used simulated samples from known distributions supported by the four software. These were determined by reading the documentation of the competitors, and were normal, logistic, exponential, and Weibull. Samples with 5, 25, 100, 200, 500, 1000, and 20000 realizations were generated; sizes were determined using previous work by Herr and Krzysztofowicz [2]. The samples were randomly generated in MATLAB from the four distribution types, and the same samples were used by all four software. For each distribution type, the parameter values estimated by the software were compared to the actual parameter values.

The second experiment was ran using two real-world datasets provided by CapTech with sizes 194 and 514. Both were split into sub-samples of sizes: 194, 100, 50, 25, 5; and 514, 250, 100, 50, 25, 5. The distribution that had the smallest MAD using DFit was the Weibull distribution. So, the Weibull distribution was then fit to all of the sub-samples by each software and reviewed to determine which resulted in the smallest MAD. The parameter estimates and MADs were then recorded and compared against those from DFit.

Finally, the user experience associated with each software was evaluated on general ease of use. This was observed qualitatively by judging how easily the software can be implemented, what kind of documentation exists, and how they display the estimates of parameters and the goodness-of-fit values.

B. Results

The results of the experiments can be summarized as follows:

1) Although MATLAB's Statistics and Machine Learning Toolbox provide the parameter values closest to the true values, their parametric distribution functions do not fit the empirical distribution functions as well as DFit.
2) DFit offers a better guided experience for users, with easily accessible documentation and parameter estimates that provide the smallest MAD.

DFit automatically supplies the user with the MAD and uses it to rank the distribution types if multiple fits are done on the same data. However, the Statistics and Machine Learning Toolbox, XLSTAT, and stat::fit, do not supply MAD. The MAD is an important and preferred performance metric over the error of parameter estimates because true parameter values are not known in real-world applications. Since DFit is the only software that provided the MAD to the user, it was calculated manually for the other three software.

In the first experiment, the four distributions—normal, logistic, exponential, and Weibull—were created with known parameter values. The fits from the different software were evaluated on their absolute difference from the true parameters as well as MAD. Overall, the Statistics and Machine Learning Toolbox performed the best in terms of recovering the true parameter values. However, when comparing the performance in terms of MAD, the Statistics and Machine learning Toolbox and XLSTAT performed the worst, having the largest MAD values across the board.

In the second experiment, DFit provided the smallest MAD in all cases except for the sub-sample of size 5 from the sample of size 194. Parameter estimates could not be compared because true parameter values were not known. In this set of experiments, the Statistics and Machine Learning Toolbox and stat::fit performed the worst, with XLSTAT performing in the middle. Also, when comparing the average MAD across the different fits, DFit's were the smallest (TABLE I).

In both experiments, with the simulated data and with the real-world data, overall, DFit performed the best in having the smallest MAD values. There were only three instances (the logistic distribution estimated from samples of sizes 200 and 1000 and from a sub-sample of size 5) where DFit did not perform better than the other software.

Despite all four software providing functionality for estimating the parameters of parametric distributions, their user experience is different. The benefit of DFit is that it guides the user through the distribution modeling methodology from the book, where multiple distribution types are fit and ranked based on MAD, with graphs available for visual analysis of goodness-of-fit. To determine the types of distributions to fit, the user can easily select all that are on the appropriate sample space. The Statistics and Machine Learning Toolbox requires the user to specify different distribution types and evaluate their goodness-of-fit on their own. With many types of distributions available, determining which are appropriate is not straightforward. Also, since it is a MATLAB package, it requires programming experience to use. XLSTAT and stat::fit offer more in guiding the user and have interfaces that don't require programming experience, but they still fall behind DFit in usability. Stat::fit can fit and rank the goodness-of-fit of multiple types of distributions, but the metric they use for these rankings is unclear. XLSTAT suffers from the same problems as the Statistics and Machine Learning Toolbox, with the exception of not needing programming experience. Formal documentation can be found online for all of the software, but DFit is the only one that is released alongside a thoroughly researched book containing the methodology behind it. Overall, for the task of distribution modeling, DFit's user experience is superior to that of the three competitors.

TABLE I: Average MAD across software.

Software	Average MAD	
	Six sample sizes from 519 to 5	Five sample sizes from 194 to 5
DFit	0.127	0.096
Statistics and Machine Learning Toolbox	0.188	0.167
Stat::fit	0.203	0.180
XLSTAT	0.153	0.142

IV. Demand Forecasts

The last task was preparing forecasts of the demand for Professor Krzysztofowicz's book and for the number of users of DFit over the next three years. As both of these random variables are unique and not represented by past data, as well as the fact that data on book sales and software usage are generally kept private, they are forecasted judgmentally. Ascertaining the knowledge necessary to make these judgments was done through market analyses, which culminated in assessed quantiles for each of the two random variables. Parametric distributions were then fit to these quantiles using DFit, quantifying the uncertainty in the demand for the book and DFit.

A. Methodology

The book presents the fundamentals of probabilistic forecasting and optimal decision making under uncertainty, as well as supporting material on mathematical modeling, probabilistic reasoning, statistical estimation, judgmental assessment, rational decision making, and numerical calculations. As such, it is relevant to upper-level undergraduate and first-year graduate courses from a wide variety of science and engineering disciplines, including but not limited to: Systems and Industrial Engineering, Operations Research, Mathematics, Statistics, Economics, and Management Science. Its adoption by courses from these disciplines provides the first primary source of demand, which is the students taking those courses. The second primary source of demand is libraries that contain this type of engineering and management science content. The third primary source of demand is researchers of forecasting and decision problems and consultants in the areas of predictive and decision analytics.

To analyze the demand coming from students, upper-level undergraduate and first-year graduate courses from engineering and business schools were researched to identify the typical class size and common required books. Highly ranked programs in the aforementioned disciplines were used for this research, which were identified using U.S. News & World Report rankings. Within the United States, the Best Engineering Schools and Best Business Schools rankings were used. As the book was written in English, programs outside of the United States were limited to those offered in English, for which the Best Global Universities in the United Kingdom, Canada, and Australia rankings were used. Each program was analyzed for upper-level undergraduate and first-year graduate courses in probability and decision theory, where the goal was to gauge how frequently these courses are offered, how many students enroll in them, and how successful the books they require tend to be. The result of this research informs judgment towards the number of students that may contribute to the demand for the book.

To analyze the demand coming from libraries, an estimate was assessed of the number of libraries that contain English books related to probability and decision theory. These are mostly limited to those at engineering and business schools, as they are the ones most likely to own a copy of a book with this type of content. This informs judgment towards the number of libraries who may contribute to the demand for the book.

To analyze the demand coming from researchers and consultants, the value of an author's name-recognition in probability and decision theory books was researched. This is because the number of researchers and consultants whose work is related to these problems is too difficult to estimate. Authors were identified by searching for popular books, which are those that are well-known or that arose in the analysis of student demand, and the value of their name-recognition was gauged by comparing the number of citations they have received on papers to the number of citations they received on their book. This then informs judgment towards the number of consultants and researchers who may contribute to the demand for the book.

These analyses were then used to assess the demand for the book in terms of five p-probability quantiles, for $p = 0.01, 0.25, 0.5, 0.75, 0.99$. These assessed quantiles were then used to fit a parametric distribution function with DFit. This distribution function constitutes a probabilistic forecast of the demand for the book over the next three years.

The number of users of DFit is closely tied to the demand for the book, as it is referenced throughout and is supposed to be used alongside the book. However, it is not necessarily restricted to purchasers of the book. So, in forecasting the number of users of DFit, the potential number of users not coming from the book was also of interest. This was analyzed using the Google Trends of related terms to see how many people are making searches that suggest they are interested in distribution-fitting software. Then, another set of five p-probability quantiles was assessed for the number of users of DFit. As before, these were used to fit a parametric distribution function with DFit, which serves as the probabilistic forecast of the number of DFit users over the next three years.

B. Results

The quantiles for both the book and DFit were assessed judgmentally based on the market analyses of the primary sources of demand. For students, it was found that most Universities with engineering programs, both inside and outside of the U.S., were likely to have several courses at the upper-level undergraduate and first-year graduate level that are related to quantitative decision making and probabilistic modeling. These most often were within Management Science, Industrial and Systems Engineering, and Operations Research departments, which tended to be interdisciplinary and accessible to students from a variety of fields. Class sizes were generally not greater than 30, with the few exceptions coming from the most popular programs. The required books were infrequently listed.

For libraries, it was found that there are likely one to three thousand that may purchase copies of books relating to probabilistic forecasting and quantitative decision making. In optimistic scenarios, it would be expected that a large fraction of these would purchase a copy of the book, while pessimistic scenarios would have a small fraction.

For researchers and consultants, it was found that an author's name-recognition, measured through their number of citations on papers, was very relevant to the number of citations they receive on books. Authors of popular probability and decision theory books tended to have hundreds, if not thousands, of citations on their most popular papers. As Professor Krzysztofowicz has also received several hundred citations across his top papers, it would be expected that his name-recognition is significant enough to help drive demand from researchers and consultants.

For organic traffic to DFit, it was found that terms relating specifically to the act of fitting probability distributions were often searched with similar frequency to broader terms such as "probability" or "statistics". It is important to note, though, that non-exact terms such as "data fitting" were much more popular than more specific terms like "probability distribution fitting". So, there is a significant number of search engine users that are interested in fitting distributions to data, but they may not know exactly what they are looking for. If DFit is well-optimized for these types of searches, then organic traffic could potentially serve as a significant source of users for DFit.

The assessed quantiles of the 3-year demand for the book are:

1) The lowest quantile ($p = 0.01$) is assessed at 100 copies, which reflects a *highly* pessimistic scenario where the book fails to gain significant traction beyond a very niche audience. This could occur if the book is not adopted by any courses outside of the University of Virginia and only a few researchers and consultants find it relevant.
2) The first quartile ($p = 0.25$) is assessed at 1000 copies, which represents a conservative estimate. It accounts for some course adoptions and interest from researchers and libraries, but not widespread recognition or usage.
3) The median ($p = 0.5$) is assessed at 3000 copies, which indicates a moderate success level, where the book is adopted by several courses across the targeted disciplines and countries, garners interest from a reasonable number of libraries, and is recognized by a fair share of researchers and consultants.
4) The third quartile ($p = 0.75$) is assessed at 8000 copies, which suggests a more optimistic scenario where the book is widely adopted in courses, has significant library presence, and is well-regarded among researchers and consultants.
5) The highest quantile ($p = 0.99$) is assessed at 15000 copies, which represents an *highly* optimistic scenario where the book becomes a leading reference in the field, is adopted by a large number of courses globally, and achieves substantial sales to researchers and consultants.

The assessed quantiles for DFit are closely related to the assessed quantiles for the book, as it is expected that effectively everyone who purchases the book will use DFit. However, each p-probability quantile for DFit is larger than the corresponding p-probability quantile for the book to account for users coming from word-of-mouth and search engines. As p increases, the quantiles reflect more optimistic scenarios, so the size of the increase from the book's p-probability quantile to DFit's p-probability quantile is expected to increase. Based on this, the assessed quantiles of the 3-year demand for DFit are:

1) ($p = 0.01$) 200 users,
2) ($p = 0.25$) 1500 users,
3) ($p = 0.5$) 5000 users,
4) ($p = 0.75$) 12000 users,
5) ($p = 0.99$) 20000 users.

Using DFit, all available distribution functions on a bounded below interval were fit to the two sets of assessed quantiles with a lower bound $\eta = 0$. The selected distribution functions were chosen based on their MAD and visual goodness-of-fit to the assessed quantiles. The exponential distribution $\mathrm{EX}(\alpha, \eta)$ [1] was selected for both forecasts, with mean $\alpha = 4792.03$ for the book and $\alpha = 7187.80$ for DFit. Fig. 2 and 3 show these distribution functions overlaid on their corresponding assessed quantiles. They serve as probabilistic forecasts of the 3-year demand for the book and DFit.

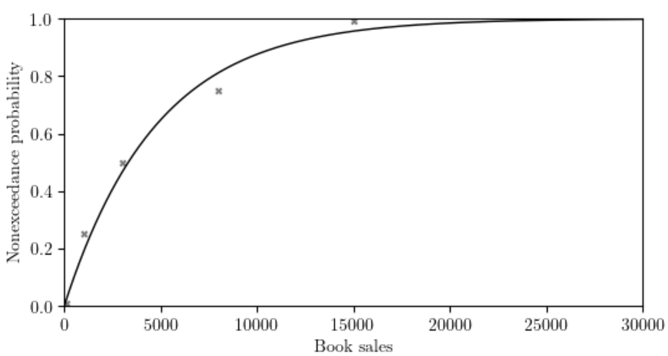

Fig. 2: The book's forecasted demand and assessed quantiles.

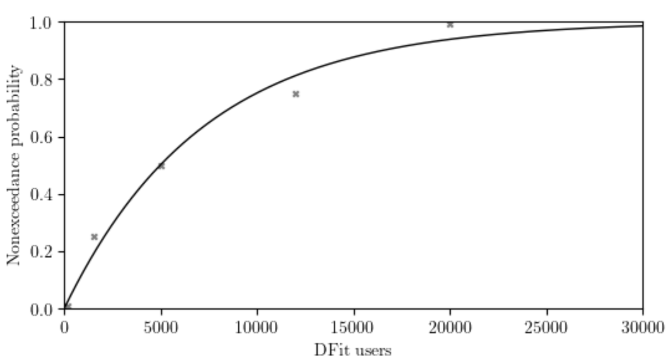

Fig. 3: DFit's forecasted demand and assessed quantiles.

V. Summary and Conclusions

The migration of DFit, the software implementing a distribution modeling methodology, to a cloud-based infrastructure has been a critical step towards enhancing the software's scalability and performance in preparation for the release of the book. This addressed the limitations of the previous tech stack and set a foundation for future improvements and scalability.

Performance evaluation of DFit against competing distribution-fitting software demonstrated its superiority in terms of a goodness-of-fit of the distribution to data. This evaluation highlighted the strengths of DFit's Uniform Distance method, which outperformed others in fitting real-world data. The analysis also identified opportunities to improve DFit's user experience, distinguishing it from competitors through better documentation, user guidance, and the ability to fit many families of distributions at once to find the best model.

Forecasting the 3-year demand for both the book and DFit utilized market analyses and judgmental techniques for assessing quantiles to quantify the uncertainty about the demand. This approach revealed a broad range of potential outcomes, from conservative to highly optimistic scenarios, based on course adoptions, library purchases, researcher and consultant interest, and organic search traffic. These forecasts serve as a key deliverable for CapTech.

In conclusion, this report presents a compelling business case for CapTech through the modernization of DFit, a comprehensive performance evaluation, and judgmental demand forecasts. The joint venture between Professor Krzysztofowicz, Wiley & Sons, Ltd., and CapTech Consulting not only enhances the academic and practical applications of probabilistic forecasting and optimal decision making, but also positions DFit as a unique and strong tool for modeling probability distributions of continuous random variables.

Acknowledgments

We would like to thank CapTech and their representative, Wray Mills, for their sponsoring of this project and continual support.

References

[1] R. Krzysztofowicz, *Probabilistic Forecasts and Optimal Decisions*. Wiley, 2024.

[2] H. D. Herr and R. Krzysztofowicz, "Addendum to Bayesian ensemble forecast of river stages and ensemble size requirements," *Journal of Hydrology, 515, 304-306*, 2014.