SOFIA
UNIVERSITY

ST. KLIMENT
OHRIDSKI
EST. 1888

Faculty of Economics and Business Administration

Mihail Veselinov Yanchev

Summary

Of a Dissertation on the Topic of:

Modeling Economic Uncertainty: Methods, Evaluation, and Applications of Probabilistic Forecasting

In fulfillment of the requirements for obtaining the academic degree
"Doctor of Philosophy" in 3.8 Economics

Academic Advisor: **Assoc. Prof. DSc Anton Gerunov**

Sofia

2023

Table of Contents

Introduction

Uncertainty is a fundamental characteristic of the future and the present. Making optimal decisions related to future outcomes requires on the one hand a definition of possible outcomes, and on the other an attempt to categorize and measure the uncertainty surrounding them. This is true both on an individual level as well as on an organizational level. Many economic and business decisions require both a prediction of future outcomes as well as a measurement of the underlying uncertainty. According to the prominent economist and thinker Frank H. Knight (1921), the notion of risk encompasses all uncertainty that can be measured and in order to separate the two notions, uncertainty should refer only to the immeasurable. This definition has dominated economic thinking for the past century and was instrumental in the development of the risk management domain. Economists, statisticians, and engineers have all considered and contemplated the concept of uncertainty and the ways to measure it and interesting categorizations and dichotomies have emerged.

A somewhat different sociological view on uncertainty is presented by Beck (1992) who argues that modernity has brought new levels of complexity in the global socio-economic systems, which in turn has created new risk factors like financial crises, recessions, and climate change, which compound with natural risks which societies have faced historically. In this sense, economic uncertainty arises from the behavior of industries and governments and the interplay between them. Fiscal and monetary policies, trade and investment flows and international and national political developments are all determining factors. On the macroeconomic level risks are often related to shocks and the accumulation of imbalances. On an enterprise level, the macroeconomic risks are still very relevant, but one could also identify risks related to developments in a specific market or a supply chain, or the organization itself.

A growing body of research has been focused on the measurement and modeling of firm-level and economy-wide risks. The concept of value-at-risk which originates in the financial risk management (Markowitz, 1959) domain has been recently applied to other areas like macroeconomics, resulting in an analytic approach called growth-at-risk which has been used by institutions like the IMF in the task of global risk monitoring (Prasad et al., 2019). Similarly, the current dissertation aims to define a framework and an approach, which allows for both forecasting of future outcomes and the measurement of uncertainty surrounding the forecasts for economic indicators of any type, whether they are on a microeconomic or a macroeconomic level. This framework would be especially useful when the decision-maker or the forecaster faces a realization of a low-probability event or an unexpected shock. In parallel, the study draws inspiration from the rapidly developing field of machine learning and deep learning to propose a novel approach to probabilistic forecasting.

The study depends on empirical results within the realm of economics. The conclusions of the study bear important implications for economic decision-making in the face of uncertainty both on the enterprise and the national level. Moreover, the outcome of the study is a framework that can be applied in a wide range of areas like macroeconomics, energy economics, financial economics, and others.

I. Dissertation Overview

Probabilistic forecasting has clear advantages compared to the more established point forecasting, which is predominant in economics. In the context of rare events like the coronavirus pandemic and the Russian invasion in Ukraine and the resulting extreme economic volatility, probabilistic forecasting becomes indispensable. The main goal of the current dissertation is to determine the state of the art in probabilistic forecasting across the fields of economics, statistics, and machine learning and propose novel improvements, which can have practical benefits in the task of economic forecasting. In order to achieve this goal, a multidisciplinary study was performed and a new approach to probabilistic forecasting was proposed and applied to several problems of interest among economists.

Relevance and Significance of the Research

Since the beginning of the 21st century, there have been two major global events, which have posed a great challenge for both forecasters and decision-makers. The first is the global financial crisis of 2007-2009 and the following great recession, whose far-reaching repercussions affected many economies around the world. While the crisis was a realization of low probability risk (see Makridakis et al. 2009; Chen 2019), in retrospect it was evident that it was caused by a build-up of systemic risk, which turned out to be visible in the data (Altunbas et al., 2017).

The second event is the coronavirus pandemic of 2020, which forced many governments around the world to implement lockdowns, which in effect caused a sharp economic recession throughout the world. While similar to the global financial crisis, the pandemic was also a realization of a low-probability event (Antipova, 2021), it was a completely unexpected shock, which could hardly be considered predictable. Still, its effects on the economy could be quantified in the short term as is shown in the results from the current study, which is an important lesson for the future.

As this text is written, a war is waged in Ukraine. Apart from the humanitarian crisis and the existential consequences of this violent conflict, there are also serious economic effects like the threat to the global food supply and the unprecedented energy crisis which unfolded in 2022. Similarly to the other two examples, the uncertainty such events introduce affects nations, enterprises, and individuals and their ability to make informed decisions. Therefore, it is imperative to be able to forecast and measure uncertainty in the face of such low-probability events and unexpected shocks in order to make optimal decisions with respect to future outcomes.

The current study aims to formulate a framework, which can be used by decision-makers in both government and private organizations when it comes to forecasting indicators of importance. Moreover, the study aims to establish the use of probabilistic forecasts as a primary approach in situations characterized by elevated degrees of risk and uncertainty.

Object and Subject of Study

The object of study is economic uncertainty. This includes measurable degrees of uncertainty, which relate to the economic conditions within a specific region or globally, as measured by various economic indicators.

The subject of the study is forecasting economic uncertainty. Moreover, the focus of the study is the development of a new neural network architecture for probabilistic forecasting, which relies on the concept of quantile regression and the use of artificial neural networks.

Research Objectives and Tasks

The goal of the dissertation is to develop a general neural network architecture for modeling and forecasting economic risks and establish probabilistic forecasting as necessary in situations characterized by rare events and extreme shocks like economic crises and natural disasters. In order to achieve the research goal, the following research tasks are defined:

1. To perform an in-depth literature review on the topics of uncertainty in economics and machine learning, economic forecasting, density forecasting, and deep learning methods for probabilistic forecasting
2. To develop a deep learning architecture for economic modeling and forecasting, which can allow for uncertainty quantification and use in different economic contexts and time series forecasting in general
3. To evaluate the proposed deep learning architecture and its relative performance with respect to established benchmarks in various applications.
4. To apply the proposed deep learning architecture for the purposes of nowcasting the pandemic-related recession in several small open economies
5. To apply the proposed deep learning architecture for the purposes of forecasting the natural gas prices in Europe after the Russian invasion of Ukraine
6. To apply the proposed deep learning architecture for the purposes of constructing inflation fan charts for Bulgaria

The following tasks lead to the goal of the dissertation and demonstrate how the resulting framework can be applied to real-world cases.

Hypotheses

The primary research hypothesis of this study is that a general deep learning architecture designed for modeling and forecasting economic uncertainty can be formulated, which is expected to outperform a set of benchmarks, particularly in situations related to rare events and unexpected shocks. The working hypotheses are the following:

1. The proposed deep quantile-based probabilistic regression framework can outperform various statistical and machine learning benchmarks across a number of tasks like nowcasting the pandemic-related recessions, forecasting natural gas prices on the Balkan Gas Hub, and inflation forecasting for Bulgaria
2. A general family of distributions like the Sinh-Arcsinh family of distributions is suitable for modeling various economic indicators like GDP, inflation, and gas prices and the resulting forecasting densities have superior performance.
3. Methods borrowed from the machine learning and deep learning domains like using multiple loss functions, estimating multiple quantiles simultaneously, and augmenting traditional loss functions with custom terms, which address the quantile crossing problem can lead to performance improvements compared to established procedures.
4. Data for GDP, country-specific sentiment indicators and financial indicators related to overall stock market indices and bond yields covering the global financial crisis and the consequent recession can be useful in nowcasting the pandemic-related recessions

All working hypotheses are tested empirically. Hypotheses 1 and 2 required the analysis of various unconditional distributions of interest and an evaluation of their moments. Hypothesis 2 additionally requires the use of the Kolmogorov-Smirnov test to evaluate the compatibility of the proposed Sinh-Arcsinh family of distributions to the observed distributions of economic indicators of interest. The testing of hypotheses 3 and 4 employs techniques for performance evaluation and comparison between competing approaches.

Scope of Study

The study is focused on the modeling and forecasting of economic risks and especially in situations related to rare events and unexpected shocks like the global financial crisis of 2007-2009 and the coronavirus pandemic of 2020. Methodologically, the study relies heavily on established procedures in economic forecasting as well as refers to established and novel approaches from the fields of machine learning and deep learning.

In terms of the time scope of the empirical analysis, the study covers roughly the period after 2000 up to the present moment of writing the dissertation. Geographically, the study is concerned with both global and regional events and impacts. In the case study on nowcasting the pandemic-related recession, the geographical scope includes Bulgaria, Lithuania, Estonia, and Romania. The case study on forecasting natural prices is focused on Bulgaria and its interconnections with the European natural gas market. In the last case study focused on inflation forecasting the geographical scope is narrowed down to Bulgaria only.

Methodology of Study

The study relies on the scientific method, by applying careful observation and data collection, objectivism, and rigorous skepticism, formulating hypotheses through induction, verification of deductions based on the formulated hypothesis via measurement-based testing, and iterative refinement or rejection of the initial hypotheses on the basis of findings. For the purposes of the review of relevant literature, the methods of analysis and synthesis, as well as induction and deduction have been employed. On the other hand, the interdisciplinary approach allows for the understanding and evaluation of conclusions and results from various scientific fields like economics, econometrics, statistics, risk management, and machine learning.

The analysis of specific data modeling and forecasting procedures requires methods from the domains of statistics, econometrics, and machine learning. For the purposes of probabilistic and density forecasting several classes of procedures have been reviewed and used, from fully parametric methods like autoregressive conditional heteroskedasticity models, to fully non-parametric models like kernel density estimators to hybrid semiparametric procedures. Statistical hypotheses are formulated and formal tests are performed where necessary in order to support a given conclusion or justify the researcher's choices. Relevant evaluation techniques are performed and relevant metrics are calculated in order to assess the performance of competing procedures. These quantitative methods allow for the testing of the working hypotheses and the achievement of the dissertation's goals.

Data Sources

Since the main research hypotheses in this dissertation are validated empirically, the data is a fundamental prerequisite in this study. Most of the empirical analysis is done on data from primary data sources, which are indicated in the text, where it is relevant. Since these primary data providers have strict methodologies in collecting and disseminating data (e.g. Eurostat, IMF, NSI), the quality of the data was considered optimal and no further data quality improvement was performed. Some of the data used comes from cited sources, which is also clearly indicated.

All calculations relevant to the dissertation analysis are done using the programming language Python 3.x and R 4.x. The open-source framework TensorFlow (Abadi et al., 2016) originally developed by Google is used for all implementations of artificial neural networks. Normal computer hardware is used for most computations related to results in the dissertations. In rare use cases like artificial neural network hyperparameter optimization, a more specialized setup using GPU hardware was used.

Value and Originality of the Study

The listed dissertation contributions have potential value and utility for the fields of economics and machine learning. The relevant sub-fields would be respectively economic forecasting and deep learning.

Firstly, the discussion of uncertainty, its definition, and classification are useful in clarifying the term and concept in the context of economic forecasting. This leads to a common understanding, reduces semantic ambiguity, and allows for a clear distinction, which is instrumental in quantifying uncertainty. Since a study focused on economic probabilistic forecasting could suffer in terms of ambiguity and lack of precision if the term "uncertainty" is used as self-explanatory, it is valuable to use a precise definition.

The proposed DQPR model could be useful in a range of forecasting tasks in economics – both macroeconomic and microeconomic. Since forecasting is important for decision-making on any level of aggregation and uncertainty quantification has a proven information value, this approach is a valuable tool in such tasks. Since, the tasks of quantifying aleatoric and epistemic uncertainty are separable, one could focus only on the measure of uncertainty of interest. Moreover, the DQPR could be used in areas outside of economics, where time series forecasting and uncertainty quantification would be essential since the approach is very general in terms of its implementation.

The applications of the DQPR are novel in both scope and topic. The topics of the pandemic-related recessions and the volatility of the natural gas markets in Europe concern very recent events, which could have repercussions for decades to come. Demonstrating how to perform forecasting in periods characterized by elevated uncertainty would be beneficial for similar scientific studies in the future.

Limitations of the Study

The study has its limitations, which should be considered and discussed. Probabilistic time series forecasting is a very broad area of study across multiple disciplines like economics, econometrics, statistics, engineering, and machine learning to mention a few. Therefore, this study does not claim to be absolutely exhaustive in all methods and approaches to probabilistic forecasting, but instead, it focuses on the ones deemed the most influential and practical. Moreover, sub-topics like artificial neural networks or deep learning, Bayesian inference, and Bayesian neural networks are themselves quite vast, and it is hard to claim exhaustiveness in them as well. However, from all topics considered, it was aimed to distill the most important and influential, the most widely used, the most established, or on the other hand the most promising novel methods across the disciplines mentioned above.

With respect to the proposed novel deep quantile-based regression model, it is important to mention that although a considerable number of different architectures have been considered both by manual and automated iteration, one could not claim exhaustiveness in the search. Moreover, it is important to consider that for every specific task, the optimal architecture might be quite different and therefore, this architectural optimization is very much task-specific. On the other hand, no thorough analysis of the asymptotic properties of this approach has been carried out in order to fully verify its pros and cons. Also, no thorough cross-sectional comparison has been carried across both tasks and multiple methods, as this has been secondary to applying the approach to the problems at hand.

Finally, with respect to the applications of the proposed probabilistic forecasting approach, it should be mentioned that the focus was on the economy of Bulgaria and similar countries from Eastern Europe. Therefore, the approach has not been tested in the context of countries, which have a richer set of data like the USA, UK, or other EU countries like Germany or France. This was motivated mainly by the knowledge and acquaintance with the Bulgarian economy, but also the general lack of similar studies in the field focusing on Bulgaria or Eastern Europe.

Avenues for Future Research

The discussed limitations of the study open up several avenues for future research. Firstly, scientific fields outside the ones traditionally related to economics can be explored for advances and views on the topic of probabilistic time series forecasting. Such fields less related to economics where important and interesting conclusions and advances related to probabilistic forecasting have been made are meteorology, neuroscience, and medicine. On the other hand, due to the quick turnover of ideas and papers in fields like artificial neural networks and deep learning, it is appropriate to regularly review the newest literature in search of newer architectures, algorithms for optimization, uncertainty quantification techniques, or Bayesian inference methods.

With respect to the proposed DPQR, further research into its asymptotic properties would be appropriate. Moreover, it would be interesting to test across a broader array of tasks – even ones outside the scope of economics. For example, it would be interesting to apply to tasks of forecasting positive count data, where the Poisson distribution would be appropriate. On a similar note, its performance could be compared to a larger set of similar probabilistic algorithms. The hurdle there would be that most purely probabilistic methods are still not as accessible and easy to use as the traditional point forecasting methods.

Finally, it would be interesting to try and replicate the results from the studies performed on the pandemic-related recessions and the natural gas price during the war in Ukraine for a broader geographical scope. For the former topic of the pandemic-related recessions, it would be interesting to apply the proposed DQPR model to the aggregate EU time series, other separate EU members' data, or US data in order to compare its performance with other methods in existing studies. For the latter topic of forecasting natural gas prices, it would be interesting to apply the approach to data from other global gas hubs in the US and Asia. Such comparative studies can reveal the pros and cons of the approach in different contexts and further its development.

II. Structure of the Dissertation

The following dissertation consists of three chapters, introduction, conclusion, references and two appendices. The chapters follow logically the research objectives and tasks stated previously.

Table of Contents

The first chapter consists of two parts. The first part focuses on the definition and classification of economic uncertainty for the purposes of economic forecasting. It delves deeper into classifying uncertainty into aleatoric and epistemic and ways to quantify it. The second part of the first chapter is an in-depth literature review of economic forecasting and its development in the last century. The chapter concludes with a review of various methods used for the generation, calibration, and evaluation of probabilistic forecasts.

The second chapter establishes a novel neural network architecture for probabilistic forecasting based on deep learning and inspired by a proven technique from the economics literature – the deep quantile-based probabilistic regression (DQPR). A couple of alternative architectures are considered and empirical tests are performed to evaluate the training performance of the proposed model. A review of Bayesian inference in deep learning is performed and a method for disentanglement of aleatoric and epistemic uncertainty is presented. Finally, an approach to evaluate the proposed model against several benchmarks is selected and ways to perform sensitivity analysis are discussed.

The third and final chapter is concerned with specific applications of the proposed DQPR model in three separate case studies. The first case study deals with nowcasting the pandemic-related recessions in Bulgaria, Lithuania, Estonia, and Romania. The proposed framework is compared to a linear version of the same model, and the advantages of using deep learning are demonstrated in an empirical exercise.

The second case study demonstrates how the proposed framework can be applied to forecasting natural gas prices on the Balkan Gas Hub using data from the leading Europe TTF gas hub in the Netherlands. The performance of the proposed framework is compared to several statistical and machine learning methods and it is demonstrated that it outperforms the benchmarks.

The third case study is a demonstration of the ability to construct inflation fan charts using the proposed framework – a tool regularly used by central banks around the world. Apart from constructing density forecasts for inflation in Bulgaria over different forecasting horizons, it is also demonstrated how aleatoric and epistemic uncertainty can be disentangled in practice.

Finally, the conclusions summarize the contributions of the study, its limitations, and avenues for future work on the subject.

III. Main Results of the Dissertation

Chapter 1: Economic Forecasting and Uncertainty

Uncertainty – Definition and Classification for the Task of Economic Forecasting

The concept of uncertainty has been a topic of great interest to scholars across both natural and social sciences. It has played an important role across a diverse set of fields and thus over time the term "uncertainty" has adopted various meanings. The first recorded attempts to define uncertainty, were by the Greek philosophers of the school of Athens (Bernstein, 1996). The ancient Greek word εικος (eikos), which could be translated as probable or plausible, was defined by Socrates as "likeness to truth". Aristotle also theorized on the topic of decision-making in his *Nicomachean Ethics*, yet he did not delve deep into the concept of uncertainty, but rather accepted luck or chance as a given. On the other hand, while games of chance seem to be as old as history with some early known examples from Ancient Egypt dating back to 3500 BC, it was not until the Renaissance, that gambling was used as a foundation to study uncertainty (Bernstein, 1996).

In 1654, a French nobleman Chevalier de Méré challenged the famous mathematician Blaise Pascal to solve a puzzle, which has confused mathematicians for some two hundred years, when it was posed by the monk Luca Paccioli. The puzzle was how to divide the winnings of an unfinished game of chance between two players, while one of them is ahead. With the help of another brilliant mathematician, Pierre de Fermat, Pascal laid the foundation of what we know today as the theory of probability. In consequence, various great thinkers and scholars have continued to explore and revisit the concept of uncertainty like Bayes, Bernoulli, and Galton to name a few. At the beginning of the 20th century, the concept of uncertainty already played a critical role in the analysis in various fields from physics to psychology and this is also when some of the most prominent economic thinkers laid out some fundamental groundwork on the concept of risk and uncertainty.

Research focused on modeling and forecasting uncertainty can suffer from ambiguity and lack of precision, without clearly defining the concept of uncertainty and by using the term as self-explanatory. This text aims to establish a working definition and classification of uncertainty for the task of economic forecasting. This is necessary in order to arrive at a common understanding of the term, reduce semantic ambiguity and define a clear distinction when it comes to quantifying forecast uncertainty. Recently, in the fields of engineering and machine learning a separation of uncertainty into aleatoric and epistemic has grown in popularity. Aleatoric uncertainty roughly refers to the inherent stochasticity in the environment or its measurement, while epistemic uncertainty refers to the limitations of the knowledge of the observer. The practicality and ambition for clarity and unambiguous separation make this classification appealing and considered suitable in the context of economic forecasting. Therefore, two fundamental sources on uncertainty by John Maynard Keynes and Frank H. Knight, which define and explore the concept in the economics literature are reviewed from the perspective of the classification of uncertainty into aleatoric and epistemic. Consequently, the concepts of aleatoric and epistemic uncertainty are explored and the possible ambiguity and interaction between them are discussed. Finally, a working definition and classification of uncertainty are laid out and refined for practical use in the context of economic forecasting.

For some time, uncertainty had no place in economics (Davidson, 1999). Classical economics theory dealt with agents possessing perfect information about the outcomes of their decisions, and uncertainty was simply ignored. In the year 1921, two seminal works were published by two up-and-coming economists, who eventually became the founding fathers of two prominent schools of economics: John Maynard Keynes published his dissertation on probability theory and Frank H. Knight also published a revised version of his dissertation, which dealt with uncertainty and judgment. These two modern thinkers and their seminal works laid the foundations of what later became two separate schools of economic thought – the Keynesian school of macroeconomics (Snowdon and Vane, 2015) and the Chicago school of microeconomics (Emmett, 2009).

On the surface, it appears like these foundational works have developed a similar theory of uncertainty, which has led some to call it “the Knight-Keynes uncertainty concept” (e.g. Davidson, 1972; Hodgson, 2011). However, Packard et al. (2021) who have performed a critical review of the historical records and the works of Keynes and Knight, argue that fundamentally the two thinkers differed in their political views, scientific epistemologies, their ontological beliefs, and ultimately their views on uncertainty. The authors believe that historically, Keynes’ and Knight’s theories of uncertainty and their political philosophies, in general, have been wrongfully homogenized, which is in stark contrast to the differences in the schools of thought each of them laid the foundations of. Although it is undeniable there are similarities between their concepts of uncertainty, there are also nuanced differences, which led them to different conclusions about the nature of uncertainty and how one should deal with it.

Keynes was an objectivist, a positivist, and a determinist, who believed in an objective deterministic reality, which should be studied via empiricism and rationalism. In this light, the problem of uncertainty arising from the limitations, ignorance, and irrationality of an actor, can be continually mitigated through systematic scientific inquiry and the constant pursuit of new evidence. However, such a view would reject the notion that certain aspects of reality might be inherently uncertain even in the case of perfect knowledge.

According to Packard et al. (2021), Knight’s view of uncertainty, in light of his worldview and especially his opposition to positivism in economic analysis, is an aleatoric one – related to the inherent stochasticity and unpredictability of processes, which is also irreducible by the accumulation of evidence. Others, among which Friedman (2007), interpret Knightian uncertainty as epistemic and thus similar to Keynes’ view, an interpretation which has dominated mainstream economics. However, Knight (1921) himself implies on numerous occasions that in economics you have “a larger proportion of factors … of the variable and fluctuating sort” and states that “it is a world of change in which we live, and a world of uncertainty.”

It appears that Knight believed that uncertainty was inherent in reality, which points to uncertainty in the aleatoric sense. However, he often referred to the subjective, the psychological, and the knowledge of the individual, which means he also perceived uncertainty in the epistemic sense. In his own words, he seems to extend his view in the epistemic sense, not only to the future but to the present as well, which seems to coincide with ideas from psychology. Perhaps, in his rejection of positivism and the treatment of social science and economics in particular as an exact science, Knight achieved a more broad and comprehensive view of uncertainty, compared to Keynes, who believed in the predictability of human behavior and perceived uncertainty as almost exclusively epistemic in nature.

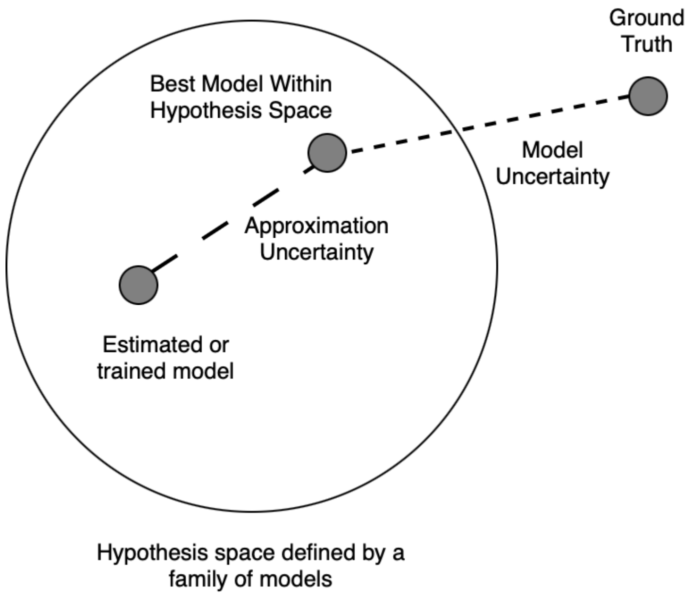
While Frank H. Knight's definition and dichotomy of risk and uncertainty remains hugely influential in economics, its direct application to economic forecasting seems limited from the contemporary standpoint. Using Knight's classification in the context of forecasting can only be performed ex-post, after the realization of the event of interest, because at the time of forecasting, one could not identify what type of situation one is facing. Therefore, Knight's classification is still useful in an ex-post analysis, but has limited value at the time of forecasting. In fact, one could even say that at the time of forecasting, forecasters always act as if they are in a situation of risk, where full information of the possible outcomes and the associated probabilities is known. Therefore, it is appropriate to turn to a more contemporary classification of uncertainty, which was already mentioned, but was not elaborated on – the division of uncertainty into aleatoric and epistemic. This alternative classification is deemed more practical for the context of economic forecasting because it can be applied both at the time and forecasting and ex-post and allows for a more detailed quantitative analysis.

The division of uncertainty into aleatoric and epistemic seems to have originated from the field of engineering (Hora, 1996; Faber, 2005; Kiureghian and Ditlevsen, 2009), although it is hard to find a scientific study that mentions the origins of the terms. Due to its practicality, it has been used in computer science and machine learning (Dutta, 2013; Shaker and Hüllermeier, 2020; Hüllermeier and Waegeman, 2021; Lai et al. 2022) and in economics as well, although rarely (Dequech, 2004; Packard and Clark, 2020; Curto, Acebes and González-Varona, 2022).

Aleatoric (or statistical) uncertainty refers to the uncertainty related to the inherent stochasticity or randomness in data-generating processes or in the outcome of an experiment (Hora, 1996; Hüllermeier and Waegeman, 2021). The term comes from the Latin word alea, which means bone or dice, since bones were used as dice in gambling games (Lewis and Short, 1879). It is assumed that with the current knowledge about a data-generating process, there is a degree of uncertainty that cannot be reduced by accumulating more evidence or changing or refining the statistical model of the given process. However, this uncertainty can be identified and quantified. A subtle convenience of the concept of aleatoric uncertainty is that it can be justified despite the ontological view of its user. It fits and can be used in the contexts of both determinism and indeterminism.

The word epistemic originates from the Ancient Greek word επιστημη (epistémē), which means knowledge. Epistemic uncertainty arises from the lack of knowledge of the observer, economic decision-maker, or forecaster regarding the data-generating process (Hüllermeier and Waegeman, 2021). Epistemic uncertainty coincides very much with Keynes' view of uncertainty, which could be reduced by the accumulation of greater evidential weight or knowledge about the system in general. Thus, in contrast to aleatoric uncertainty, the main characteristic of epistemic uncertainty is that it can be reduced in principle (Hora, 1996). If this concept is applied to a forecasting task, epistemic uncertainty might relate to uncertainty related to the family of statistical or machine learning models chosen for a given task or the amount of data available. One could reduce this uncertainty, by picking a better model of the data-generating process or by collecting more relevant data, in order to estimate the model parameters more precisely.

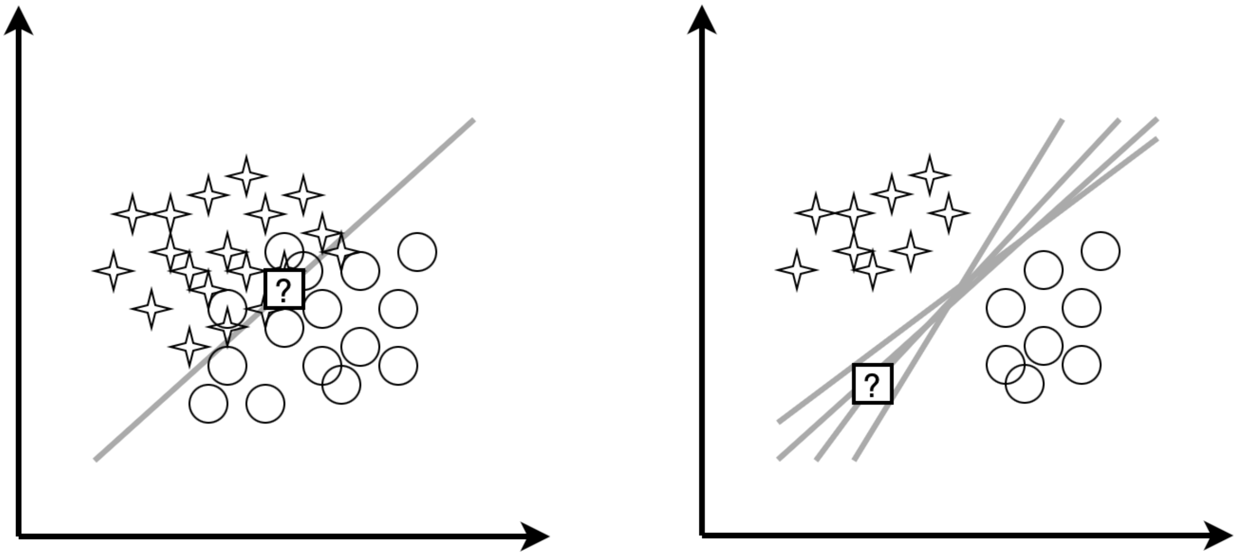
Figure 1: Types of Epistemic Uncertainty



Source: Author, based on Hüllermeier and Waegeman (2021)

Epistemic uncertainty can be further reduced into at least two sub-categories: model uncertainty and approximation uncertainty (Hüllermeier and Waegeman, 2021) as shown in Figure 1. Approximation uncertainty relates to the uncertainty surrounding the model parameters and can be expressed as the difference between the chosen hypothesis or model and the optimal hypothesis within the chosen hypothesis space (family of models). Model uncertainty refers to the choice of the hypothesis space or family of models in general, and can be expressed as the difference between the ground truth (or the population model) and the optimal hypothesis within the hypothesis space. If the task of economic modeling or forecasting is concerned, model uncertainty refers to both the choice of a family of models, but also the predictors used for modeling. For example, using a linear model, when modeling a quadratic relationship might be a source of epistemic uncertainty of the model sub-category. A missing variable bias might again be a source of epistemic uncertainty of the same kind. An example of approximation uncertainty in a simple linear regression context would be uncertainty surrounding a regression coefficient, which is usually expressed via a confidence interval. According to Hüllermeier and Waegeman (2021), given a consistent estimator, asymptotically one could eliminate approximation uncertainty by increasing the number of observations.

Figure 2: Aleatoric and Epistemic Uncertainty



Source: Author, based on Hüllermeier and Waegeman (2021)

Figure 2 demonstrates how epistemic uncertainty differs from aleatoric uncertainty in the context of a simple classification model. On the left-hand side, at the point denoted by a question mark, the prediction is aleatorically uncertain, due to the overlap of the two classes in this region around the decision boundary. On the right-hand side, the point denoted by a question mark is a case of epistemic uncertainty, due to the lack of knowledge about the model parameter, which in turn is caused by the lack of enough data.

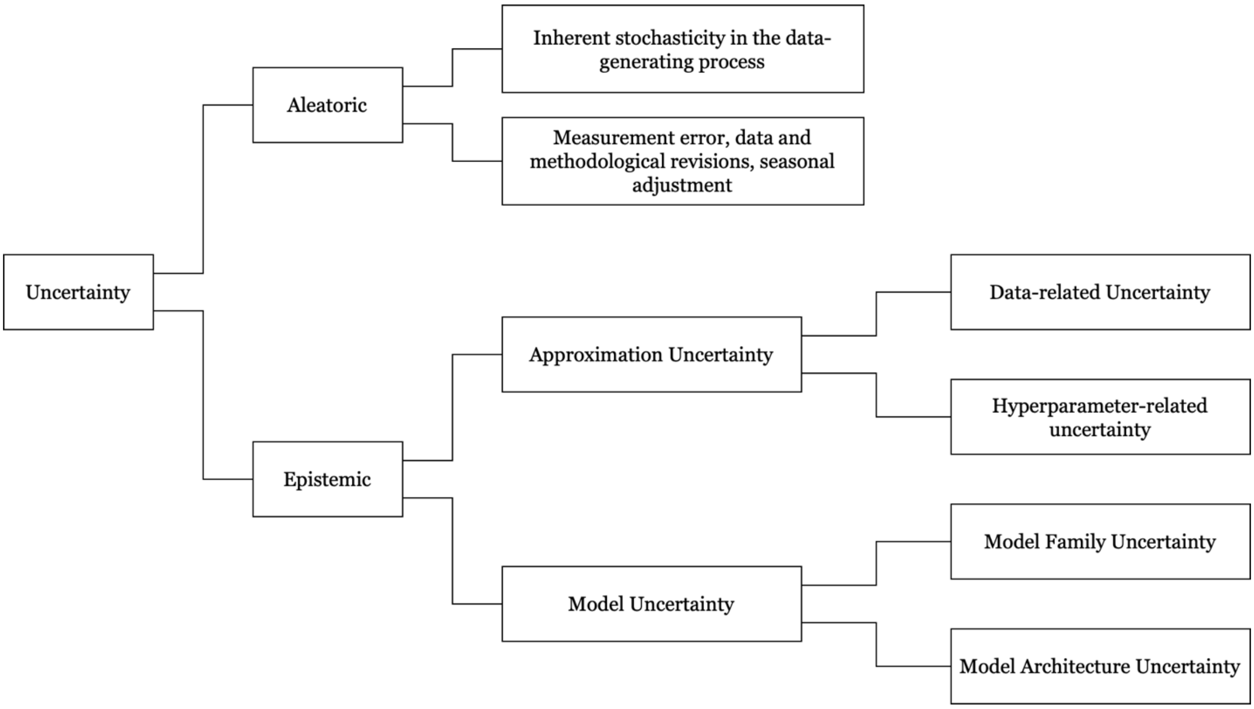
Based on the discussion so far, uncertainty could be defined as the lack of certain knowledge or understanding about the realizations of a given situation or event. Uncertainty can be considered a fundamental characteristic of the future, but as Knight (1921) points out it could extend to the present as well, especially when complex interactions are concerned. Uncertainty stems both from the stochasticity inherent in the environment as a whole or in a given data-generating process, as well as from the limited knowledge of the observer or forecaster, who attempts to model the said data-generating process.

When aleatoric uncertainty is considered, it can be expressed in the following way. First and foremost, $\mathcal{D}^T = (Y, X, X_{T-h})$ is the information set available to the forecaster. Y is a vector of the target variable or dependent variable. X is a vector of the predictors or the independent variables up to time T and X_{T-h} is a vector of predictors available after time T in order to generate forecasts Y_{T-h} for a forecast horizon with length h. We can measure the aleatoric uncertainty in a simple way using the unconditional or empirical distribution of Y given by $f(Y)$ however this allows to make only naive forecasts. The predictive distribution $f(Y_{T+h}|\mathcal{D}^T)$ is what is referred to as a density forecast since it describes the distribution of Y for future values and is conditional on the information set.

Assuming that $f(Y_{T+h}|\mathcal{D}^T)$ is generated using a hypothesis or model $H(\theta)$ with a vector of parameters θ, the epistemic approximation uncertainty can be defined as the uncertainty around the parameters, expressed by the conditional distribution $f(\theta|\mathcal{D}^T)$. As was mentioned previously, asymptotically increasing the size of the information set $N \to \infty$ would in principle eliminate this approximation uncertainty. However, another important aspect of approximation uncertainty is hyperparameter optimization. Given a fixed hypothesis space, optimizing the hyperparameters of the learning algorithm can reduce the distance toward the optimal model within the hypothesis space and thus reduce epistemic approximation uncertainty. Therefore, two subtypes of approximation uncertainty are data-related uncertainty and hyperparameter-related uncertainty.

Given the simple definition given above, Figure 3 presents the classification of uncertainty in the context of an economic forecasting task.

Figure 3: Classification of Uncertainty for Economic Forecasting



Source: Author

While it seems that the advantage of density forecasting over point forecasts is established, it is important to analyze uncertainty according to its source. Hüllermeier and Waegeman (2021) argue that uncertainty analysis "for individual instances, is arguably important and practically more relevant than a kind of average accuracy or confidence, which is often reported in machine learning". They give the example of medical diagnosis, where a patient would be interested in the reliability of the prediction in her/his case rather than some average reliability measure of the model. Kull and Flach (2014) also argue that assigning reliability scores to each instance is much more powerful compared to assigning an aggregate reliability score. The additional information provided by a quantification of uncertainty in this transductive way, would affect and perhaps improve decision-making due to a greater transparency and information value.

One could see how this view can be applied to economic forecasting as well, in order to provide greater informational value when communicating forecasts.

In this critical review, the theories on uncertainty of Keynes and Knight were reviewed from the point of view of a classification of uncertainty, which originates from engineering and is becoming predominant in machine learning. Its main premise is separating uncertainty as one inherent in the environment or related to its measurement (aleatoric) and one arising from the limitations of the forecaster and her/his knowledge (epistemic). Due to the practical convenience and transparency of this approach to uncertainty, it is adopted for the task of economic forecasting. This classification of uncertainty is enriched and adjusted for the context of economic forecasting and outlines the subclasses of aleatoric and epistemic uncertainty. Conclusively, it is established that a forecaster should not avoid an assessment of uncertainty and should attempt to dissect the uncertainty in order to increase the informational value of her/his forecasts.

Economic Forecasting and the Development of Density Forecasting

Towards the end of the 19th century, there was a transition from point estimates to distribution estimates in the field of statistics according to Stigler (1975). Gneiting (2008) describes a similar shift in interest from point forecasts to probabilistic forecasts across many fields, and economics makes no exception. The review of forecasting literature done by Diebold and Lopez in 1996 reveals that when it comes to forecast evaluation, the topic of point forecasts evaluation dominated the field at the time. Few articles were concerned with the evaluation of prediction intervals (Chatfield, 1993; Christoffersen, 1998) or probability forecasts (Wallis, 1993; Clemen et al., 1995). Furthermore, Diebold et al. (1998) believe that until the advent of quantitative finance and risk management, there was little demand for interval or density forecasts within the economics field. The practice of forecasting in itself is an attempt to study an uncertain future, and probabilistic forecasts expressed as probability distributions over expected future realizations are a prime way to measure the degree of uncertainty.

Currently, in the field of forecasting the simplest way to measure uncertainty related to a forecast is via confidence and prediction intervals. Chatfield (1993) describes interval forecasts as consisting of upper and lower limits associated with a predefined probability. These upper and lower limits define the range in which a future value of the random variable would fall with some level of confidence. Hansen (2006) elaborates that interval forecasts are often constructed around point forecasts as an additional measure of uncertainty. Indeed, as many forecasting methods are tailored towards generating point forecasts, calculating intervals is a straightforward way to quantify the uncertainty around such forecasts.

Another concept, quantile regression, which originally dates back to the 18th century, was more recently re-introduced by Koenker and Bassett (1978) and applied in economic analysis in various studies. Fitzenberger, Koenker, and Machado (2002) presented a number of economic studies which utilized quantile regression. According to them, it was not until the 1990s that the technique gained larger popularity among economists and econometricians. Some noteworthy studies the authors present in the book *Economic Application of Quantile Regression* (2002) are Fitzenberger et al. (2001) who studied the wage structure in West Germany, García et al. (2001) who use quantile regression to investigate gender wage differences in Spain and Buchinsky (2001) who analyzed returns to education among women in the US.

As Koenker and Bassett (1978) show a task of sorting can be turned into an optimization problem. Just as finding a sample mean can be done by minimizing the sum of squared errors, finding the median can result from minimizing the sum of absolute errors. Koenker and Bassett (1978) further elaborate to show that an asymmetrical loss function which gives different penalties to positive and negative residuals, can yield any quantile for a given sample. Solving for the following equation (1) yields the τ-th quantile as its solution:

$$\boldsymbol{min}_{\xi \in \mathbb{R}} \sum_{\boldsymbol{i}=0}^{\boldsymbol{n}} \boldsymbol{\rho_\tau}(\boldsymbol{y_i} - \boldsymbol{\xi})$$

(1)

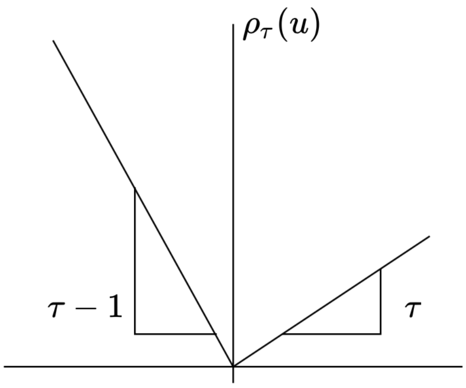
Where $0 < \tau < 1$ and $\rho_\tau(\cdot)$ is the titled absolute value function, which can be seen in Figure 4, for a sample of size *n*. In this equation, if τ is set to equal 0.5, the equation will yield the median. Therefore, if the scalar ξ in equation (1) is replaced with a parametric function $\xi(x_i, \beta)$ and τ is set to equal 0.5, one could obtain the estimate of the conditional median function.

$$\boldsymbol{min}_{\xi \in \mathbb{R}} \sum_{\boldsymbol{i}=0}^{\boldsymbol{n}} \boldsymbol{\rho_\tau}(\boldsymbol{y_i} - \boldsymbol{\xi}(\boldsymbol{x_i}, \boldsymbol{\beta}))$$

(2)

Setting τ to different values will lead to the estimation of different conditional quantile functions.

Figure 4: Tilted absolute value function



Source: Author, based on Koenker (2005)

In general, we would model the relation between the conditional quantile of y_{t+h} and a vector of predictors X and optionally their lags, for a given time period t and a forecasting horizon h. In order to estimate the quantile regression of y_{t+h} on X, the regression coefficients β_τ for a given τ is chosen to minimize the weighted absolute value of errors:

$$\hat{\beta}_\tau = \mathrm{argmin}_{\beta_\tau \in \mathbb{R}^k} \sum_{t=1}^{T-h} (\tau \cdot 1_{(y_{t+h} \geq X\beta_\tau)} |y_{t+h} - X\beta_\tau| + (1-\tau) \cdot 1_{(y_{t+h} < X\beta_\tau)} |y_{t+h} - X\beta_\tau|)$$

(3)

where $\mathbf{1}(\cdot)$ is the indicator function, which subsets negative and positive errors, and T is the total length of the time series. The output value from the model is the quantile of y_{t+h} conditional on the model input X:

$$\hat{Q}_{y_{t+h}|X}(\tau|\mathrm{X}) = X\beta_\tau$$

(4)

This method allows one to estimate a quantile regression model to estimate any arbitrary quantile, conditional on the predictors. However, if one would like to estimate several different quantiles, one might run into the so-called crossing problem, which multiple scholars have run into and tried to address in one way or another (see Koenker, 1984; Cole and Green, 1992; He, 1997; Bondell et al., 2010; Rodrigues and Pereira 2020). Among the more interesting solutions are the ones proposed by Bondell et al. (2010) and Rodrigues and Pereira 2020.

In a seminal paper, Adrian et al. (2019) used a two-step procedure of constructing conditional quantiles using a quantile regression model and consequently fit a probability distribution to the estimated quantiles. The authors studied the conditional US growth distribution with an emphasis on financial conditions and their dynamics during economic recessions. The authors identified several stylized facts about the conditional distribution of growth for the USA, among which a strong negative correlation between the conditional mean and variance and a significant relationship between current financial conditions and future shifts in the lower tail of the conditional distribution. Similar conclusions were confirmed by De Santis and Van der Veken (2020), who performed a similar exercise including data from the beginning of 2020 and a separate dataset covering the Spanish flu pandemic period across a number of countries. Figueres and Jarociński (2020), confirm the same stylized facts identified by Adrian et al. (2019) for the Euro Area.

Quantiles of the conditional distribution of GDP growth in this framework are expressed as functions of the observed predictors. After generating the conditional quantiles, one could fit a probability distribution function to them in order to generate a density forecast. Adrian et al. (2019) propose using a skewed t-distribution for this purpose. In order to estimate the four parameters related to the skewed t-distribution, the problem can be formulated as a least squares optimization problem, using the estimated conditional quantiles[1] and the inverse cumulative probability function:

$$\{\hat{\mu}_{t+h}, \hat{\sigma}_{t+h}, \hat{\nu}_{t+h}, \hat{\alpha}_{t+h}\} = \text{argmin}_{\mu,\sigma,\nu,\alpha} \sum_{j=1}^{J} \left(\hat{Q}_{y_{t+h}|\mathrm{X}}(\tau_j|X) - F^{-1}(\tau_j; \mu, \sigma, \nu, \alpha)\right)^2$$

(5)

where $\hat{\mu}_{t+h} \in \mathbb{R}$ (mean or location shift), $\hat{\sigma}_{t+h} \in \mathbb{R}^+$ (standard deviation or scaling parameter), $\hat{\nu}_{t+h} \in \mathbb{R}$ (skewness parameter), and $\hat{\alpha}_{t+h} \in \mathbb{R}^+$ (kurtosis or tail weight parameter). $F^{-1}$is the inverse cumulative distribution function and $\hat{Q}_{y_{t+h}|X}(\tau_j|X)$ is the estimated quantile of y_{t+h} for a given τ and conditional on X. This method can be used to estimate a density based on the conditional quantiles, as well as the unconditional or observed quantiles of the actual GDP growth.

A full probabilistic forecast (or density forecast) is a forecast expressed as a probability distribution, instead of a single value, which would be considered a point forecast. Point forecasts are often a central feature of a probability distribution like the conditional mean or conditional median, which can be arrived at by optimizing the loss functions of respectively mean squared errors and mean absolute errors. Density forecasts can be expressed as the parameters that describe a probability distribution, or as they are formally called the moments of the distribution. For many families of probability distributions like the normal skewed distribution, the skew t-distribution, and the Sinh-Arcsinh distribution these include the mean (or location), the variance (or scale), the skewness and kurtosis (the last two are sometimes referred to as shape parameters). In the time series context, which is predominant in economics and econometrics a density forecast over horizon h, is expressed as the forecasted moments of a probability distribution for each time within the horizon.

Following Bassetti et al. (2019), the basics of density forecasting could be established using the context of a multiple linear regression model without an intercept for convenience:

$$y_t = X_t^T \beta + \varepsilon_t$$

(6)

where $t = 1, \ldots, T$ and $\varepsilon_t \sim i.i.d.(0, \sigma^2)$. β is a (m × 1) vector of coefficients, σ^2 is the variance of the error term ε_t, X_t is a (m × 1) vector of covariates or predictors, which can include exogenous variables z_t and lagged values of the dependent variable, $Y_{t-p}, p > 0$. A direct method to compute a density forecast is to assume the distribution for the error term and ignore parameter uncertainty. A usual assumption is the one of normality - $\varepsilon_t \sim N(0, \sigma^2)$. This would account for aleatoric uncertainty, but ignore the epistemic one.

[1] The .05, .25, .75 and .95 quantiles are used for the estimation of the conditional distribution.

The h-step ahead density forecast, conditional on the information available in the information set up to time T would be:

$$f(Y_{T+h}|\mathcal{D}^T) = N(X_T^T\hat{\beta}, \hat{\sigma}^2)$$

(7)

where $\hat{\beta}$ and $\hat{\sigma}^2$ can be computed either analytically or numerically. In this case, the variance is fixed for a given estimation of the model.

There are numerous methods for generating probabilistic forecasts, but no comprehensive toolkit for this. In economics, the autoregressive conditional heteroskedasticity (ARCH) model, seminal work in economics by Engle (1982), aimed to model and forecast volatility on the stock market and is still used widely in explicitly modeling conditional variances. A different approach of modeling explicitly all moments of the distribution of a target variable is defined by the generalized additive models for location, scale and shape - GAMLSS (Rigby and Stasinopoulos, 2005). In this model, the moments of a given distribution are modeled via separate equations, which are estimated jointly either via maximum likelihood estimation or Bayesian methods. Within the field of machine learning and specifically deep learning, multiple approaches based on artificial neural networks have been proposed. Some notable examples are Gal and Ghahramani (2016), who propose a theoretical framework called Monte Carlo dropout, Salinas et al. (2019) present the DeepAR model which is an autoregressive recurrent neural network, which performs probabilistic forecasting and is specifically tailored to forecast a large number of time series, and Alexandrov et al. (2020) present their forecasting package for Python called GluonTS utilizing transformer and wavenet architectures for probabilistic forecasting..

With density forecasting, the task of evaluating a forecast is harder, because one needs to compare forecast or predictive densities with a single ground truth value. Therefore, numerous studies have been focusing on developing and refining methods for density forecast evaluation. As Gneiting et al. (2007) define two separate aspects in which density forecast needs to be evaluated – calibration and sharpness. Calibration refers to the statistical consistency between the density forecasts and the observed, while sharpness is understood as the concentration of the density forecasts. A higher concentration of the density forecasts is better, subject to calibration. As Mitchell and Wallis (2011) point out, sharpness is a property of the predictive distributions alone, while calibration is a property of the forecast-observation pairs.

Scoring rules assign numerical scores to probabilistic forecasts based on the predictive distribution and the realization of the forecasted variable. They conveniently summarize the predictive performance, when the quality of a probabilistic forecast is evaluated. Such scoring rules include the logarithmic score and the continuous ranked probability score (CRPS). However, it is also worth evaluating the accuracy of the central features of the predictive distributions, like the mean and median. This can be performed by treating these central features like point forecasts. Therefore, traditional metrics for point-forecast accuracy like MSE, RMSE, or MAE are absolutely valid in this respect. See Gneiting (2011) for a discussion on the topic of the evaluation of point forecasts.

Tracing the historical development of economic forecasting allows one to understand the dichotomy between structural and non-structural modeling and forecasting. The former is driven and strongly underpinned by theory and is often used for scenario simulations and testing theoretical propositions, while the latter proved to be more practical and accurate in the forecasting profession. An enormous body of literature proposed numerous non-structural methods and procedures for generating time series forecasts, and numerous studies dealt with the properties of these methods and their evaluation. As of recently, many new methods for probabilistic forecasting have been developed and refined, due to the evident necessity to quantify the uncertainty surrounding forecasts.

A clear trend towards probabilistic forecasting is observed in the scientific literature in general, due to the evident advantages over point forecasts discussed as early as the late 60s of the 20th century, but even more often after 2000. Many studies in economics and even in machine learning still focus on point forecasting. However, many researchers as well as institutions like central banks have already recognized the advantage of using density forecasts. On the other hand, the practical tools for generating and evaluating density forecasts are still not as accessible and established as the instruments now widely used for point forecasting.

Chapter 2: Probabilistic Forecasting Using Artificial Neural Networks

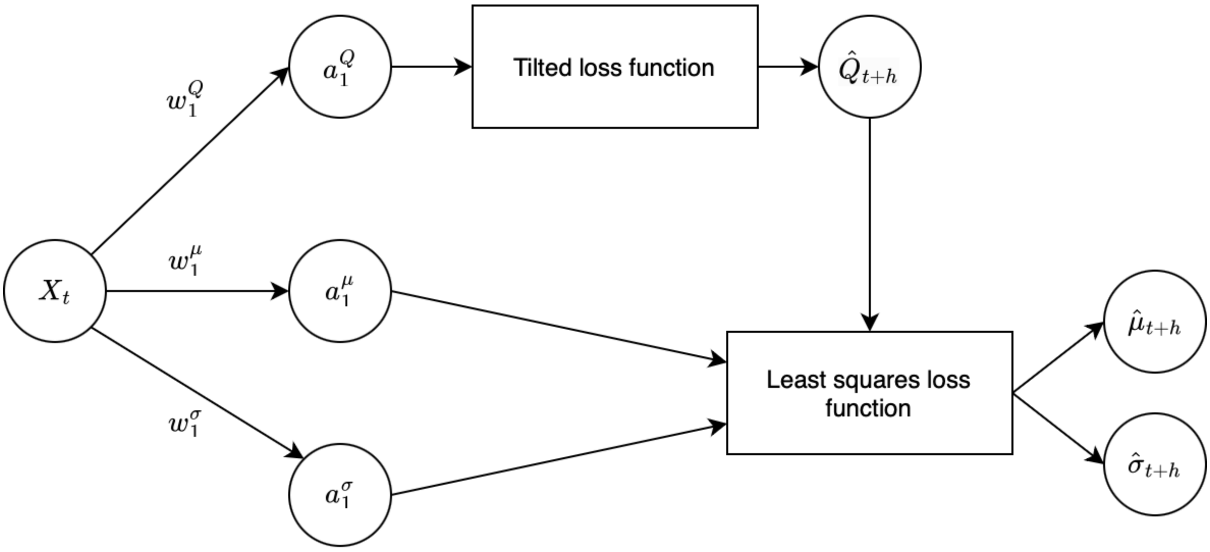
Quantile-based methods are often used for interval or density forecasting. There is a vast and growing economics literature that deals with quantile-based methods, their estimation, evaluation, and applications to various forecasting tasks. On the other hand, there are interesting developments on the same topic in the machine learning field. Lately, the machine learning field in general, and the deep learning subfield in particular, have been a source of many innovations. Deep learning is the subfield concerned with artificial deep neural networks. Therefore, it is only natural to transfer these innovations and apply them to the field of economics. See Cook and Hall (2017) for a great overview of different artificial neural network architectures in the context of economic forecasting.

A novel artificial neural network architecture is proposed for the purposes of probabilistic forecasting of time series, which is based on the estimation of conditional quantiles and outputs predictive densities. It is inspired by a two-step procedure used in a seminal paper by Adrian et al. (2019), but implements conditional quantile estimation within a neural network architecture, employs simultaneous estimation of quantiles similarly to Rodrigues and Pereira (2020), and outputs predictive densities in a single inference step.

The motivation behind this proposition was to address two issues. First, it is practically convenient to have the two steps in the procedure described above contained in a single model. The simultaneous estimation of an arbitrary number of conditional quantiles eliminates the necessity to estimate a separate quantile regression for each quantile level. As discussed below, this addresses the notorious quantile crossing problem. On the other hand, it is convenient to estimate the predictive density in a single inference step instead of two separate steps. Secondly, compared to other probabilistic models which estimate the parameters of a predictive distribution directly, one could argue this approach is more explainable and transparent. Any change in the variance or shape parameters could be related to the observed behavior of the quantiles. Also, any change in the quantiles could be explained by their relationship to the chosen predictors.

In order to simultaneously estimate conditional quantiles and output predictive densities, the network architecture depends on the use of two loss functions – the tilted loss function used for the estimation of the quantiles and a least squares loss function for the estimation of the final predictive density. This novel architecture is called a deep quantile-based probabilistic regression or DQPR for short. The simplest version of the proposed architecture would include the estimation of a single quantile (for example the median) and a simultaneous estimation of a predictive density following the normal distribution, as is seen in Figure 5.

Figure 5: Simple Version of the DQPR



Source: Author

Here X_T is the set of predictors or regressors and, for this formulation, it is assumed that it contains only the first lag of the target variable. Therefore, X_t is a $(n \times m)$ vector where n is the number of observations, m is the number of predictors, which in this case is equal to 1. The target variable y is a $(n \times 1)$ vector. The network is defined by the following equations at inference time:

$$\hat{Q}_{t+h} = a_1^Q = g(X_t w_1^Q + b_1^Q)$$

(8)

$$\hat{\mu}_{t+h} = a_1^{\mu} = g(X_t w_1^{\mu} + b_1^{\mu})$$

(9)

$$\hat{\sigma}_{t+h} = a_1^{\sigma} = g(X_t w_1^{\sigma} + b_1^{\sigma})$$

(10)

Where a_1^Q, a_1^{μ}, and a_1^{σ} are the activations, which are equivalent to the outputs of the neural network - $\hat{Q}_{T+h}$ which is the conditional median, $\hat{\mu}_{T+h}$ the conditional mean of the predictive density and $\hat{\sigma}_{T+h}$ the conditional variance of the predictive density. $g()$ is an arbitrary activation function, which in this case is a linear activation or in other words, no transformation of the product of inputs and weights. The weights w_1^Q, w_1^{μ}, and w_1^{σ} are $(m \times 1)$ vectors and the $b_1^Q, b_1^{\mu}, b_1^{\sigma}$ are bias terms or intercepts, which are scalars.

The tilted loss function (also known as pinball or quantile loss) used to estimate a single conditional quantile is defined by:

$$\mathrm{Q}L = \sum_{t=1}^{T-h} (\tau \cdot \mathbf{1}_{(y_{t+h} \geq X_T w_1^Q)} \left| y_{t+h} - X_t w_1^Q \right| + (1-\tau) \cdot \mathbf{1}_{(y_{t+h} < X_T w_1^Q)} \left| y_{t+h} - X_t w_1^Q \right|)$$

(11)

Where $\mathbf{1}()$ is the indicator function used to subset negative and positive residuals. In the case that there are multiple quantiles of interest, let $\{\tau_j\}_{j=1}^J$ be a set of J quantile levels.

One way to estimate the different conditional quantiles would be to fit a quantile regression for each level. This is not only inefficient, but can also cause the notorious crossing problem, which quantile regressions often have (Rodrigues and Pereira, 2018) The quantile loss function can be generalized for the case of multiple quantile levels, in order to reflect the sum of the individual loss function of the different quantile levels.

$$\mathrm{Q}L = \sum_{j=1}^{J} \sum_{t=1}^{T-h} (\tau_j \cdot \mathbf{1}_{(y_{t+h} \geq X_T w_1^{Q_j})} \left| y_{t+h} - X_t w_1^{Q_j} \right| + (1-\tau_j) \cdot \mathbf{1}_{(y_{t+h} < X_T w_1^{Q_j})} \left| y_{t+h} - X_t w_1^{Q_j} \right|)$$

(12)

In order to further address the crossing problem, Bondell et al. (2010) propose adding an additional term, which would be called here crossing loss:

$$CL = \sum_{j=1}^{J-1} max(0, X_t w_1^{Q_\tau} - X_t w_1^{Q_{\tau+1}})$$

(13)

In this formulation, if a lower quantile has a value greater than a higher quantile, the loss function would have a value of their difference, otherwise, it would have a value of zero. Therefore, this term could be added to the tilted loss defined above.

Finally, the least squares loss function which is responsible for estimating the predictive density parameters is as follows:

$$\mathrm{LSL} = \sum_{j=1}^{J} (\hat{Q}_{y_{t+h}|\mathrm{X}}(\tau_j|X_t) - F^{-1}(\tau_j; \mu, \sigma))^2$$

(14)

where $\hat{\mu}_{t+h} \in \mathbb{R}$ (mean or location shift), $\hat{\sigma}_{t+h} \in \mathbb{R}^{+}$ (standard deviation or scaling parameter). $F^{-1}$is the inverse cumulative distribution function and $\hat{Q}_{y_{t+h}|X}(\tau_j|X)$ is the estimated quantile of y_{t+h} for a given τ and conditional on X_t. If a four-parameter family of distribution like the Sinh-Arcsinh distribution is used, the loss function would be extended in the following way:

$$\mathrm{LSL} = \sum_{j=1}^{J} (\hat{Q}_{y_{t+h}|\mathrm{X}}(\tau_j|X_t) - F^{-1}(\tau_j; \mu, \sigma, \nu, \alpha))^2$$

(15)

where $\hat{\nu}_{t+h} \in \mathbb{R}$ (skewness parameter), and $\hat{\alpha}_{t+h} \in \mathbb{R}^{+}$ (kurtosis or tail weight parameter).

The proposed neural network architecture can be trained by minimizing the following "total loss" objective function:

$$\mathrm{TL} = \mathrm{QL} + \mathrm{CL} + \mathrm{LSL}$$

(16)

One could modify the function, by scaling its components with specific weights. These weights can be fixed during training or changed using an arbitrary rule. For example, what was tested in a specific implementation is for the sum of the tilted and crossing loss to have a higher weight in the first 70% of the epochs of the training process and then scale down their weight during the last 30% of the epochs. Such a modification would look like this:

$$\mathrm{TL} = \delta(\mathrm{TL} + \mathrm{CL}) + (1 - \delta)\mathrm{LSL}$$

(17)

where $0 < \delta < 1$ if both terms should effectively enter the final loss function.

One of the most common optimization algorithms used in machine learning, as well as statistics is gradient descent as expressed below:

$$w_i^k = w_i^k - \gamma \frac{\partial TL}{\partial w_i^k}$$

(18)

$$b_i^k = b_i^k - \gamma \frac{\partial TL}{\partial b_i^k}$$

(19)

where γ is the learning rate, which is a fundamental parameter, which can have a significant effect on the outcome of the optimization. $\frac{\partial TL}{\partial w_i^k}$ and $\frac{\partial TL}{\partial b_i^k}$, are the gradients or the derivatives of the loss function with respect to the trainable model parameters – the weights and the biases. The gradients are obtained by applying the chain rule.

The DQPR uses the Sinh-Arcsinh distribution introduced by Jones and Pewsey (2009) to generate conditional distributions. This is a four-parameter distribution that can account for location, scale, skewness, and tail weight and is a generalization of the normal distribution. The reason for choosing it was its convenient implementation in TensorFlow and similar properties to the skewed t-distribution.

The final output dense layers of the DQPR model uses specific parameterization in order to ensure that the scale and tail weight parameters are positive numbers, which is a prerequisite for the implementation of the Sinh-arcsinh distribution. The parameterization is the exponential linear unit plus 1 in order to ensure non-negativity.

$$\begin{cases} x + 1 & x \geq 0 \\ -(e^x - 1) + 1 & x < 0 \end{cases}$$

(20)

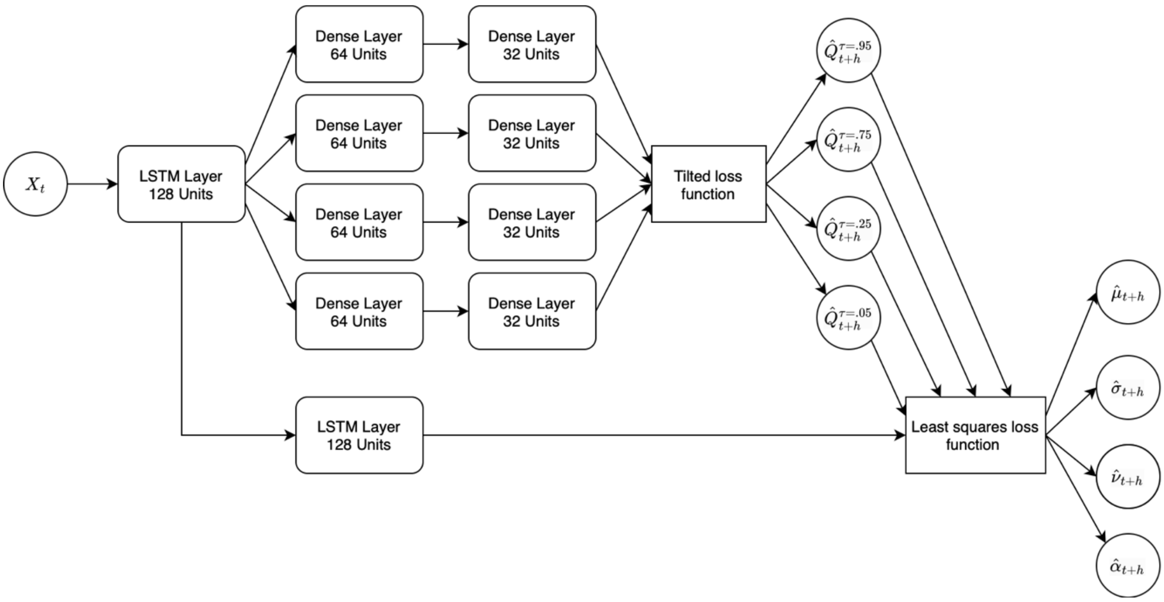
This parametrization is necessary and ensures convergence of the optimization algorithms, as well as the ability to generate a conditional distribution of the Sinh-Arcsinh family.

In order to combine two loss functions into the model, a type of dynamic weighting was implemented. Different versions of the weighting were tested and used for the three case studies. For the first case study on nowcasting the pandemic-related recessions, the model version relied on a high weight (95%) for the tilted absolute value function for 90% of the training epochs (model training duration) and a low weight (5%) for the remainder of the training. Respectively, the weight for the second least squares function, which defines the conditional distribution parameters, is kept low (5%) during 90% of the duration of the training and switched to a high value (95%) for the last 10% of the duration of the training.

For two case studies on forecasting the natural gas price on the Balkan Gas Hub and constructing inflation fan charts for Bulgaria, the model relied on a high weight (90%) for the tilted absolute value function for 60% of the training epochs (model training duration) and low weight (10%) for the remainder of the training. Respectively, the weight for the second least squares function, which defines the conditional distribution parameters, stays low (10%) during 60% of the duration of the training and switches to a high value (90%) for the last 40% of the duration of the training.

By applying this architecture and experimenting with it for the purposes of case studies presented in the next chapter, two extended versions of the architecture were arrived at. Figure 6 depicts the changes compared to the simpler version presented above for the architecture used for the first case study on nowcasting the pandemic-related recession.

Figure 6: Extended Version of the DQPR for Nowcasting the Pandemic-related Recession



Source: Author

Both extended architectures are based on the estimation of four separate quantiles (.05, 0.25, 0.75, and .95 percentiles). For this purpose, branching of the network is used for the estimation of each separate quantile, where each branch contains two dense layers. Eventually, all conditional quantiles are estimated simultaneously using a common cost function, as described above. Two types of activation functions are used in this extended architecture – the sigmoid and the hyperbolic tangent functions. The sigmoid function is defined below:

$$\sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1} = 1 - \sigma(-x)$$

(21)

The dense layers use the hyperbolic tangent activation function as defined below:

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}$$

(22)

This transformation scales the output of the dense layer to lie between 1 and -1. Therefore, the output of a dense layer in the diagram above is defined in the following way:

$$a_i^k = \tanh(a_i^k w_i^k)$$

(23)

where a_i^k is the activation of the *i*-th neuron of the *k*-th layer.

Another feature of the extended architecture is its reliance on recurrent layers. In this case, long short-term memory (LSTM) layers were used after both vanilla recurrent and gated recurrent units (GRU) were tested in different contexts. See Hochreiter, Schmidhuber (1997) on LSTM layers, Cho et al. (2014) on GRU layers, and Bengio et al. (1994) on vanilla recurrent layers and the issues associated with them. Also, a good overview of the current state and implementations of time series models based on LSTM layers can be found in Hewamalage, Bergmeir, and Bandara (2021).

This generalized architecture allows one to use multiple predictor variables and multiple lags for each variable. On the other hand, it allows training a model to predict multiple steps ahead as well, which makes it flexible in terms of usage. The DQPR model was implemented using the TensorFlow library and the ADAM optimizer was used for the model training (Kingma and Ba, 2014). The TensorFlow library uses automatic differentiation and gradient descent through time for recurrent neural networks (Rumelhart, Hinton, Williams, 1986; Williams and Zipser, 1992).

As already discussed, epistemic uncertainty is defined as the limitation in the knowledge of the observer or forecaster regarding the data-generating process of interest. A fundamental assumption of epistemic uncertainty is that it can be reduced through the accumulation of additional knowledge or data. In more practical terms this might include accumulating more observations, finding new predictive variables, or testing out better models.

According to the proposed classification of uncertainty, epistemic uncertainty can be further divided into two other subtypes of uncertainty: model uncertainty which relates to the choice of a family of models and the model architecture, and approximation uncertainty, which relates to the choice of hyperparameters and the uncertainty related to the data sample.

In practice, one could quantify the approximation uncertainty by estimating the uncertainty surrounding the model parameters, and it is easier to explore the chosen hypothesis space via a search technique. However, since the ground truth is unknown, the only way to quantify the model uncertainty is by testing various families of models and model architectures. In the context of an artificial neural network, one could say that changing the architecture more significantly is effectively proposing a new family of models, therefore some hyperparameter optimization in this context can be thought of as changing the hypothesis space.

In the context of probabilistic forecasting with artificial neural networks, the main topic of interest are methods of quantifying parameter uncertainty in neural networks. A concise review of relevant literature includes four methods popular in the deep learning literature.

Dropout (Srivastava et al. 2014), is a technique for regularization, where a random proportion of neurons or activations in a layer of a neural network are set to equal zero, effectively excluding their effect on the final output. Dropout is implemented by the construction of a dropout mask vector, where each element is drawn independently from a Bernoulli distribution with a pre-defined proportion, which is then multiplied with the activations of a given layer. In contrast to Dropout, which is a process usually enabled during model training, Monte Carlo Dropout (Gal and Ghahramani, 2016) or MC Dropout is the same concept, but applied during inference. In this way, different activations are disabled at every inference pass and the model becomes stochastic. According to Gal and Ghahramani (2016) each forward pass of the neural network generates a sample from the Bayesian posterior distribution and by generating a high number of samples (100 is a good starting point proposed by the authors) one could approximate the posterior. The method has its critics, but in practice, it is very easy to implement and computationally extremely cheap. Dropout can be applied to any type of neural network architecture and layer type, including recurrent layers.

Due to its ease of implementation and transparency and computational efficiency, it was decided to apply Monte Carlo Dropout to quantify the parameter uncertainty. Dropout and by extension MC Dropout is readily available to use in TensorFlow for any type of layer and architecture. Additionally, recurrent layers can be the subject of recurrent dropout, which is a strategy for regularizing the recurrent layers. In this way, the neural network is rendered Bayesian or pseudo-Bayesian in the following manner. The model is used to infer the predictive distribution denoted by:

$$p(y|X, \mathcal{D})$$

where y is a target variable, X is a vector of predictor variables and $\mathcal{D}$ is the information set $\mathcal{D} = (y_t, X_t)_{t=1}^{T}$. The goal is to learn the distribution over the model parameters also known as the parametric posterior distribution $p(\Theta|\mathcal{D})$. As previously mentioned, MC dropout works by randomly setting neuron outputs or activations to zero, which regularizes the neural network during inference. Each forward pass of the neural network corresponds to a different sample from the approximate parametric posterior distribution $q(\Theta|\mathcal{D})$:

$$\Theta_k \sim q(\Theta|\mathcal{D})$$

(24)

where Θ_k represents a so-called dropout configuration or a set of parameters resulting from a single forward pass k of the stochastic neural network. Sampling from the approximate parametric posterior distribution $q(\Theta|\mathcal{D})$ allows for Monte Carlo integration of the model's likelihood, which uncovers the predictive distribution as follows:

$$p(y|X) \approx \int_{\Omega} \underbrace{p(y|X,\mathcal{D})}_{likelihood} \underbrace{q(\Theta|\mathcal{D})}_{p.posterior} d\Theta$$

(25)

$$p(y|X) \approx \frac{1}{K}\sum_{k=1}^{K} p(y|X,\Theta_k) \;\; s.t. \Theta_k \sim q(\Theta|\mathcal{D})$$

(26)

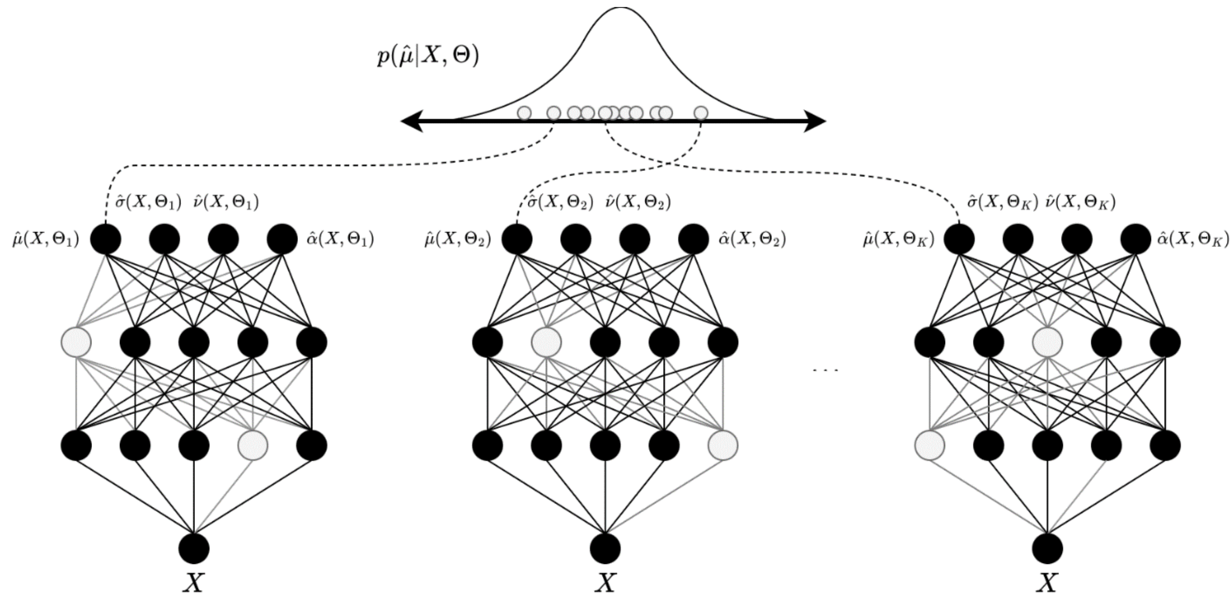
For simplicity, the likelihood is often assumed to be Gaussian distributed, but so far we have used the four-parameter Sinh-Arcsinh family of distribution:

$$p(y|X,\Theta) = F(\hat{\mu}(X,\Theta), \hat{\sigma}(X,\Theta), \hat{\nu}(X,\Theta), \hat{\alpha}(X,\Theta))$$

(27)

The parameters are output by the simulations or the K forward passes of the neural network. Figure 7 illustrates how MC dropout yields a different output by randomly disabling neuron activations (gray circles) and leaving other activations on (black circles) with each forward propagation. Multiple forward passes with different dropout configurations Θ_k yield a predictive distribution over the mean $p(\hat{\mu}|X,\Theta)$.

Figure 7: Bayesian Inference with MC Dropout



Source: Author, Davis et al. (2020)

Two important aspects of the MC dropout implementation are the dropout proportion used and the number of forward propagation samples generated at inference time. For the dropout proportions, Gal and Ghahramani (2016) advise it to be between 10% and 50%. For this implementation, it was decided to use 10%. The authors advise that the number of forward passes can be between 30 and 100. It was decided to use 100, due to the low computational cost of increasing the number of samples.

Kendall and Gal (2017) propose a method for separating aleatoric and epistemic uncertainty for the case of MC Dropout in their study focused on computer vision. Valdenegro-Toro and Mori (2022) review and generalize the method for Flipout and neural network ensembles. Assume a Bayesian neural network outputs a predictive distribution $p(y|X)$, which for simplicity is assumed to be Gaussian. The model outputs two quantities $\mu_k(X)$ and $\sigma_k^2(X)$, which are respectively the mean and variance for every forward pass in a total number of K forward passes. The model parameters are sampled from the approximate parameter posterior $\Theta_k \sim q(\Theta|\mathcal{D})$, which produce different predictions $\mu_k(X)$ and $\sigma_k^2(X)$ for every forward pass, given the same input.

$$p(y|X) \sim \mathcal{N}\left(\mu_*(X), \sigma_*^2(X)\right)$$

(28)

$$\mu_*(X) = \frac{1}{K}\sum_{k=1}^{K} \mu_k(X)$$

(29)

$$\sigma_*^2(X) = \frac{1}{K}\sum_{k=1}^{K}(\sigma_k^2(X) + \mu_k^2(X)) - \mu_*(X)^2$$

(30)

The last equation represents the total predictive uncertainty $\mathbb{V}(y|X)$. The total predictive uncertainty can be separated into two subcomponents using the law of total variance:

$$\sigma_*^2(X) = \frac{1}{K}\sum_{k=1}^{K}\sigma_k^2(X) + \frac{1}{K}\sum_{k=1}^{K}\mu_k^2(X) - \mu_*(X)^2$$

(31)

This can be reformulated as follows:

$$\mathbb{V}(y|X) = \underbrace{\mathbb{E}[\mathbb{V}(y|X,\Theta)]}_{\substack{Aleatoric \\ Uncertainty}} + \underbrace{\mathbb{V}(\mathbb{E}[y|X,\Theta])}_{\substack{Epistemic \\ Uncertainty}}$$

(32)

The last derivation states that across all forward pass samples with different dropout or parameter configurations Θ, the mean of the variances represents the aleatoric uncertainty, whereas the variance of the means corresponds to epistemic uncertainty.

The third chapter of the dissertation is dedicated to three different applications of the DQPR model. In these empirical studies, it is imperative to test and establish the value of the proposed framework in real-world use cases. For this purpose, the performance of the DQPR is measured using relevant indicators and is then compared to benchmarks of varying complexity – from the simplest to the more advanced. For each specific exercise, a different set of benchmarks are used as each of the empirical exercises has a slightly different goal.

The first two relevant indicators used throughout the empirical studies are widely-used measures of point-forecast accuracy. Both of them focus on the central features of the predictive distribution. The root mean squared error (RMSE) measures the accuracy of the conditional mean of the forecast distribution, and the mean absolute error (MAE) evaluates the accuracy of the median. The continuous ranked probability score (CRPS) which is a generalization of the mean absolute error in the context of density forecasts is a measure of both sharpness and calibration (Gneiting et al., 2007) and is considered here as the main performance indicator of interest throughout the empirical studies. The RMSE, MAE, and CRPS are used throughout all three of the empirical studies.

Five separate benchmark models are used throughout the empirical studies. The first one is an AR(1) model with constant variance using the first lag of the target variable and an intercept. The constant variance is calculated using the residuals from the training sample. This benchmark is used across all three of the empirical studies as it is the simplest way to produce a density forecast, albeit with a constant variance.

The second benchmark is a GARCH(1,1) model with an AR(1) mean model, using again only the lag of the target variable. This benchmark is used only in the second empirical study on forecasting the natural gas prices on the Balkan Gas Hub.

The third benchmark is a LSTM-based probabilistic regression implemented using the TensorFlow framework. The model is implemented with the assumption of a Normal distribution, which allows for modeling conditional mean and conditional variance. This benchmark is also used only in the second empirical study on forecasting the natural gas prices on the Balkan Gas Hub.

The fourth benchmark uses the original procedure by Adrian et al. (2019) described earlier in this chapter. It involves estimating a separate quantile regression model for each quantile of interest. In this case, they are four quantiles – .05, .25, .75, .95. Each quantile is estimated using a linear model. After the conditional quantiles are generated, the skewed t-distribution is fitted to the estimates in order to generate the final predictive density. This benchmark is used in the case study of nowcasting the pandemic-related recessions in order to compare the performance as various improvements to this original procedure are made.

The fifth benchmark is an improved version of the original procedure by Adrian et al. (2019), which has two main differences. The first is that the four quantiles are estimated simultaneously, as proposed by Rodrigues and Pereira (2020). The second is an additional term in the cost function, which explicitly addresses the crossing quantiles problem proposed by Bondell et al. (2010). This benchmark is used in the case study of nowcasting the pandemic-related recessions in order to track the improvements to compare the performance between the non-simultaneous quantile estimation and the DQPR.

Lastly, in order to perform a sensitivity analysis of the DQPR model and derive relative variable importance, the local interpretable model-agnostic explanations (or LIME for short) method proposed by Ribeiro et al. (2016). This method learns an interpretable linear model around every prediction, which can provide local explainability. Global explainability is obtained using the local explainability over a given sample and calculating the mean absolute contributions to the predictions for each variable.

Chapter 3: Applications of the Deep Quantile-Based Probabilistic Regression

Nowcasting the 2020 Pandemic Lockdown Recession in Small Open Economies

Recessions are not rare events, according to An et al. (2018). The authors analyzed data on 153 recession episodes across 63 countries between 1992 and 2014 and found that countries on average are in a recession 12% of the time. However, recession events and their timing and magnitude remain hard to predict for both experts and statistical models

(Lewis and Pain, 2014). On the other hand, more impactful events like the great recession that occurred between 2007 and 2009 and the recent recession caused by the coronavirus pandemic lockdown are an even greater challenge for forecasters and decision-makers as they represent realizations of low probability risks (Makridakis et al., 2009; Chen, 2019; Antipova, 2020). While the great recession was caused by a build-up of systemic risk, which in retrospect turned out to be visible in the data (Altunbaz et al., 2017), the coronavirus pandemic lockdown was caused by an unusual and unexpected shock. Therefore, this latest crisis can be considered one of the biggest challenges for the forecasting profession in recent decades.

In an answer to such challenges, the International Monetary Fund (IMF) among other institutions, has been using a framework for quantifying macroeconomic risks to growth, which has become known as growth-at-risk (Prasad et al., 2019). Since models designed to forecast a central feature of the distribution of interest like the mean or the median are unable to capture asymmetries between upside and downside risks, the assessment of the uncertainty surrounding point forecasts becomes necessary (Clemens, 2004). One way to address this necessity, which is supported by a growing body of research recently and is at the core of the IMF growth-at-risk framework, is to model empirically the future growth distribution on the basis of current macroeconomic and macro-financial conditions. While different models have been used to achieve this task including Bayesian VAR models (Carriero, 2020), stochastic volatility models (Iseringhausen, 2021), and GARCH models (Brownlees, Souza, 2021), this paper focuses on methods based on quantile regression.

In an influential paper, Adrian et al. (2019) use a two-step procedure of constructing conditional quantiles using a quantile regression model and consequently fit a probability distribution to the estimated quantiles. The authors studied the conditional US growth distribution with an emphasis on financial conditions. They identified several stylized facts about the conditional distribution of growth for the USA, among which a strong negative correlation between the conditional mean and variance and a significant relationship between current financial conditions and future shifts in the lower tail of the conditional distribution. The same conclusion was confirmed by De Santis and Van der Veken (2020), who performed a similar exercise including data from the beginning of 2020 and a separate dataset covering the Spanish flu pandemic period across a number of countries. Figuerez and Jarociński (2020), confirm the same stylized facts identified by Adrian et al. (2019) for the Euro Area.

The case study applies a novel approach inspired by the semi-parametric two-step procedure used by Adrian et al. (2019) and De Santis and Van der Veken (2020). The proposed approach consists of a one-step model, which is based on artificial neural networks and outputs the parameters of the conditional growth distribution. The model depends internally on the estimation of conditional quantiles and for this purpose, it is based on two separate loss functions, which are dynamically weighted. The improvements proposed here lie in four separate areas:

1. A simultaneous generation of quantiles, as proposed by Rodrigues and Pereira (2020), in order to alleviate the quantile crossing problem;
2. The introduction of quantile crossing loss to the tilted loss function, which further prevents quantile crossing as proposed by Bondell et al. (2010);
3. Using artificial neural network architecture based on long short-term memory (LSTM) layers (Hochreiter, Schmidhuber, 1997) to model non-linear relationships between the predictors and the target variable and better capture the recession related to the pandemic lockdown compared to a linear model;
4. Combining the two steps of the procedure into a single model, which is being optimized by minimizing two loss functions simultaneously – the tilted absolute loss function used for estimating the conditional quantiles and a least squares loss for evaluating the final conditional distribution parameters.

This combination of improvements is called a deep growth-at-risk model or more technically deep quantile-based probabilistic regression (DQPR in short) for the purposes of this case study. Initially, the focus of the study was on the macroeconomic developments in Bulgaria, but after preliminary results were generated it was decided to test the proposed approach on three other small open European economies, relatively similar in terms of size and structure of the economy. Therefore, the proposed procedure was tested on data for Bulgaria, Estonia, Lithuania, and Romania, covering the coronavirus pandemic lockdown period and the recession related to it, and achieved a better out-of-sample performance across four of them compared to three separate benchmarks.

The case study is structured as follows. The next section covers the data used in this study, while the third section summarizes the empirical results. The last section contains a discussion of the results and the conclusions of the study.

For this analysis, the target variable of interest is the quarterly growth rate of the seasonally and calendar-adjusted chain-linked volumes of GDP. The available final release of the GDP data is used as of the writing of this text. Apart from the lags of the target variable, a list of leading indicators of financial conditions and economic activity was compiled in order to be used as candidate predictors. The choice of leading indicators was following an approach similar to Adrian et al. (2019), De Santis and Van der Veken (2020), Figuerez and Jarociński (2020), and Prasad et al. (2019). It was imperative that they are available for a longer time frame and an initial choice for a starting year of the samples was the year 2000 as this ensured a long enough training sample and the opportunity to put aside a test sample. Currently, there are a lot of interesting leading indicators which can be used for similar macroeconomic forecasting tasks, but their main disadvantage is the lack of accumulated historical data. Moreover, it was decided to include only indicators, which are available for a specific quarter by the end of the same quarter, in order to be able to use the current values of the predictors in time reference to the GDP growth values, which are released later on. Therefore, short-term indicators which are released with a significant delay were not included in the modeling data set, despite their relevance, because they have limited use in the nowcasting of GDP growth.

All variables were normalized with a mean of 0 and a standard deviation of 1 prior to use in the model. The total sample covers the period 2000Q1 to 2021Q4 and was divided into a training and a testing sample. The first 64 quarters were used for model training (2000Q1 to 2016Q2) and the last 22 quarters were used for validating the model performance (2016Q3 to 2021Q4). A rolling window approach was followed to construct the training and test samples. The final specifications are depicted in a stylized way in Table 1.

Table 1: Final Model Specifications

Country	Specifications
Bulgaria	$GDP_t = f(GDP_{t-1}, SENTIMENT_t, SOFIX_t, US\ BOND\ YIELD_t)$
Estonia	$GDP_t = f(GDP_{t-1}, SENTIMENT_t, OMXTGI_t)$
Lithuania	$GDP_t = f(GDP_{t-1}, SENTIMENT_t, LT\ BOND\ YIELD_t)$
Romania	$GDP_t = f(GDP_{t-1}, SENTIMENT_t, US\ BOND\ YIELD_t)$

Source: Author

In the listed specifications, the left-hand side describes the target variable, and the right-hand side the set of predictors. For Bulgaria and Estonia, the inclusion of the domestic stock price indices leads to optimal performance. In the case of Bulgaria, the inclusion of the US 10-year government bond yield carried additional predictive power. Similarly, for Lithuania, using the Lithuanian 10-year government bond yield results in the best model performance. For Romania, the use of the US 10-year bond yield leads to better performance compared to using domestic indicators.

The DQPR model was tested against three separate benchmarks. The first one is an AR1 model, which uses the conditional mean generated by the model and a constant conditional variance calculated over the training sample as parameters of the Normal distribution. The two other benchmarks are based on the original two-step approach. Both benchmarks use linear quantile regression, but the first one estimates the conditional non-simultaneously, while for the second the conditional quantiles are generated simultaneously and the use of the crossing loss is applied, similarly to the DQPR model. The skewed t-distribution is used for the generation of the conditional distribution of these two benchmarks. The performance of the models is based on the pseudo-out-of-sample performance over the test sample covering the pandemic lockdown recession occurring in 2020Q1 and/or 2020Q2. The main indicators which were used to measure and compare the performance are RMSE and MAE for the point-forecast accuracy and the CRPS as the main indicator of the quality of the density forecasts. For the DQPR, the median forecast is used in both the calculation of the RMSE and MAE.

The DQPR model achieves superior accuracy compared to the benchmark, both during the negative growth period which every country experiences between 2020Q1 and 2020Q2, and the total testing sample. The model also recognizes the downside risks reliably, given the increased spread of the distribution during periods of negative growth.

The benchmarks fail at recognizing both the timing of the downside as well as the downside risks across all countries. However, it seems to perform satisfactorily during upturn periods. All models perform poorly when it comes to forecasting the recovery after the initial slump and recognizing upside risks. All models perform well during non-recession periods, but the DQPR model is better at modeling the recession caused by the unexpected shock of the global pandemic lockdown. By utilizing a high number of LSTM units, the DQPR model manages to recognize a highly non-linear relationship between the predictors and the target. During part of the initial experimentation, it was observed that reducing the number of LSTM units reduces its performance during the recession periods and makes its performance more similar to the benchmarks. More detailed performance results can be seen in Table 2.

Table 2: Performance Evaluation

Country	Model	RMSE	MAE	CRPS
Bulgaria	AR1	2.3857	1.2188	1.2146
	Linear Non-Simultaneous Quantile Estimation	2.0498	1.7895	1.6312
	Linear Simultaneous Quantile Estimation	2.6840	1.3597	1.7928
	DQPR	**1.4828**	**0.8782**	**1.0693**
Estonia	AR1	1.9221	1.1094	1.1026
	Linear Non-Simultaneous Quantile Estimation	1.6027	1.0163	1.3550
	Linear Simultaneous Quantile Estimation	2.3538	1.5348	1.6579
	DQPR	**1.3445**	**0.8066**	**1.0182**
Lithuania	AR1	2.2336	1.0367	1.0311
	Linear Non-Simultaneous Quantile Estimation	1.5220	0.7641	0.9111
	Linear Simultaneous Quantile Estimation	2.0726	0.9709	0.9086
	DQPR	**1.1068**	**0.5047**	**0.7649**
Romania	AR1	2.9932	1.4178	1.4122
	Linear Non-Simultaneous Quantile Estimation	1.8490	1.1164	1.3224
	Linear Simultaneous Quantile Estimation	2.6532	1.3937	1.6711
	DQPR	**1.7456**	**1.0263**	**1.0648**

Source: Author

The DQPR consistently leads for RMSE, MAE, and CRPS across all countries. The non-simultaneous quantile estimation benchmark achieves better performance compared to the other two benchmarks. However, the model suffers significantly from the quantile crossing problem, which is slightly alleviated when specific distribution parameters are fit to the estimated quantiles. The improvement introduced by the DQPR compared to the best benchmark across performance indicators is between 5.6% and 33.9%. The summary of the percent improvement per country and per indicator can be found below.

Table 3: Performance Improvement of the DQPR Against the Best Benchmark

	RMSE	MAE	CRPS
Bulgaria	27.7%	27.9%	12.0%
Estonia	16.1%	20.6%	7.7%
Lithuania	27.3%	33.9%	15.8%
Romania	5.6%	8.1%	19.5%

Source: Author

The DQPR model produces conditional distributions across all countries, which confirm some of the findings of Adrian et al. (2019). Both symmetric conditional distributions during expansions and negative skewness during periods of recession can be observed. Additionally, a negative correlation between the conditional mean and variance of the growth distribution is evident as well. These results were confirmed for all four of the countries in the sample, both in-sample (on the training set) as well as out-of-sample (on the test set).

In order to perform a sensitivity analysis and extract the variable importance from each model, the average absolute marginal contributions of each variable to the model output have been calculated using the LIME method across the test sample as well as for 2020Q2 when the recession was observed across all countries. The explainability results are presented in terms of relative importance in Table 4.

Table 4: Relative Variable Importance

Country / Indicator	Test Sample	2020Q2
Bulgaria		
GDP	71.28%	54.28%
Sentiment Indicator	18.55%	38.01%
SOFIX	5.25%	3.24%
US Bond Yield	4.92%	4.46%
Estonia		
GDP	68.19%	75.30%
Sentiment Indicator	20.87%	24.43%
OMXTGI	10.94%	0.26%
Lithuania		
GDP	46.06%	4.08%
LT Bond Yield	31.20%	17.02%
Sentiment Indicator	22.74%	78.90%
Romania		
US Bond Yield	36.79%	23.33%

GDP	36.18%	10.44%
Sentiment Indicator	27.03%	66.23%

Source: Author

It is observed that for all countries except Romania, the GDP has the highest average importance over the test sample. For Romania, the US Bond Yield is the most important variable on average. For Bulgaria and Estonia, the sentiment indicator is the second most important variable, while for Lithuania it is the Lithuanian Bond Yield, and for Romania the GDP. When 2020Q2 is considered, a significant increase in the importance of the sentiment indicator is observed across all countries, and for Lithuania and Romania it has the highest relative importance for the given period. This points out the fact that the sentiment indicator has a greater role in faithfully predicting economic recessions compared to the proxy indicators of financial stress, which is a surprising finding.

The COVID-19 pandemic and the recessions many countries experienced due to implemented lockdowns posed an unprecedented challenge to decision-makers and forecasters. Both private enterprises and government institutions had to adapt to this shock quickly and implement policies to tackle the consequences, based on limited foresight. While it is virtually impossible to anticipate such an event as the coronavirus pandemic and its consequences ahead of time, one could forecast or nowcast its effects on the economy through leading indicators, which could help the decision-making process.

The current study demonstrates that a parsimonious model using country-specific sentiment indicators as well as country-specific and global financial variables can successfully nowcast recessions caused by unexpected shocks like the coronavirus pandemic. The comparative performance of the artificial neural network DQPR model proves that it is a useful tool in modeling macroeconomic risks related to the 2020 coronavirus pandemic lockdown in four small open economies in Europe. Its ability to model highly non-linear relationships makes it superior to a set of linear benchmarks in this case.

For Bulgaria and Lithuania, the DQPR model manages to predict very accurately the negative growth of GDP in 2020Q2, when the strongest economic effects of the lockdowns were felt. In the case of Estonia, the DQPR model does not accurately predict the start of the recession in 2020Q1, but manages to predict very accurately the negative growth in 2020Q2. However, it is not clear whether the growth dynamics in this quarter are not a result of seasonal adjustment. For Romania, the DQPR model fails to predict the full extent of the lockdown recession in 2020Q2, but still outperforms significantly the linear benchmarks in terms of MAE and CRPS. Apart from its disadvantage in much lower prediction accuracy with respect to predicting the pandemic crisis, the linear benchmarks achieve satisfactory performance in nowcasting growth during upturn periods. Overall, the DQPR achieves a performance improvement against the best benchmark of up to 33.9%.

A disadvantage shared by both the proposed DQPR model as well as the linear benchmarks is their limited ability to predict the upturn after the initial decline in economic growth. This result is observed across all countries and at first glance the problem is with the so-called shape of the 2020 recession, which in all countries of interest seems to have a V-shape. The models are trained on the recessions caused by the global financial crisis, which had either a U or a W-shape for the countries of interest, which might be why the models fail to anticipate a quick and strong recovery after only a quarter or two of decline in growth. The inclusion of indicators of delayed consumption might be a way to account for the strong recovery, as the recession was not the result of a decline in income, but the inability to spend due to the policy of lockdown.

With respect to the indicators used across the four countries of interest, it is evident that both country-specific and global factors of financial stress carry predictive power with respect to economic growth and specifically in the task of predicting the pandemic lockdown recession. For Bulgaria and Estonia, it was shown that the use of domestic stock price indices' close values leads to optimal results. In the case of Bulgaria, the inclusion of the US 10-year government bond yield carried additional predictive power. For Lithuania, the inclusion of the Lithuanian 10-year government bond yield resulted in the best-performing model specification. For Romania, the US 10-year bond yield carried more predictive power with respect to predicting the pandemic recession, compared to the country-specific financial indicators. Additionally, across all four countries, it is demonstrated that a parsimonious model containing few indicators yields optimal performance.

Using the LIME approach, relative variable importance was estimated for the DQPR model across the four countries. When the whole test sample is considered, the GDP has the highest average importance over the test sample, except for Romania. For Romania, the US Bond Yield is the most important variable on average. For 2020Q2 specifically a significant increase in the importance of the sentiment indicator is observed across all countries and for Lithuania and Romania, it has the highest relative importance for the given period. This result shows the significance of using the country-specific sentiment indicator in forecasting the pandemic-related recession.

The DQPR model combines a couple of recent improvements proposed by researchers working on quantile regression models, which allows it to mitigate known problems like crossing quantiles. The first improvement is the simultaneous estimation of quantiles, which allows one to estimate an arbitrary number of quantiles within one estimation procedure and using a single loss function. This both speeds up the process of generating the conditional quantiles, but also is shown to alleviate the crossing problem. The second improvement is the explicit inclusion of a crossing loss term within the loss function, which additionally mitigates the issue. Moreover, combining the two steps of the original procedure into a single model creates one internally consistent procedure without sacrificing its flexibility. Working in the context of an artificial neural network allows one to construct a custom model with two loss functions and combine the two steps of the original estimation procedure into a one-step procedure.

Additionally, the current analysis confirms that there is a negative correlation between the conditional mean and variance of the distribution of growth as well as symmetric conditional distributions during expansions and negative skewness during periods of recession for the four economies analyzed in this study, in to the stylized facts Adrian et al. (2019) identified for the US, and Figuerez and Jarociński (2020) confirmed for the Euro Area as well.

Probabilistic Forecasting of Natural Gas Prices on the Balkan Gas Hub Using Deep Learning

Natural gas is one of the major energy commodities of global importance. Its uses include heating, electricity generation, fuel for transportation as well as being an input good in the production of plastics, fertilizers, and fabrics. The reliance on natural gas in Europe is significant. According to Eurostat, the share of natural gas in the overall energy mix of the EU was 23.7% in 2020 and natural gas accounted for 31.7% of the final energy consumption of households in the same year. In terms of supply, a large portion of the natural gas in the EU is imported with the imports dependency being 83.6% for 2020, according to Eurostat public database. Although the EU has made significant steps to diversify the supply, it still imported around 40% of its total gas consumption in 2021 from Russia, according to the European Commission.[2]

Price formation of natural gas in Europe has undergone significant changes in the past two decades. Before 2000, natural gas prices in the EU were almost entirely determined by long-term contracts indexed to the price of crude oil. Gas prices followed crude oil price trends, which provided a somewhat stable reference price. However, this system did not reflect supply-demand dynamics, and the market participants in the EU were unable to take advantage of periods of lower-cost supply. Over the last decade, gas prices in the EU have gradually moved away from oil indexation toward "gas-on-gas" competition, where prices reflect multiple sellers and buyers of natural gas on spot markets. Zhang et al. (2018) claim that the hub-based pricing mechanism is associated with less extreme volatility because it is not as vulnerable to speculation. The Title Transfer Facility (TTF) in the Netherlands developed as the most liquid hub and relevant price benchmark in the EU. Eventually, similar hubs were created in other countries like the Balkan Gas Hub in Bulgaria, which was launched in the beginning of 2020. Due to a significant degree of market integration in the European gas markets (Broadstock et al., 2020) prices follow similar trends across the various gas hubs (Berrisch and Ziel, 2022).

In terms of price dynamics, the natural gas market has been historically volatile due to supply and demand imbalances, costly infrastructure, storage levels, weather, policy changes, and political events (Siddiqui, 2019). The reliance on a single large supplier like Russia has proven to carry a significant risk as well (Weisser, 2007). There have been supply disruptions historically (in 2006 and 2009), but with the onset of the Russian

[2] European Commission: https://commission.europa.eu/news/focus-reducing-eus-dependence-imported-fossil-fuels-2022-04-20_en

invasion of Ukraine in February 2022, a series of political decisions by the Russian government and changes to service provision by the Russian supplier Gazprom led to unprecedented episodes of volatility and spikes in prices, mostly related to limited and uncertain supply. Amid colder weather, concerns about the launch of Nord Stream 2, and growing tensions between Russia and Ukraine, the supply of gas on the Yamal line fell at the beginning of February 2022, which led to the first in a series of unprecedented price surges on the European gas markets (European Commission, 2022). After the start of the Russian invasion of Ukraine on 24 February 2022, the EU imposed heavy sanctions on Russia and Gazprom required payments in Rubles on their long-term contracts which led to another temporary, but significant spike in prices. Eventually, a series of supply disruptions to various countries in Europe led to similar spikes and volatility on the natural gas markets (European Commission, 2022).

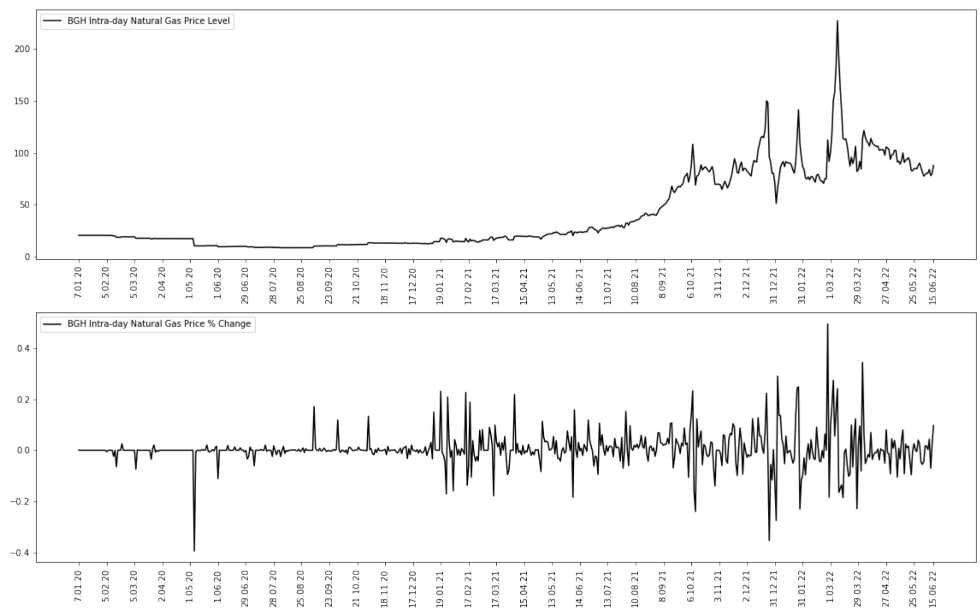
The current energy crisis in Europe leads to unprecedented levels of uncertainty, where short-term planning is both important and difficult. In the context of the European liberalized markets, the ability to make accurate predictions of the natural gas price is important for various market participants and drives their decisions about the quantities and timing of purchases on the spot markets. According to Siddiqui (2019), the increasing reliance on futures as a hedging tool against price-driven risks is attributed to the inability of market participants to forecast the spot prices accurately. Multiple recent studies have addressed the issue of natural gas prices forecasting in the past (Siddiqui, 2019; Su et al. 2019a, Su et al. 2019b; Jianliang et al., 2020), but they were exclusively focused on point forecasting. There are only a couple of studies, which focus on the issue of forecasting the natural gas price in the context of probability density forecasting (Berrish and Ziel, 2021; Ding et al., 2022). Probabilistic forecasts carry information about the whole predicted distribution instead of only a central value, and in this sense, they are an assessment of future risks. In the context of the natural gas price in Europe, where periods of higher volatility are observed, probabilistic forecasts would be more informative and better suited compared to point forecasts, especially when heteroskedasticity is observed and changes in the probability distribution of the underlying data generation process are possible.

The current study is focused on day-ahead probabilistic forecasting of the intraday natural gas prices on the Balkan Gas Hub (BGH), with the aim of modeling and predicting risk, and proposes the use of a deep quantile-based probabilistic regression model to achieve this goal. This neural network architecture was previously used for nowcasting the economic crisis caused by the global pandemic lockdown across several countries in Eastern Europe (Yanchev, 2022) and was adapted to the forecasting task at hand. The study uses a historical sample of daily prices starting from the launch of the BGH in 2020 until December 2022. Additional explanatory variables have been used like the TTF 1-month futures, the URALS crude oil spot prices, the European Emissions Allowance (EUA) prices, as well as data on Gazprom import volumes through several pipelines. The proposed model out-of-sample performance is tested against two statistical benchmarks and one other novel deep learning method. The proposed model achieves the best out-of-sample performance in terms of several indicators compared to the benchmarks.

The case study is structured as follows. The following section covers the data used in this study, while the third section summarizes the empirical results. The last section contains a discussion of the results and the conclusions of the study.

In the current study, the target indicator of interest is the daily percentage change of the Balkan Gas Hub intraday natural gas price. The period of interest is between the beginning of 2020 when the Balkan Gas Hub was first launched and December 2022. From Figure 8 one could see how the prices on the gas hub evolved over time both in level and in percent change. Relatively, the period until the end of 2020 is marked by low volatility and a barely noticeable trend. After the beginning of 2021 there is a noticeable increase in volatility and at the end of 2021 there is an abrupt shift to a defined upward trend and clusters of stronger volatility. From September until December 2022, factors like high levels of gas storage across the EU and newly finished infrastructure across several countries including Bulgaria led to prices gradually coming down.

Figure 8: Dynamics of the Balkan Gas Hub Intraday Price in Level and in Percent Change



Source: Author, Balkan Gas Hub

Apart from the first lag of the target variable, several leading indicators were short-listed due to their potential predictive power in forecasting the target variable. The leading indicators were picked following an approach similar to Berrish and Ziel (2021) and Ding et al. (2022). A list of the selected leading indicators can be found in Table 5.

Table 5: Selected Leading Indicators

Indicator	Description and Source
Russian Natural Gas Pipeline Import Volume in the EU	Daily frequency, Source: ENTSOG Transparency Platform
TTF 1-Month Futures	Daily frequency, Source: Investing.com
European Emissions Allowance Yearly Futures	Daily frequency, Source: Investing.com
URALS Crude Oil Spot Prices	Daily frequency, Source: Investing.com

Source: Author

All indicators used were transformed to daily percent change. The first lag with respect to the target variable was used for all variables. All variables were normalized with a mean of 0 and standard deviation of 1 prior to use in the model. The total sample covers the period 07.01.2020 to 30.12.2022 and was divided into a training and a testing sample. The first 668 days were used for model training and the last 75 quarters were used for validating the model performance in a pseudo-out-of-sample exercise. A rolling window approach was followed to construct the training and test samples. In the initial months of the launch of the Balkan Gas Hub, there were multiple days without any trades for which no intraday prices were realized. In order to remove the missing values, an imputation was performed where each missing value was replaced with the last available value.

The performance comparison between the selected models is evaluated in a pseudo-out-of-sample exercise. The test sample covers the period between 18.03.2022 and 30.12.2022, which is marked by clusters of volatility and significant spikes in the natural gas price. Several indicators were used to measure and compare the performance of the models, both in terms of their point-forecast accuracy and their density forecast accuracy. Table 6 presents the RMSE, MAE, and CRPS for all models.

Table 6: Out-of-sample Performance Evaluation

Model	RMSE	MAE	CRPS
AR(1) with Constant Variance	0.1177	0.0787	0.0621
AR(1) - GARCH(1,1)	0.1180	0.0785	0.0605
LSTM-based Probabilistic Regression	0.1091	0.0734	0.0570
Deep Quantile-based Probabilistic Regression	**0.1079**	**0.0712**	**0.0566**

Source: Author

The DQPR leads across all indicators of performance, with the LSTM-based probabilistic regression being second, again, across all indicators. In terms of CRPS, the two models perform almost identically, and therefore, both models are well-suited for the task at hand. The AR(1) with constant variance benchmark outperforms the AR(1)-GARCH(1,1) benchmark in terms of all indicators, and the latter has the worst overall performance. The improvement in performance introduced by the DQPR against the best benchmark varies between 0.7% for the CRPS and 3.0% for the MAE.

Table 7: Performance Improvement of the DQPR Against the Best Benchmark

	RMSE	MAE	CRPS
Improvement in %	1.1%	3.0%	0.7%

Source: Author

In order to perform a sensitivity analysis and extract the variable importance from the DQPR model, the average absolute marginal contributions of each variable to the model output have been calculated using the LIME method across the test sample. The explainability results are presented in terms of relative importance in Table 8.

Table 8: Relative Variable Importance

Indicator	Variable Importance
TTF 1-Month Futures	60.07%
BGH Intraday Natural Gas Price	19.84%
European Emission Allowances	14.32%
URALS Crude Oil Price	3.42%
Russian Natural Gas Pipeline Import Volumes in the EU	2.35%

Source: Author

The TTF 1-month futures have the highest variable importance across the test sample, followed by the BGH intraday price and the European emission allowances. The URALS crude oil price and the Russian pipeline imports have much more limited variable importance. These results prove that using data from the leading European TTF hub has a high value-added for forecasting natural gas prices on a regional hub like the BGH.

The year 2022 posed a significant challenge for Europe with the return of armed conflict on its territory. In terms of the economic impact, the biggest effect was felt on the energy markets and more specifically the natural gas market. A series of political decisions by the Russian government and changes to the service provisions by the Russian supplier Gazprom followed the first months of the invasion of Ukraine. The effect of these actions and the overall escalation of the conflict led to unprecedented price spikes and volatility on the European natural gas market. As of the writing of this article, the markets have reached a state of relative calm, but only after another series of events, which took place in Q3 of 2022 like the reduction or termination of Gazprom's supply to more European countries (including France and Latvia), and the explosions of the Nord Stream 1 and Nord Stream 2 pipelines, which rendered them completely damaged and inoperative.

The increased risk on the gas market affected all market participants – businesses and governments alike. The current study attempts to provide a solution to operating in a market with elevated risk via one-day-ahead probabilistic forecasting of the gas intraday prices on the Balkan Gas Hub. The study proposes an approach to generating probabilistic density forecasts using a neural network architecture called deep quantile-based probabilistic regression (DQPR). The method includes the estimation of four conditional quantiles and the consequent estimation of the parameters of the Sinh-Archsinh distribution, which is a four-parameter distribution that can account for location, scale, skewness, and tail weight and is a generalization of the normal distribution. The proposed approach is compared to two statistical benchmarks and one other novel deep learning approach. The two statistical benchmarks are an AR(1) model with constant variance, estimated over the training sample, and an AR(1)-GARCH(1,1) model which models both the conditional mean and variance. For both benchmarks, an assumption of normality and symmetry is made. The deep learning approach is an LSTM-based probabilistic regression, which allows for the direct estimation of the parameters of a normal distribution.

Several leading indicators were used with the proposed DQPR model and the LSTM-based probabilistic regression. These indicators include the TTF 1-month futures, the Russian pipeline gas import volumes to the EU, the URALS spot price, and the European Emission Allowances yearly futures. The performance comparison shows that the proposed DQPR model outperforms the rest of the models across RMSE, MAE, and CRPS and achieves a performance improvement of up to 3.0% against the LSTM-based benchmark. These results establish the DQPR model as a reliable method for forecasting natural gas prices.

Looking at the relative variable importance for the DQPR, the most important indicator is the TTF 1-month futures, followed by the BGH intraday prices and the European emissions allowances. These results prove that using data from the leading European TTF hub has a high value-added for forecasting natural gas prices on a regional hub like the BGH.

The current study has several novel contributions. It is the first study to focus on forecasting the natural gas prices on the Balkan Gas Hub. It employs a novel and useful strategy in forecasting prices on one of the multiple gas hubs in Europe, by using the leading TTF gas exchange as a predictor. The study reworks the novel DQPR model from its initial application to modeling recession risks to the task of generating density forecasts of natural gas prices.

Future work on the subject can focus on extending the model specifications with additional indicators of supply and demand, longer-term forecasting, an application of data from other gas hubs of global importance, and experimenting with a wider range of machine learning algorithms.

Advanced Applications: Measuring Aleatoric and Epistemic Uncertainty in Forecasting Consumer Inflation in Bulgaria

Inflation forecasting has been of great interest to economists for the past century for both theoretical and practical purposes. Inflation is one of the very important and perhaps fundamental markers of the health of the economy, and theories of its relationship with unemployment called the Phillip's curve have been historically of great interest to theoretical economists (see Fuhrer et al., 2009 or Kasabov et al., 2017 for a summary). On the other hand, measuring and forecasting inflation has been central in determining monetary policy and inflation targeting, in which central banks like both the ECB and US Federal Reserve have been engaged. Inflation has many facets like consumer prices, producer prices, wages, import and export prices, housing prices, and inflation expectations, which are important for the decision-making process of firms, consumers, and governments.

After the global financial crisis of 2008 and 2009, and the recession which followed and spread throughout Europe and most of the world, inflation in Europe remained low and in some countries even negative. Bulgaria also experienced a prolonged period of negative annual HICP inflation between August 2013 and December 2016, according to data from the NSI and Eurostat. According to the ECB (Ciccarelli and Osbat, 2017), the drivers of the low inflation in Europe after 2012 included both internal and external factors for the various countries but also found clear patterns across countries. Low or decreasing external prices of commodities were probably the major driver of low or negative inflation. On the other hand, following the recession of 2009-2010, Europe fell into the sovereign debt crisis, which affected some of the southern members of the EU like Spain, Italy, and Greece, which affected aggregate demand in these countries, but also the whole EU. Therefore, subdued aggregate demand after the Great Recession led to this long period of low or even negative inflation for some of the EU member states.

In contrast, in the period 2020-2023, inflation skyrocketed across the world due to several factors. The coronavirus SARS-COV-2 pandemic caused unprecedented disruptions in the global economy in terms of both supply and demand. This was caused by pandemic lockdowns instituted across many countries simultaneously, which caused sharp local recessions, but also disrupted dramatically global value chains (Barlow et al., 2022). This disruption was so strong that it led to global imbalances which are yet to be resolved like the global semiconductor shortage (Attinasi et al., 2021), shortage and price increases of metals and food products[3], and shipping container shortage (Toygar et al., 2022).

On top of the global disruptions due to the coronavirus pandemic, in February 2022, the Russian invasion of Ukraine set the stage for an unprecedented crisis in the European energy markets and a rise in global food prices. At the start of the war conflict, Russia and the state-owned company Gazprom gradually discontinued natural gas exports to

[3] Based on Food and Metals price indices collected by the IMF.

multiple European countries, as a response to massive financial and trade sanctions imposed by the EU and the USA. These actions affected directly natural gas, electricity, and crude oil markets in Europe and had spillovers across the world (Adolfsen et al., 2022). Additionally, the war in Ukraine threatened the global wheat food security as the two sides of the conflict are two of the largest global producers (Nasir et al., 2022). Predictably, these developments were translated into very large increases in both producer and consumer prices in 2022 across Europe and the World. (sentence on producer and consumer price increases in percentages). In this volatile environment, the task of forecasting inflation has become increasingly harder, but also increasingly necessary for all economic agents.

It was previously discussed that probabilistic forecasting is superior to point forecasts, especially when extraordinary volatility or asymmetries of the risks are involved. Many central banks have historically adopted publishing fan charts, which represent the uncertainty surrounding inflation or growth forecasts. Central banks like the FED, the ECB, the Bulgarian National Bank, and many others publish inflation and growth fan charts regularly. The fan chart is usually represented as a central forecast and multiple bands of confidence similar to prediction intervals around this central forecast. The further out away from the central forecast, these bands are, the lower the probability of the realization. A characteristic of these fan charts is they expand or widen as the forecasting horizon increases, due to the compounding of the uncertainty in non-stationary processes. In other words, the further into the future one tries to predict, the larger the uncertainty surrounding the prediction.

This case study aims to demonstrate an approach for constructing an inflation fan chart using a Bayesian version of the DQPR model, which accounts for both aleatoric and epistemic uncertainty. This involves a quantification of both the uncertainty inherent to the data-generating process or its measurement (aleatoric) and the uncertainty related to the model parameters and model Selection and its approximation properties (epistemic). This is achieved using an approximation of Bayesian inference in the context of artificial neural networks using the Monte Carlo dropout (MC dropout) technique. The structure of the study is the following: the next section is a detailed description of the used data and the last two sections are a summary of the results and a discussion outlining the conclusions.

Various sources were used to collect the data for this exercise. The overall harmonized index of consumer prices for Bulgaria downloaded from Eurostat is used as a target variable for this exercise. Monthly data on unemployment based on Labor Force Survey adjusted series are also downloaded from Eurostat. The data on external commodity prices is taken from the IMF data portal. The following commodity prices were selected for inclusion as explanatory variables:

- Food price index (2016=100) including cereal, vegetable oils, meat, seafood, sugar, and other foods

- All metals index (2016=100) includes aluminum, cobalt, copper, iron ore, lead, molybdenum, nickel, tin, uranium, and zinc
- Crude oil price index (2016=100), simple average of three spot prices – Brent, West Texas, and Dubai Fateh
- TTF day-ahead natural gas price in US dollars

Two additional endogenous factors, which are strongly related to inflation, were considered as well: the output gap and inflation expectations. The output gap is calculated as the deviation from the potential output, and the potential output is calculated using an HP filter (Giorno et al., 1995). The inflation expectations are represented by lagged values of the inflation target variable, as well the consumer sentiment indicator for the perception of inflation. The consumer sentiment measure used was the consumer perception of inflation over the next 12 months.

The general choice of explanatory variables was inspired by the study on the Phillips curve in Bulgaria by Kasabov et al. (2017). The authors point out that apart from factors like the output gap and inflation expectations, external supply-side shocks like international oil prices, other import prices, and the effective exchange rate are important determinants of inflation. Moreover, during the most recent period after 2020, during which Bulgaria recorded steadily accelerating inflation, there were clear and strong supply-side shocks in terms of energy, food, and metal prices.

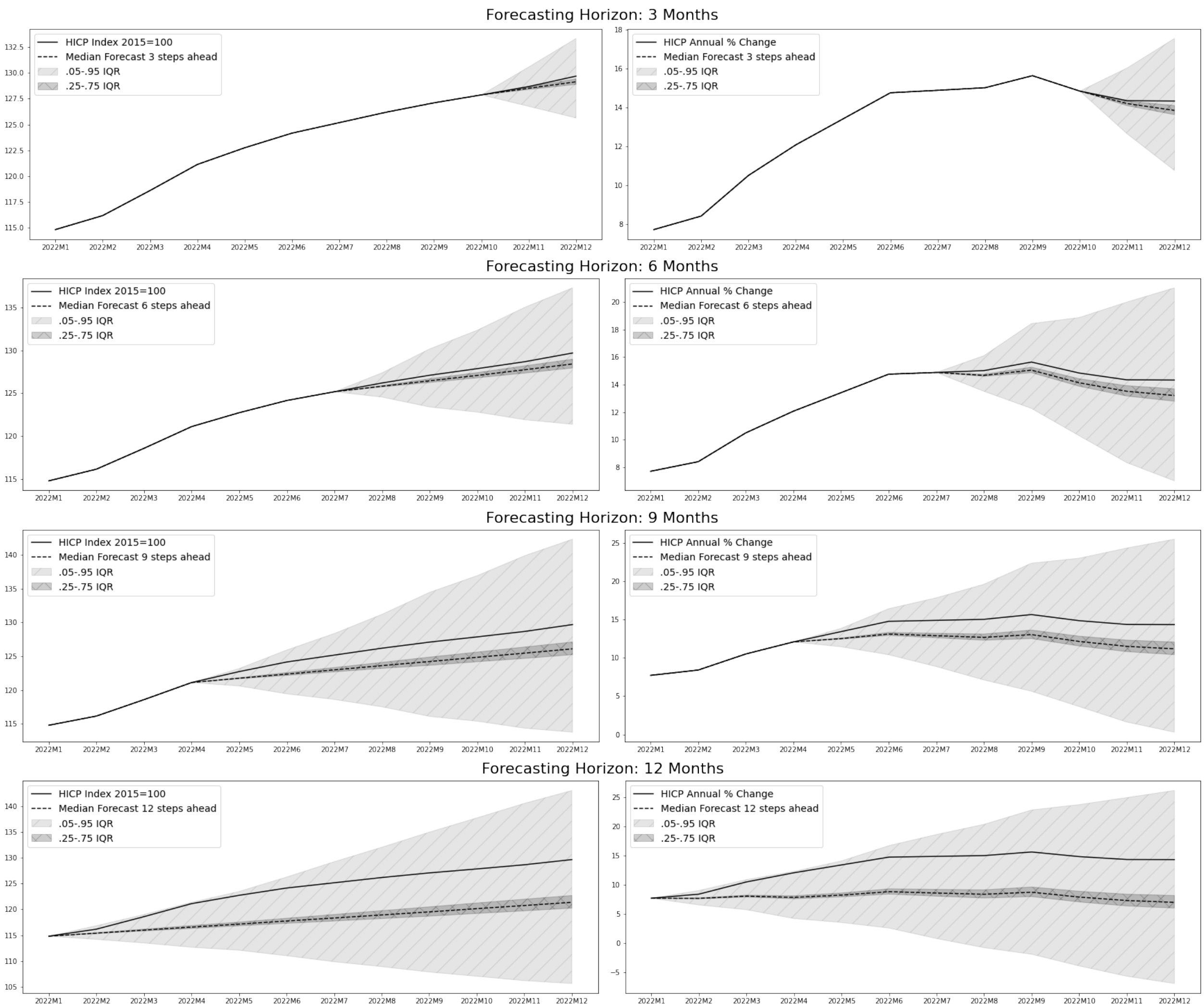
Different combinations of the indicators listed above were tested and the specification, which yielded the best performance was the one containing the lags of unemployment, the HICP, the metals price index, and the food price index.

Two separate procedures were performed in the scope of this exercise. The first involved a 1-month ahead forecasting over the test set for the purposes of uncertainty disentanglement. The second was focused on forecasting with several different forecast horizons for the purposes of constructing inflation fan charts. For the latter procedure, the actual values of the explanatory variables were used as assumptions when generating the forecasts. In the previous chapters, this part of the dataset was denoted by X_{T-h} - the explanatory variables available after time T. For both forecasting procedures, 100 samples of the parameters of the predictive distributions are generated. Additionally, the fan charts were generated for a linear AR1 benchmark using a Normal distribution as well and were compared to the proposed DQPR model across several indicators of performance.

The results from the 1-month ahead forecasting over the test set demonstrate the disentangling of the total variance of the forecasts into their aleatoric and epistemic components and compares various model specifications in terms of the size of the total variance, the relative size of the epistemic uncertainty, and the out-of-sample model performance. Firstly, epistemic uncertainty is significantly smaller compared to the aleatoric in terms of magnitude across all specifications. It is evident that a smaller total variance, as well as a smaller relative size of the epistemic variance, are related to better out-of-sample model performance, although this relationship is not unambiguous. The rationale would be that lower total variance is associated with sharper forecasts, while generally, the less well-defined the model specification, the larger the total and epistemic variances are. However, it is seen that there are exceptions, like the specification containing the lags of the HICP, the food price index, and the consumer sentiment regarding inflation, which is better in terms of out-of-sample performance compared to the one containing the lags of the HICP and the food price index, despite the larger total variance and the larger relative size of the epistemic variance. Such analysis can be further developed for comparing competing models.

The second procedure involved generating the density forecasts used for the construction of the fan charts. Four different forecasting horizons were used – 3 months, 6 months, 9 months, and 12 months. The forecasts are generated as a monthly percent change and are then converted to the HICP index and the annual percent change. The DQPR model outputs 100 samples of the distribution parameters for each time step, and the final distribution parameters used for the creation of the fan charts are the means of the parameters across the samples.

Figure 9: Inflation Fan Charts



Source: Author

The fan charts presented in Figure 9 have three main components, the central forecast, which is based on the median, the .25 - .75 inter-quantile range (IQR), and the .05 - .95 inter-quantile range. The two inter-quantile ranges separate two distinct areas of lower and higher uncertainty. It is evident from the plots that the .25 - .75 IQR is much narrower, which points out to predictive distributions with greater kurtosis. Additionally, all plots point to skewness, which is significant in some cases. In most plots, the skewness is clearly positive, with a larger area covered by the two IQRs and especially the .05 - .95 IQR. This points to strong upside risks to inflation on the basis of the explanatory variables. The actual inflation values fall outside the .25 - .75 ICQ in every case, and despite the strong upside risks to inflation, the median forecast underestimates inflation for every forecasting horizon. This underestimation is more emphasized as the forecasting horizon increases.

For both the index and the annual percent change, the widening of the fan chart is observed as the forecast horizon increases, which corresponds to increasing uncertainty. This is expected as both the index and the annual percent change are not stationary time series. This is a property of density forecasts expected almost religiously. However, as mentioned previously, this property of the density forecast depends very much on the characteristics of the time series in question.

Table 9: Forecasting Performance

Model	**Forecasting Horizon**	**RMSE**	**MAE**	**CRPS**
AR1	3 months	0.0036	**0.0025**	**0.0020**
	6 months	0.0043	0.0031	**0.0023**
	9 months	0.0048	0.0039	0.0028
	12 months	0.0050	0.0043	0.0029
DQPR	3 months	**0.0033**	**0.0025**	**0.0020**
	6 months	**0.0038**	**0.0029**	**0.0023**
	9 months	**0.0033**	**0.0027**	**0.0021**
	12 months	**0.0039**	**0.0032**	**0.0025**

Source: Author

Table 9 contains more detailed performance evaluation results. Several indicators of performance are observed across all forecasting horizons for the DQPR and the AR1 benchmark. These include the RMSE, the MAE, and the CRPS. The relatively high accuracy of the forecasts across all horizons might be attributed to the use of the actual values of the explanatory variables as assumptions in constructing the density forecasts. The AR1 and DQPR have identical performance for the MAE and CRPS in the 3-month horizon and the CRPS for the 6-month horizon, but as the forecasting horizon increases, the advantage of the DQPR is more pronounced. The improvement introduced by the DQPR against the benchmark varies between no improvement and 31.3% across the different forecasting horizons. The table below presents a summary of the improvement per indicator and per forecasting horizon.

Table 10: Performance Improvement of the DQPR Against the AR1 Benchmark

	Forecasting Horizon	RMSE	MAE	CRPS
Improvement in %	3 months	8.3%	0.0%	0.0%
	6 months	11.6%	6.5%	0.0%
	9 months	31.3%	30.8%	25.0%
	12 months	22.0%	25.6%	13.8%

Source: Author

In order to perform a sensitivity analysis and extract the variable importance from the DQPR model, the average absolute marginal contributions of each variable to the model output have been calculated using the LIME method across the test sample. The explainability results are presented in terms of relative importance in Table 11.

Table 11: Relative Variable Importance

Indicator	Variable Importance
Unemployment	38.40%
Metals Price Index	28.76%
Food Price Index	23.68%
HICP	9.16%

Source: Author

The results point out that unemployment has the highest relative importance, but the external prices indices for food and metals prices cumulatively hold over 52% of the relative performance. The lag of the HICP has the lowest relative importance. This points to a model driven significantly by external factors, but the core factor of unemployment is as much crucial for the superior performance of the DQPR.

The period following the outbreak of the coronavirus pandemic has been characterized by high inflation across Europe and the world. The first effects were felt as the pandemic-related lockdowns caused supply bottlenecks for many goods. Eventually, as the pandemic-related recessions subsided demand recovered rapidly, but the supply could not adjust accordingly. A notable example was the global semiconductor shortage, which affected multiple large industries like the computer, automobile, and appliances industries. Eventually, the Russian invasion of Ukraine was the trigger for an unprecedented energy crisis in Europe, which affected energy prices globally. Energy inflation had strong spillovers to other sectors of the economy (Yagi and Managi, 2023) and fueled expectations for high and persistent inflation (Kilian and Zhou, 2022).

The current study addressed the topic of inflation forecasting in a high-inflation environment. Fan charts are an established tool for communicating inflation forecasts and the uncertainty surrounding them, used regularly by central banks around the world. Therefore, it was decided to apply the proposed deep quantile-based probabilistic regression (DQPR) in the task of constructing fan charts, which can account for both skewness and kurtosis. The overall architecture of the model is almost identical to the one used for forecasting the natural gas prices on the Balkan Gas Hub. It was extended using the Monte Carlo Dropout technique in order to allow for the estimation of parameter uncertainty, which is a subcomponent of epistemic uncertainty.

The focus of this exercise was modeling and forecasting consumer price inflation in Bulgaria, measured by the HICP. For this purpose, a monthly sample covering the period between 2000 and 2022 was collected, including both endogenous and exogenous factors with respect to the Bulgarian economy. The final set of indicators used in the model were unemployment, food prices index, metals price index, and the lag of the HICP.

Using 80% of the total sample the DQPR model was trained and evaluated in a pseudo-out-of-sample exercise on the remaining 20%. Forecasts with four different forecasting horizons were generated – 3 months, 6 months, 9 months, and 12 months. The resulting fan charts constructed with these forecasts exhibit strong positive skewness, which points out to strong upside risks to inflation in the observed period. On the other hand, the central median forecast underestimates actual inflation in the period covered by the test sample. The DQPR performance was compared to an AR1 benchmark and while the two models have close to identical performance in the 3-month horizon, as the forecasting horizon grows the advantage of the DQPR is more significant and the model achieves an improvement in performance compared to the benchmark of up to 31.3%.

Additionally, to the creation of the fan charts, uncertainty was quantified and separated into aleatoric and epistemic using the 1-step ahead forecasts on the test samples. From this procedure total variance was decomposed into its aleatoric and epistemic components and various model specifications were compared in terms of total forecast variance, relative size of epistemic variance, and out-of-sample performance. Although lower total variance and lower relative epistemic variance are associated with better out-of-sample performance, the relationship is not unambiguous. However, such analysis can be further developed for the purposes of comparing competing models.

Finally, a relative variable importance was calculated for the DQPR and the results point to a model driven heavily by external factors like food and metals prices. However, unemployment has the highest relative performance, which supports the theoretical underpinnings of the Philip's curve.

The current chapter demonstrates that deep learning techniques like the DQPR can be used for inflation forecasting and for communicating the forecasts in the form of a fan chart. Additionally, this approach allows for a more in-depth analysis of uncertainty, consisting of the quantification and the disentanglement of aleatoric and epistemic uncertainty. Future efforts might focus on a more thorough quantification of epistemic uncertainty and more specifically the estimation of model uncertainty.

IV. Scientific and Applied Contributions

The contributions of the disseration can be separated into three broad categories – scientific contributions, scientific-applied contributions, and methodological contributions. The main fields of the contributions are economics, econometrics, and machine learning.

Among the scientific contributions are:

- A novel method to improve economic forecasts is proposed, that leverages a neural network architecture for probabilistic time-series forecasting, termed deep quantile-based probabilistic regression (DQPR).

Among the scientific-applied contributions are:

- The proposed DQPR model outperformed a set of benchmarks when applied to nowcasting the pandemic-related recessions in the four Eastern European countries, which is novel in both scope and results for the economics literature.
- The proposed DQPR model outperformed both statistical and deep learning benchmarks when applied to forecasting natural gas prices on the Balkan Gas Hub during a period of extreme volatility, which is novel in both scope and results for the economics literature.

Among the methodological contributions are:

- A Bayesian version of the DQPR model is developed and applied in constructing an inflation fan chart for Bulgaria, as well as quantifying and disentangling aleatoric and epistemic uncertainty.
- The LIME algorithm for interpretable machine learning is applied to the DQPR model in order to perform sensitivity analysis and gain insights into global and local model explainability.

The above-mentioned contributions lead to new results in both the fields of economics and machine learning and a clear methodology for practical applications.

V. Relevant Publications

Parts of the dissertation have been published as stand-alone studies in two journals. There are plans to prepare at least two more publications on the topic of the dissertation.

1. Mihail Yanchev, 2023. "Uncertainty - Definition and Classification for the Task of Economic Forecasting," Bulgarian Economic Papers bep-2023-03, Faculty of Economics and Business Administration, Sofia University St Kliment Ohridski - Bulgaria // Center for Economic Theories and Policies at Sofia University St Kliment Ohridski, revised Mar 2023.

2. Mihail Yanchev, 2022. "Deep Growth-at-Risk Model: Nowcasting the 2020 Pandemic Lockdown Recession in Small Open Economies," Economic Studies journal, Bulgarian Academy of Sciences - Economic Research Institute, issue 7, pages 20-41.

The presented publications meet the requirements of Art. 12 of the Development of Academic Staff in the Republic of Bulgaria Act (DASRBA), art. 35 of the Regulations for application of DASRBA (RADASRBA) and fulfill the national quantitative requirements under Art. 1a, para. 1 of RADASRBA for Area 3: Social, economic, and legal sciences, Professional field 3.8 Economics.

VI. Conclusion

The research presented in this dissertation relied on concepts and findings from the fields of economics, econometrics, statistics, and machine learning. Its main purpose was to enrich the knowledge of probabilistic forecasting in economics and encourage the adoption of probabilistic forecasting over point forecasting within the field. Through its main contribution, the dissertation complements the economic forecasting toolkit with a deep learning approach to probabilistic forecasting, which has proven its reliability and transparency in three challenging empirical tasks. Judging by empirical results, the novel approach to modeling uncertainty proposed in the dissertation is a valuable tool for forecasting in the context of rare events like the recent coronavirus pandemic or the war in Ukraine. The general approach of probabilistic forecasting and the specific methodology presented in the dissertation could be valuable to the decision-making process in both the public and the private sectors.