Dimensionality Reduction for Enhanced Economic Diagnosis: A BGPLVM Approach

Luis Sebastian Weekes

MSc in Data Science
The University of Bath
2024

Dimensionality Reduction for Enhanced Economic Diagnosis: A BGPLVM Approach

Submitted by: Luis Sebastian Weekes

Copyright

Declaration

This dissertation is submitted to the University of Bath in accordance with the requirements of the degree of Master of Science in the Department of Computer Science. No portion of the work in this dissertation has been submitted in support of an application for any other degree or qualification of this or any other university or institution of learning. Except where specifically acknowledged, it is the work of the author.

Abstract

Determining the state of the economy is a difficult task. There are many factors and perspectives that need to be considered for holistic analysis. Unfortunately, many of these variables and perspectives are often forgotten or ignored. This paper experiments with the Bayesian Gaussian Process Latent Variable Model (BGPLVM), a dimensionality-reduction technique, to simultaneously account for several variables for economic analysis. This approach reduces variable selection bias and enables a robust economic diagnosis. Dimensionality-reduction techniques, particularly those that address non-linear mappings and uncertainty, remain underutilized in economics, presenting opportunities for advancements in handling high-dimensional data, data visualization, and model performance in the field. Two data sets were used for training six models—a data set comprised of mainly macroeconomic features and a data set representative of the experiences of the average individual. A PCA and two BGPLVM models (with different kernels) were trained on each data set. The fit of the models were examined and the positioning of the latent points (years) were analysed. The models successfully distinguished between years based on their economic profiles. This paper introduces a novel framework, using BGPLVMs, that enhances economic diagnosis by providing a holistic and targeted analysis, potentially guiding more informed policy decisions.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to give thanks to my family (**Myrilu, Bryan, and Rebecca Weekes**) and acknowledge the support of my partner—**Sarah Catherine Howard** and her family.

I would not have been able to write this paper without the shaping experiences that I had during my time at **Kenyon College**. In particular, I want to express my gratitude towards the marvelous professors I had in the Mathematics and Economics departments for pushing me academically, believing in me and indulging in my interdisciplinary pursuits. Furthermore, I would not have been able to write this paper without the mentorship of my dear friend **Andrew Nguyen** who I met at Kenyon College.

Thank you to the Computer Science Department at the University of Bath for the opportunity of deepening my knowledge in the field of Data Science and special thanks to **Emelie Barman** for guidance on this paper. I am grateful for the friends and colleagues I made during this year.

Thank you to anyone who listened to and conversed with me in relation to this paper.

Chapter 1

Introduction

Economic policies can be a matter of life and death. The setting of macroeconomic policy is crucial for financial stability, reduced economic volatility, and enhanced economic growth which in turn have major impacts on the well being of people – particularly the most vulnerable Ames and Izquierdo (2020). Regrettably, the state of the economy which influences the setting of macroeconomic policies is often of debate. For instance, the main purpose of (Berge and Jordà, 2011) is to evaluate the economic classification performance of the National Bureau of Economic Research in the United States. In fact, (Berge and Jordà, 2011) argue that there are no formal definitions for economic activity and that the true states of economic activity are un-observable–even in retrospect. Therefore, it is paramount that tools (or more specifically, models) that can account for the multitude in variables at an economist's disposal are developed. Such tools would aid economists in effectively diagnosing the state of the economy and develop policies to suit unique and nuanced events.

Unsupervised learning is the process of finding patterns in the underlying structures of data–without labels or rewards (Ghahramani, 2003). Since periods of economic activity are not easily labelled, unsupervised learning could help provide a structured analysis for the likeness of economic time periods based on a multitude of metrics. The majority of unsupervised learning models focus on learning probabilistic representations of data (Ghahramani, 2003). Unsupervised learning is helpful in outlier detection and monitoring – both crucial in economics. Classification is another common task in unsupervised learning. Two classic examples of unsupervised learning are clustering and dimensionality reduction. Dimensionality reduction techniques can be employed in order to engineer variables that account for a multitude of real variables–potentially enabling economists to factor several variables into their analysis.

1.1 Macroeconomics

Economics is the study of scarcity and its implications on how resources are managed to maximize outputs. Some of these outputs include goods and services amongst other outputs of concern to society (University at Buffalo, 2024).The management of these resources can have drastic impacts on public welfare (Ames and Izquierdo, 2020). In Data Science terms, a major goal of economists is to find equilibria in the economy or to optimize the economy itself. There are two main branches in economics–microeconomics and macroeconomics. Microeconomics is the study of firms and individuals and how they make choices in light of scarcity (Krugman and Wells, 2020). This paper will focus on macroeconomics which is the study of the structure and performance of national economies and of the policies that governments use to affect those performances (Abel, Bernanke and Croushore, 2011).

One of the central objectives in macroeconomics is to identify factors that drive fluctuations in a nation's economic activity (Abel, Bernanke and Croushore, 2011). In order to examine this question effectively, economists must first understand the state of the economy. A popular method of describing the state of the economy is the **business cycle**. A business cycle is a repeated sequence of economic growth, decline and subsequent recovery (Abel, Bernanke and Croushore, 2011). Periods of economic growth are called **expansions**, while periods of economic contraction are called **recessions**, but if the period of economic contraction is particularly poor they are called **depressions**. This view, however, is rather simplistic (Stiglitz, Sen and Fitoussi, 2009). It primarily focuses on economic growth and does not take into account more nuanced variables. For instance, a more complex economic phenomenon is **stagflation**, where both unemployment and inflation rates are high. Stagflation is particularly rare, because usually unemployment and inflation are negatively correlated (Abel, Bernanke and Croushore, 2011).

Part of what makes macroeconomics so interesting (yet challenging) is that there are many different ways to measure economic activity. Economists vigorously debate on the state of the economy and which variables are of importance (Berge and Jordà, 2011). Some of the most common variables used to study the economy are Gross Domestic Product (GDP), interest rates, inflation and unemployment—the first of which is the most widely used measure of economic activity (Stiglitz, Sen and Fitoussi, 2009). GDP is a measure of how much the national economy is producing; it can be measured from three equivalent perspectives: total production, total income or total expenditure (Abel, Bernanke and Croushore, 2011). Despite GDP being the most commonly used metric for assessing the economy, there are serious issues with the metric. For instance, GDP does not account for wealth distribution or material living standards (Stiglitz, Sen and Fitoussi, 2009). Interest rates, which are usually set by central banking authorities, define the rate at which entities can borrow money—essential for businesses and regular people. Inflation is a metric that accounts for the rate at which prices are increasing. Unemployment is the ratio of those unemployed and looking for employment compared to the labor force. While these variables capture different aspects of the economy, there are many other variables, like consumer sentiment, that capture nuances in the economy and better represent the experiences of the everyday person that are often overlooked (University of Michigan, 2024).

1.2 Dimensionality Reduction

In modern times we are blessed by the abundance of data; not only can we capture many more observations than before, but we are also able to capture many more features for given observations. Data with many features or variables are called high-dimensional (data). Two major issues arise with high-dimensional data. First, the geometrical properties of these data sets are not intuitive and are very different to those in two- or three-dimensions. Furthermore, data analysis tools are also usually designed for lower-dimensional data—as it is hard to see in more than three dimensions (Verleysen and François, 2005).

The curse of dimensionality is a common topic of discussion, as it pertains to high-dimensional data. The basic challenge is that with each extra dimension to be learned by a model, the number of observations needed by the model to learn effectively grows exponentially (Verleysen and François, 2005). The curse of dimensionality is also associated with any unfavorable effects on learning algorithms arising from high-dimensionality (Verleysen and François, 2005).

Neil Lawrence had an amusing (yet admittedly controversial) stance on the curse of dimensionality in one of his lectures. Lawrence stated, " It [the curse of dimensionality] does not exist in the data. It exists in your model. You can create models that exhibit the curse of dimensionality by assuming independence (Lawrence, 2014b)." In other words, Lawrence implied that there are models (more specifically, dimensionality-reduction models) that can deal with the curse of dimensionality effectively.

Consequently, dimensionality-reduction techniques have been developed to address the breakdown in interpretability, accuracy, and efficiency of models where there is high dimensional data. In essence, dimensionality reduction is the process where you map data from a high dimensional space to a low dimensional space while trying to capture as much information from the data as possible (Jia et al., 2022). In effect, they transform data matrices from $n \times d$ dimensions to $n \times q$ dimension where $q << d$. As a result, new features derived from the original features are generated, capturing a significant portion of the data's underlying information.

Principal Component Analysis (PCA) is the usual starting point in relation to dimensionality-reduction models. The objective of PCA (Jolliffe et al., 1986a) is to reduce the dimensionality of data while capturing as much signal from the data as possible. PCA achieves this by transforming the data into a new set of uncorrelated variables named principal components. Principal components are computed via singular-value decomposition (SVD). The solution to the SVD problem produces ranked eigenvectors (principal components) based on their eigenvalues.

$$C = U\Sigma V^\top$$

SVD is the decomposition of a covariance matrix into three distinct matrices where C is the $M \times N$ covariance matrix, U is an $M \times M$ matrix containing the left-singular vectors, Σ is an $M \times N$ matrix containing singular values and $V^\top$ is an $N \times N$ matrix containing the right singular vectors (Jolliffe et al., 1986a).

The principal components (PCs) are found in U; they are then chosen based on the size of their corresponding singular values in Σ. $V^\top$ contains the loadings of the PCs which describe the make up of the PCs in terms of the original variables. The first few of these principal components retain the most variation in the data set and hence we can analyse the data with these fewer components; this process effectively reduces the dimensionality of the data (Jolliffe et al., 1986a). In essence, PCA allows for data to be represented in lower dimensions that represent more features. For instance, PCA allows for data to be visualised in two or three dimensions (PCs) that account for a larger number variables. Regrettably, the PCs produced by PCA are linear combinations of the original features of the data and are hence unable to account for non-linear relationships between the data and reduced dimension space. This fact is a major flaw in the model since real world data is often non-linear.

Factor Analysis (FA) (Spearman, 1904) aims to explain a covariance matrix via a small number of 'factors' (Lawley and Maxwell, 1962). The main distinction between PCA and FA is that FA reduces the dimensionality in data through an explicit model, whereas in PCA, SVD is applied to obtain principal components (Jolliffe et al., 1986b). If there are reasonable contextual assumptions to be made about a dataset, FA should be used to reduce dimensions, because FA would increase the interpretability of the components. On the other hand, if no reasonable assumptions of the underlying data can be made, PCA is the better option.

Probabilistic PCA (PPCA) is a probabilistic extension of PCA. Probabilistic PCA is derived a from Gaussian latent variable model which is closely related to statistical factor analysis (Tipping and Bishop, 1999). A major difference between FA and PPCA is in the handling of noise. PPCA assumes isotropic Gaussian noise whereas FA assumes a different noise variance for each variable (Tipping and Bishop, 1999); this presents a potential weakness in PPCA. The assumption that all variables have similar noise structures is unrealistic, yet it allows for a simpler model. Moreover, PPCA's probabilistic approach is a valuable feature, because real data is messy, volatile, and involve unknown factors. PPCA can also be used to generate educated guesses in missing data points due to its probabilistic framework, which is another useful feature (Tipping and Bishop, 1999) However, PPCA is still unable to account for non-linear mappings between the data and the latent space.

The **Gaussian Process Latent Variable Model (GPLVM)** is a non-linear, probabilistic extension of PCA (Lawrence, 2005). The GPLVM utilizes Gaussian Processes in order to capture the non-linear relationships between latent variables and the data—a feature that makes the GPLVM more robust than PCA and PPCA. The GPLVM is probabilistic in that the mappings from the latent space to the data are captured by Gaussian processes, but the latent points are point estimates.

The **Bayesian GPLVM (BGPLVM)** is a Bayesian extension of the GPLVM (Titsias and Lawrence, 2010). The BGPLVM utilizes variational inference to provide posterior distributions for the latent points; enabling the model to account for uncertainty in the latent points.

1.3 BGPLVMs for Macroeconomics

Relying such few macroeconomic variables (such as GDP) to study the economy is one-dimensional, because they do not capture the economy as a whole. Dimensionality-reduction techniques offer a more robust and nuanced approach by incorporating many different variables into the analysis of economic time periods. The incorporation of many variables into the model helps to address the bias that the field of economics has in using variables such as GDP—a bias which can have negative consequences (Ames and Izquierdo, 2020). Therefore, not only can the BGPLVM account for many macroeconomic variables in the model, the model can provide a more human-centred approach to the economy through the incorporation of variables that account for living standards, such as pollution and real disposable income into the model. Furthermore, the BGPLVM would be able to provide uncertainty estimates for the positioning of each given time period in terms of the data.

Consequently, the aim of this paper is to develop a frame work, through the use of Bayesian GPLVMs, to differentiate between types of economic activity (using many different variables). In effect, this tool could assist economists in diagnosing the state of the economy holistically and with degrees of uncertainty—enabling them to focus on solutions.

Chapter 2

Literature Review

2.1 Justifying Machine Learning in Economics

Over the last decade, machine learning (ML) has been utilised more in the toolkit of economists (Storm, Baylis and Heckelei, 2020). Machine learning techniques have been leveraged more in recent times due to the availability of large Macroeconomic data sets, the rapid development of machine learning algorithms, and compute power (Storm, Baylis and Heckelei, 2020). According to some experts, machine learning techniques offer superior modeling capabilities compared to traditional econometric approaches, such as various forms of regression (Storm, Baylis and Heckelei, 2020) (Goulet Coulombe et al., 2022). However, (Storm, Baylis and Heckelei, 2020) mentioned that economic researchers are still cautious upon using certain machine learning techniques due to their concerns on transparency, interpretability, and their ability to find causal relationships. Finding causal relationships is crucial in economics, because economics deals with adjustments of inputs to change outputs that are justified by causal relationships. These reservations are mostly attributed to deep learning (Storm, Baylis and Heckelei, 2020). While these concerns are valid, the goal of this paper is to utilize machine learning techniques in an experimental fashion that find interesting patterns while being as robust as possible in retaining explanation power. Economists are also identifying that machine learning techniques are not only helpful in their predictive power, but also in their identification of causal relationships (Storm, Baylis and Heckelei, 2020).

(Storm, Baylis and Heckelei, 2020) detail weaknesses in traditional econometric approaches and how machine learning can address them. One major ML technique that traditional econometric approaches can benefit from is regularisation—which is the modification of a model in order to better generalize to unseen data (Storm, Baylis and Heckelei, 2020). Over-fitting in traditional econometric models can be problematic, because the models might not generalize enough to be useful in real-world settings. The main goal of regularisation is to prevent the model from over-fitting to the training data. An example of regularisation in machine learning is the LASSO regression (Tibshirani, 1996). The LASSO regression adds a penalizing term to the model that reduces the value of coefficients and can even set some coefficients to zero. Machine learning algorithms are highly flexible in comparison to econometric models; this flexibility allows for machine learning models to fit more nuanced relationships (Storm, Baylis and Heckelei, 2020)(Goulet Coulombe et al., 2022).

The GPLVMs used in this paper account for both regularisation and non-linearities robustly

(Titsias and Lawrence, 2010). (Goulet Coulombe et al., 2022) discuss that ML techniques' handling of non-linearities are profoundly important in data-rich environments. ML techniques substantially increase forecasting accuracy for all macroeconomic variables in studies—especially when predicting over long horizons. Furthermore, the Bayesian GPLVM also accounts for uncertainty by providing posterior distributions over latent points. These few benefits alone are enough motivation to explore machine learning in the context of Macroeconomics.

2.2 Dimensionality Reduction in Macroeconomics

2.2.1 PCA in Macroeconomics

In the context of unsupervised learning, (Storm, Baylis and Heckelei, 2020) find that dimensionality-reduction techniques such as PCA could be used to find logical groupings of economic data which could then be further analysed. (Storm, Baylis and Heckelei, 2020) mention that PCA relies on linear partitions of the variable space, but data in practice is usually non-linear.

In (Issah and Antwi, 2017), the authors' goal is to predict the well-being of a firm based on macroeconomic conditions. The goal of (Issah and Antwi, 2017) is relatively similar to this paper's goal in that time periods are being evaluated based on economic conditions. The authors employ vanilla PCA to extract components from their data; these principal components are then fed into a multiple linear regression model. (Issah and Antwi, 2017) demonstrate that using PCA on a large data set of arbitrarily selected macroeconomic variables effectively reduced the variable selection bias in the study; a shared key goal for this paper. (Issah and Antwi, 2017) also exhibit that PCA addresses the multicollinearity of many Macroeconomic variables, a common issue in economics. Even-though, (Issah and Antwi, 2017) successfully make predictions and effectively implement PCA in a macroeconomic setting, the implementation of Bayesian GPLVM would add the capacity to model for non-linearities in the data and account for uncertainty in the latent space.

(Iyetomi et al., 2020) analyse the movements of 57 macroeconomic variables are able to confirm statistically significant co-movements among the variables. Furthermore, they also identify noteworthy economic events such as the dot-com bubble in 2001 as well as the Global Financial Crisis (GFC) in October 2008. (Iyetomi et al., 2020) use Complex Hilbert Principal Component Analysis (CHPCA) and Rotational Random Shuffling (RRS) and aim to build on the efficacy of PCA in macroeconomics via CHPCA (CHPCA is an extension of PCA in which an imaginary component is introduced to the original time series via a Hilbert transformation). An imaginary number framework enables the model to account for leading and lagging information in the correlation matrix, which allows for the model to perform significantly better than vanilla PCA. While the introduction of Hilbert transformation is a substantial improvement over standard PCA, it does not account for non-linearity and uncertainty in the data.

(Barbarino and Bura, 2015) find success in forecasting economic variables by utilising PCA and linear regression. (Barbarino and Bura, 2015) take many features and reduce the dimensionality of the data via PCA to find groupings in economic activity implicitly. After reducing the dimensionality of the data, the principal components are fed into multiple linear regressions for forecasting. In contrast, this study aims to find non-linear mappings between the data and latent variables to find groupings of economic activity. (Barbarino and Bura, 2015) claimed that the linearity constraints of models such as PCA do not gravely affect the outcome of their results, but the GPLVMs used in this study account for non-linearities regardless.

2.2.2 Probabilistic PCA, Variational Inference, Gaussian Processes and other Dimensionality Reduction Techniques

In (John, Ekpenyong and Nworu, 2019), five different PCA models were evaluated based on performance in relation to the imputation of missing values in economic and financial time series data. In relation to PPCA, (John, Ekpenyong and Nworu, 2019) find that PPCA works well on data sets with a proportion of missing values between 10 and 15 percent. Also, (John, Ekpenyong and Nworu, 2019) examine Bayesian PCA (BPCA), where, instead of using a maximum likelihood estimator to account for uncertainty as in PPCA, in BPCA, Bayesian inference is used to estimate the principal components. In effect, PPCA estimates the principal components probabilistically, while BPCA actually provides posterior distributions over the components. Both PPCA and BPCA performed better than vanilla PCA in relation to normalized root mean squared error in the context of imputing missing data (John, Ekpenyong and Nworu, 2019). Furthermore, (John, Ekpenyong and Nworu, 2019) employ the non-linear Local Least Squares PCA (LLSPCA) in the imputation exercise. LLSPCA is based on the linear combination of the *k*-nearest neighbors of a missing dataset; in effect, LLSPCA uses the absolute value of the distance between variables as a measure of strength for the variables (John, Ekpenyong and Nworu, 2019). Out of all models, LLSPCA performs the best in relation to the imputation of missing values, which contrasts the claims of (Barbarino and Bura, 2015) in which accounting for non-linearities in their data was not essential. Regrettably, LLSPCA does not model for uncertainty in the components. Even though (John, Ekpenyong and Nworu, 2019), study models that address both uncertainty and non-linearity in data, models that do both such as the Bayesian GPLVM are not examined.

Variational inference adds more power to an economist's tool kit in that it can increase model flexibility by allowing for a larger number of parameters to be learnt (Storm, Baylis and Heckelei, 2020). In other words, variational inference proposes simple distributions to estimate more complex distributions. In effect, variational inference trades some accuracy for performance (Blei, Kucukelbir and McAuliffe, 2017). (Athey et al., 2018) employ variational inference to calculate latent variables in the context of the restaurant business, and in effect predict demands. In addition, (Ruiz, Athey and Blei, 2020) utilise variational inference to create a probabilistic model of consumer choice in the context of super markets. Another interesting use of variational inference in economics is where (Hoberg and Phillips, 2016) attempt to capture changes in industries over time using companies' product descriptions. Although these use cases are very interesting, variational inference is only a part of the Bayesian GPLVM model.

In (Geweke, Koop and van Dijk, 2011), there are two examples where Gaussian processes are used in application, one in economics and one in finance. (Geweke, Koop and van Dijk, 2011) examine the application of Dynamic Stochastic General Equilibrium Models (DSGEs). DSGEs are a broad class of dynamic macroeconomic models that extend the neoclassical growth model (Solow-Swan growth model) as discussed by (King, Plosser and Rebelo, 1988). In effect, the Solow-Swan growth model is a function that explains GDP in terms of technology, hours of labor and the capital stock (machinery, infrastructure, tools, etc.). (Geweke, Koop and van Dijk, 2011) illustrate how many complex problems can be solved in DSGEs if innovations in technology as Gaussian. Furthermore, (Geweke, Koop and van Dijk, 2011) provide an example where Gaussian processes are used in the field of finance to estimate continuous time models from discrete data. However, neither of these examples take advantage of clustering or dimensionality reduction.

2.2.3 GPLVMs and Most Recent Literature

A recent Ph.D. thesis, (Guo, 2023), extensively outlines the use of dimensionality reduction techniques for Macroeconomic forecasting. While (Guo, 2023) utilizes techniques similar to GPLVMs, the goal of (Guo, 2023) is forecasting. (Guo, 2023) also explores the difference between sparse and dense dimensionality-reduction techniques and their performances. Furthermore, the main model utilized in (Guo, 2023) (TVP-3PRF) does not address non-linearity or uncertainty.

(Nirwan, 2020) uses GPLVMs in an economic and financial setting. More specifically, (Nirwan, 2020) applies the GPLVM to the returns of several assets in order to create a predictor for missing asset prices and to identify structure in financial data. However, (Nirwan, 2020) does not capitalize on the features of the Bayesian GPLVM which enable for uncertainty estimates to be produced for latent points. Furthermore, the application of (Nirwan, 2020) is distinct to this paper.

(Hauzenberger et al., 2024) recently employs the use of Gaussian Process Vector Autoregressions to determine economic relationships and how they change over time. The aim of Hauzenberger et al. (2024) is relatively similar to this paper, except that GPLVM and Bayesian GPLVM models will be employed to attempt to differentiate between types of economic activity. The main difference between the models in this study and that of (Hauzenberger et al., 2024) is that GP-VARs are primarily used for modeling multivariate time series data, where the aim is to find links between multiple time series variables over time, whereas GPLVMs are dimensionality reduction techniques.

Chapter 3

Methodology

The objective of this paper is to develop a tool that could potentially diagnose economic activity. Two perspectives have been chosen as represented by the data sets used in training the models. One model will be used in a more classical macroeconomic approach; this model will account for mostly macroeconomic variables. This shall be called "the Macro Model." However, the Macro Model still accounts for a small proportion of variables that reflect the human experience of the economy such as wealth inequality and CO2 emissions. The second model uses a subset of the data used by the Macro Model. This subset of variables represents a more balanced approach to measuring yearly economic activity in terms of the human experience–this shall be called "the Human Experience Model." The Human Experience Model aims to reconcile the phenomenon of the economy doing well while the average person struggles as outlined in (Jackson, 2009).

3.1 The Models

3.1.1 Gaussian Processes

A stochastic process is a collection of random variables indexed by a set. Stochastic processes are characterized by a probability distribution that is consistently defined for every possible subset of the random variables.(Williams and Rasmussen, 1995). A Gaussian processes (GP) is a stochastic process that specifies distributions over function spaces (Lawrence, 2005). GPs can be fully specified by a mean and covariance function (often called a kernel). The kernel ensures that any finite set of points follows a joint multivariate Gaussian distribution (Williams and Rasmussen, 1995). The predictive distribution for new data is obtained via an $(n + 1)$ dimensional joint Gaussian for the n training data and the predicted value by conditioning on the training observations (Williams and Rasmussen, 1995). In effect, a distribution over functions is formed by marginalising over infinite functions proposed by a kernel that takes into account observed data; this predictive function also accounts for uncertainty in its estimates.

The squared-exponential kernel (see formula 3.1) is the most commonly used kernel (Williams and Rasmussen, 2006). The squared-exponential kernel has three hyper parameters: the length scale (L), the signal variance (σ^2), other wise-known as the amplitude, and the noise variance (σ_n^2) (Williams and Rasmussen, 2006). The length scale defines how much the correlation between points decreases as their distance increases; a shorter length scale represents complex functions that change direction frequently, while longer length scales represent smoother functions. The signal variance accounts for the vertical variation in the function. The noise variance accounts for noise in the data. The choice of a kernel is critical since they determine the generalization properties, and hence behavior of the GPs (Duvenaud, 2014). Some other commonly used kernels are the linear kernel (Formula 3.2) and periodic kernel (Formula 3.3) (Williams and Rasmussen, 2006) (Duvenaud, 2014). In order to capture complex relationships in data, kernels are combined through multiplication and addition—although this may lead to more computational costs (Duvenaud, 2014).

Squared Exponential Kernel

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{1}{2L^2}\|x - x'\|^2\right) \tag{3.1}$$

Linear Kernel

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2(\langle \mathbf{x}, \mathbf{x}' \rangle + C) \tag{3.2}$$

Where C is a constant.

Periodic Kernel

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{2}{L^2}\sin^2\left(\frac{\pi\|x - x'\|}{P}\right)\right) \tag{3.3}$$

Where P determines the distance between repetitions in a function.

3.1.2 Gaussian Process Latent Variable Models (GPLVMs)

The Gaussian Process Latent Variable Model is a non-linear, probabilistic extension of PCA (Lawrence, 2005). The use of Gaussian Processes is what enables the GPLVM to find non-linear mappings between the data and latent points. Gaussian processes are employed to map the data to the latent space by modifying the probabilistic PCA model (Lawrence, 2005). The most important modification is that the inner-product kernel, used in a modified version of probabilistic PCA, is replaced with a covariance function (kernel) that allows for non-linear functions and hence non-linear mappings (Lawrence, 2005). Furthermore, the GPLVM is probabilistic in the mapping from latent variables to data because of the probabilistic nature of GPs. Another benefit that GPLVMs have over PCA and PPCA is that GPLVMs have a built in regularizing process through the use of Gaussian priors over the latent variables in the optimization process 3.5 Lawrence (2003) Titsias and Lawrence (2010).

In order to train the GPLVM, the joint log-probability of the model must be maximised 3.4. In (Lawrence, 2005), the primary methodology of training the model was to find the Maximum A Posteriori estimate of X whilst jointly maximizing with respect to the hyper parameters.

Log Joint Probability for GPLVM

$$\log P(\mathbf{Y}, \mathbf{X} \mid \theta) = \log P(\mathbf{Y} \mid \mathbf{X}, \theta) + \log P(\mathbf{X}) \tag{3.4}$$

Prior Over Latent Variables

$$P(\mathbf{X}) = \prod_{i=1}^{N} \mathcal{N}(\mathbf{x}_i \mid 0, \mathbf{I}) \tag{3.5}$$

Gaussian Process Likelihood

$$P(\mathbf{Y} \mid \mathbf{X}, \theta) = \prod_{j=1}^{D} \mathcal{N}(\mathbf{y}_j \mid 0, \mathbf{K}) \tag{3.6}$$

Where K is the $N \times N$ covariance matrix defined by the kernel in formula 3.1.

GPLVMs can be computationally expensive because of the inversion of matrices, but sparsification algorithms can be employed to provide quicker yet accurate results (Lawrence, 2003).

3.1.3 Bayesian GPLVM (BGPLVMs)

The Bayesian GPLVM is a Bayesian extension of the GPLVM (Titsias and Lawrence, 2010). The Bayesian Gaussian Process Latent Variable Model (GPLVM) treats latent variables as random variables by assigning a prior distribution to them. The posterior distribution of these latent variables is then approximated using variational inference methods (Titsias and Lawrence, 2010). This allows for the BGPLVM to account for uncertainty in the latent space, whereas the standard GPLVM can only produce point estimates. Another improvement of the Bayesian GPLVM over regular GPLVM is that the former is less prone to over-fitting due to the latent variables being represented as posterior distributions (Titsias and Lawrence, 2010).

In the Bayesian GPLVM, the goal is to compute the marginal likelihood of the data (see Formula 3.7) in order to compute the true posterior over the latent variables $P(\mathbf{X} \mid \mathbf{Y})$ through Bayes' Theorem (see Formula 3.8).

marginal likelihood of the data

$$P(\mathbf{Y}) = \int P(\mathbf{Y} \mid \mathbf{X}, \theta) P(\mathbf{X}) dX \tag{3.7}$$

Bayes' Theorem

$$P(\mathbf{X} \mid \mathbf{Y}) = \frac{P(\mathbf{Y} \mid \mathbf{X}) P(\mathbf{X})}{P(\mathbf{Y})} \tag{3.8}$$

However, the marginal likelihood of the data (see Formula 3.7) is intractable as X is non-linear (Titsias and Lawrence, 2010). Therefore, we must approximate $P(\mathbf{X})$ via variational inference. Variational inference is the process of maximizing the Evidence Lower Bound (ELBO), which balances the log-likelihood of the observed data with a regularization term that enforces the similarity between the proposed and the posterior distributions (Jordan et al., 1999). In essence, variational inference proposes Gaussian distributions iteratively until the best approximation is found.

Another key feature of the Bayesian GPLVM is the use of inducing variables (Titsias and Lawrence, 2010). Inducing variables are a set of pseudo-inputs that are used to train the model—their corresponding function values are used to summarize the data (Titsias, 2009). In the context of the Bayesian GPLVM, inducing variables help to accurately approximate GP mappings, making the model computationally feasible.

Another benefit of Bayesian GPLVM model is the incorporation of the automatic relevance determination (ARD) squared exponential kernel in Formula 3.9 (Geiger, Urtasun and Darrell, 2009). The ARD kernel enables the model to infer automatically the dimensionality of the non-linear latent space by learning the length scales of each dimension (Lawrence, 2014a). Latent dimensions with shorter length scales are said to capture more variation in the data and are utilized in analysis. Dimensions with non-zero precision (see Formula 3.10) are included in defining the dimensionality of the model.

ARD Squared Exponential Kernel

$$k(x, x') = \sigma^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_d - x'_d)^2}{L_d^2}\right) \tag{3.9}$$

Where L_d^2 is the length scale for each dimension.

Precision in Terms of Length Scale

$$\alpha = \frac{1}{L^2} \tag{3.10}$$

3.2 The Data

The data used in this paper comes exclusively from the Federal Reserve Economic Data (FRED) website (Federal Reserve Bank of St. Louis, 2024), an open source database of economic data maintained by the research department at the Federal Reserve Bank of St. Louis. The individual data series were concatenated using the Pandas library (Team, 2020), and each variable was standardized using the StandardScaler in scikit-learn (Pedregosa et al., 2011). Below is a list of all variables used in the Macro Model. Variables used in the Human Experience Model are marked with (*).

1. **Gross Domestic Product (GDP)*** (Billions of Chained 2017 Dollars): The inflation-adjusted value of the goods and services produced by labor and property located in the United States.
2. **Unemployment*** (%): The number of unemployed as a percentage of the labor force.
3. **Inflation*** (%): The annual percentage change in the cost to the average consumer of acquiring a basket of goods.
4. **Federal Funds Rate*** (%): The interest rate at which depository institutions trade federal funds (balances held at Federal Reserve Banks) with each other overnight—essentially affects all other interest rates.
5. **CO_2.** (Million Metric Tons): Volume of CO_2. produced in a given year.
6. **GINI index*** (Index): Measures the extent to which the distribution of income or consumption expenditure among individuals or households within an economy deviates from a perfectly equal distribution. A GINI index of 0 represents perfect equality, while an index of 100 implies perfect inequality.
7. **Life Expectancy*** (Years): Indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.
8. **Michigan Consumer Sentiment Index*** (The index for 1966 is taken to be 100): Measures consumers' outlook on the economy.
9. **Government Debt** (Billions of Dollars): Money the United States Government owes to its creditors.

10. **Government Deficit** (Percent Change): The difference between government revenue and government spending.
11. **Commodity Trade Balance** (Percent Change): Net dollar amount of commodities traded.
12. **Producer Price Index (PPI)** (The index for 1982 is taken to be 100): Measures the average change in selling prices received by domestic producers for their goods and services over time, reflecting inflation at the wholesale level before it reaches consumers.
13. **Industrial Production** (Percent Change): The industrial production (IP) index measures the real output of all relevant establishments located in the United States.
14. **Money Supply 2 (M2)** (Billions of 1982-84 Dollars): M2 consists of M1 plus (1) small-denomination time deposits (time deposits in amounts of less than $100,000); and (2) balances in retail MMFs. In essence, M2 is a measure of the money supply that includes cash, checking deposits, and easily convertible near money, such as savings deposits, money market securities, and small time deposits.
15. **Exchange Rate with Japan** (Japanese Yen to One U.S. Dollar): Noon buying rates in New York City for cable transfers payable in foreign currencies (Japan).
16. **Employment-to-population ratio** (%): The ratio of all employed in comparison to the population.
17. **Housing Starts** (Percent change): As provided by the Census, start occurs when excavation begins for the footings or foundation of a building.
18. **Corporate Profits after Tax** (Percent Change): Total corporate profits after tax.
19. **Labor Force Participation Rate** (%): The percentage of the population working or actively looking for work.
20. **Current Account: Balance (Revenue Minus Expenditure) for United States** (Percent of GDP): The current account balance measures the difference between the country's total exports and imports of goods and services, plus net income from abroad (such as dividends and interest payments) and net current transfers (such as remittances).
21. **Capital Expenditure** (Percent Change): Refers to the funds that a company or government spends on acquiring, maintaining, or improving long-term assets, such as property, buildings, equipment, or infrastructure, which are expected to provide benefits over a long period.
22. **Personal Consumption Expenditures** (The index for 2017 is taken to be 100): The Personal Consumption Expenditures (PCE) Price Index is a measure of the average change in prices paid by consumers for goods and services over time, often used by the Federal Reserve to assess inflation and adjust monetary policy.
23. **Gross Private Domestic Investment** (Percent change): Refers to the total spending by private entities within a country's borders on goods and services.
24. **Real Government Consumption Expenditures and Gross Investment** (Billions of Chained 2017 Dollars): Government consumption expenditures and gross investment.

25. **Real Disposable Personal Income** (Billions of Chained 2017 Dollars): Income that people get from wages and salaries, Social Security and other government benefits, dividends and interest, business ownership, and other sources.
26. **Non-Farm Pay Rolls** (Percent Change): A measure of the number of U.S. workers in the economy that excludes proprietors, private household employees, unpaid volunteers, farm employees, and the unincorporated self-employed.
27. **Personal Savings Rate** (%): Personal saving as a percentage of disposable personal income (DPI).
28. **Velocity of M2** (Billions of Chained 2017 Dollars): Calculated as the ratio of quarterly nominal GDP to the quarterly average of M2 money stock. The velocity of money is the frequency at which one unit of currency is used to purchase domestically-produced goods and services within a given time period.
29. **Age Dependency Ratio** (%): The ratio of older dependents (people older than 64) to the working-age population (those ages 15-64).

3.3 Kernel Choice

The squared-exponential kernel will be used because of its smoothness assumptions (Williams and Rasmussen, 2006). These smoothness assumptions lend itself to modelling Macroeconomic data that change gradually over time. On the other hand, Macroeconomic data usually exhibit linear trends due to the business cycle and general growth over time (Abel, Bernanke and Croushore, 2011). Therefore, BGPLVMs models will also be trained with an additive kernel comprised of a linear and a squared exponential component in order to account for linear trends as well.

3.4 Procedure

3.4.1 Assessing Model Fit

In order to assess the fit of the models, the reconstruction error, trustworthiness measure and continuity measure for the Macro Model with the squared exponential kernel, the Macro Model with the additive kernel. The metrics were also computed for the Human Experience Model with the squared exponential kernel and the Human Experience Model with the additive model were computed.

Trustworthiness Measure

$$T(k) = 1 - \frac{2}{n \times k \times (2n - 3k - 1)} \sum_{i=1}^{n} \sum_{j \in U_k(i)} (r(i,j) - k) \tag{3.11}$$

Continuity Measure

$$C(k) = 1 - \frac{2}{n \times k \times (2n - 3k - 1)} \sum_{i=1}^{n} \sum_{j \in V_k(i)} (r'(i,j) - k) \tag{3.12}$$

Explanation of Terms:

- $T(k)$ and $C(k)$ are the trustworthiness and continuity measures, respectively.
- n is the number of data points.
- k is the number of nearest neighbors considered.
- $r(i, j)$ is the rank of point j in the original space when sorted by distance from point i.
- $r'(i, j)$ is the rank of point j in the reduced space when sorted by distance from point i.
- $U_k(i)$ is the set of points that are in the top k nearest neighbors in the reduced space but not in the original space.
- $V_k(i)$ is the set of points that are in the top k nearest neighbors in the original space but not in the reduced space.

The trustworthiness measure evaluates the degree of which local relationships in the original high-dimensional space are preserved in the reduced-dimensional space while the continuity measure assesses how well the relationships in the reduced-dimensional space reflect those in the original data space. The trustworthiness measure (Venna and Kaski, 2001) (Figure 3.11) and the continuity measure Van Der Maaten et al. (2009) (Figure 3.12) are complimentary measures that are used in assessing non-linear dimensionality-reduction techniques. Trustworthiness and continuity measures closer to 1 are positive indications that local structures are preserved after reducing dimensionality. The trustworthiness and continuity measures were included in assessing the model, as reconstruction errors are not sufficient in analysing the fit of a non-linear dimensionality-reduction technique (Van Der Maaten et al., 2009). In fact, a model can have a high reconstruction error while preserving local relationships; the trustworthiness and continuity measures account for the preservation in local relationships.

Procedure:

1. Computed the reconstruction error for models by computing the mean squared error between the projections of the latent variables to the data space and the actual data.
2. Computed the trustworthiness and continuity measures of each model. Assessed the fit of models using the chosen metrics.
3. Compared and contrasted fit between models and in comparison to PCA.

3.4.2 Identifying Dimensionality

1. Plotted precisions (Figure 3.10) of length scales, given by the ARD kernel (Figure 3.9), for each Latent dimension by training a BGPLVM with relatively high dimensionality.
2. Identified latent dimensions with significant precisions to determine dimensionality.

3.4.3 Training and Analysis

1. Trained Bayesian GPLVMs based on dimensionality given by the ARD method (Figure 3.9).
2. Plotted latent variables and labelled them by their representative year.

3. Compared positions and separations in latent points between PCA, BGPLVM and BGPLVM with an additive kernel.
4. Analysed latent points and dimensions with macroeconomic reasoning; proposed potential names for axes with support from analysis of latent points.
5. Obtained variance of latent points and normalised them in order to provide relative uncertainties.
6. Plotted latent points with relative uncertainty representations.
7. Analysed the uncertainty of latent points based on economic history and context.

3.5 Years of Interest

The latent points represent years in the latent space. Therefore, it is useful to keep in mind some years in which the economic activity was of note in order to help evaluate and analyze the models.

3.5.1 Recessionary Years

1973-1974

In late 1973 the Organization of Arab Petroleum Exporting Countries (OAPEC) decided to stop exporting oil to the United States. This oil embargo had rippling effects in the economy, causing stagflation and the decrease in the quality of life of the average American (Blinder, 1982) (Abel, Bernanke and Croushore, 2011).

1982

The deep recession that occurred in 1982 was largely in result of the handling of the extreme inflation. In order to combat the inflation, the Federal Reserve had to enact policies to slow the economy down which in turn greatly increased unemployment and decreased economic growth (Taylor et al., 1992)(Abel, Bernanke and Croushore, 2011).

1991

The 1991 recession is commonly referred to as a mild recession. This recession occurred in part due to the Federal Reserve increasing interest rates to try and combat inflation, thereby slowing the economy down. The oil price shock due to the Gulf War also contributed to the recession (Walsh, 1993). This recession was characterised by high unemployment and a slow down in the housing market.

2009

Known as the Great Recession. Thousands of people lost their homes and their jobs due to mortgage back securities failing (Abel, Bernanke and Croushore, 2011).

2020

The COVID pandemic caused a generational recession. Millions of people lost their jobs causing a massive decrease in consumption which then caused thousands of businesses to suffer.

3.6 Software Used

The Pandas (Team, 2020) library was used to concatenate data the individual data series from FRED for the data set. The GPflow (Matthews et al., 2017) and TensorFlow(Abadi et al., 2015) libraries were used to train the models. Meanwhile, the Matplotlib (Hunter, 2007) library was used to generate plots for the outputs of the models. NumPy (Harris et al., 2020) was used to manipulate vectors and scikit-learn (Pedregosa et al., 2011) was used to standardize features in the data set.

Chapter 4

Results and Analysis

4.1 Assessing Model Fit

Models	RE	RE Per Feature	Trust.	Cont.
PCA Macro Data	0.28	0.0096	0.98	0.99
Macro Model	0.2	0.0068	0.97	0.98
Macro Model *	0.28	0.0097	0.97	0.98
PCA Human Data	0.12	0.015	0.98	0.98
Human Experience	0.063	0.0079	0.98	0.98
Human Experience *	0.14	0.017	0.98	0.98

Table 4.1: Metrics to Assess the Fit of Models

- RE: reconstruction error
- Trust.:trustworthiness measure
- Cont.: continuity measure
- Macro Model *: the Macro Model with the additive
- Human Experience *: the Human Experience Model with the additive kernel

For all six models assessed, the trustworthiness and continuity measures are all approaching 1; therefore the models appear to capture the local structures of the data well. However, in relation to the reconstruction errors of the models, there are variations. For both the Macro Model and Human Experience Model with the squared-exponential kernel, the reconstruction errors are lower than their PCA and additive kernel counterparts. In contrast, the reconstruction errors for the Macro Model and Human Experience Model with the additive kernel are both higher than or the same as their PCA and squared-exponential kernel counter-parts. Therefore, the Human Experience Model and the Macro Model, both utilizing the additive kernel, may be prone to overfitting. Consequently, predictions and analyses derived from these models should be interpreted with caution. Additionally, the models trained on the macroeconomic data have noticeably lower reconstruction errors per feature—this may be due to the fact that the models have more data from features to train on and hence might be more informative.

4.2 The Macro Model

4.2.1 PCA Versus Squared Exponential BGPLVM

As per the ARD squared-exponential kernel, the Macro Model is inherently three-dimensional.

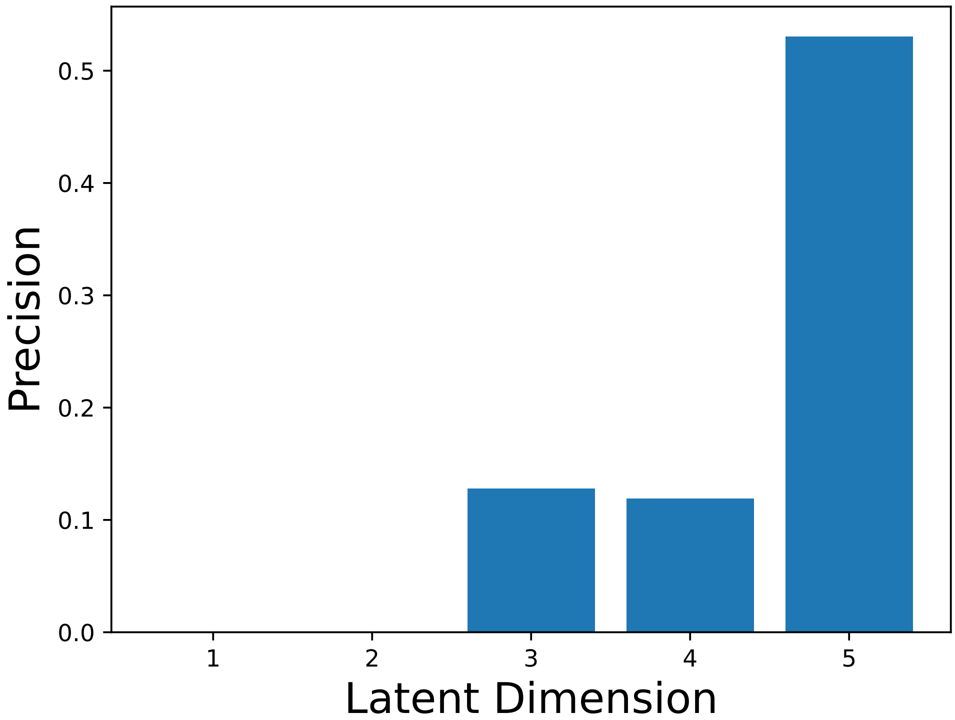

Figure 4.1: Precisions of Latent Dimensions for the Macro Model

Figure 4.1 shows that the fifth latent dimension accounts for the largest amount of variation in the data, while latent dimensions three and four account for less variation in the data.

Although the ARD kernel suggests that the data is three-dimensional, plotting the data using two-dimensions could still reveal valuable information. The two dimensional representation of the data is plotted using the two latent dimensions with the largest precisions out of the three.

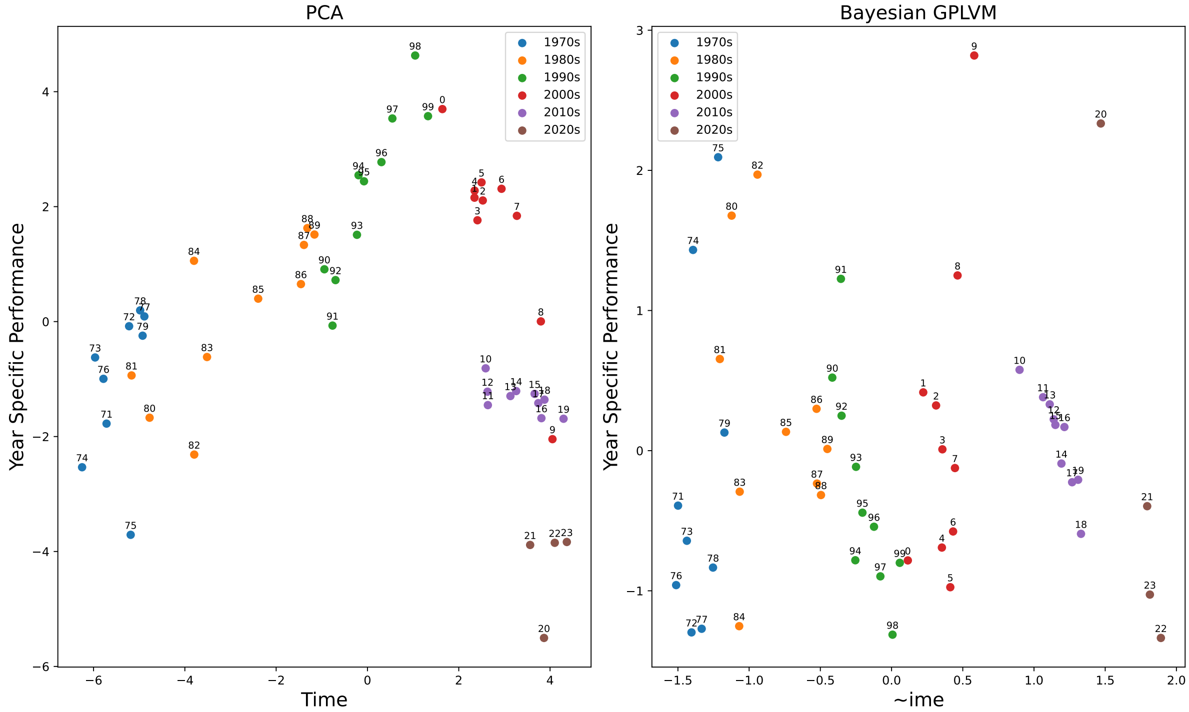

Figure 4.2: PCA VS Bayesian GPLVM (Two-Dimensional Economic Representation of Years)

Upon examining the two-dimensional BGPLVM representation of the Macro Model in Figure 4.2, the decades appear to be regularly separated along the X-axis. Therefore, it was proposed that the X-axis captures temporal changes inferred by the model, even though time was not explicitly present in the data. Furthermore, it appears as though the Y-axis accounts for the macroeconomic performance of years within each of their respective decades. For instance, within the 2020s, the Macro Model effectively recognizes that in 2020 the economy basically came to a complete halt, whereas in 2021, 2022 and 2023 the economy was in a much better place due to the opening up of the economy. However, the Macro Model was also able to account for a more subtle recession by separating 1991 from other years in the 1990s. The Macro Model also places the best performing years at the opposite end of the Y-axis. For instance, 1998 is placed solidly at the bottom of the Y-axis in comparison to the rest of the 1990s. In 1998, unemployment was at a 28-year low while inflation was at the lowest it had been in 32 years and consumer confidence was near the highest it had even been in the previous 30 years (The White House, 1998). Furthermore, the Macro Model is able to make a clear separation between the recessionary years mentioned in the methods section and relatively stable years, with a gap of approximately half a unit at $y = 1$ separating the latent points. Therefore, it was proposed that the Y-axis represents relative yearly macroeconomic performance.

Although the two-dimensional PCA representation of the macroeconomic data is able to somewhat separate the data into decades and differentiate the macroeconomic performance of years within the decades, the two-dimensional BGPLVM representation of the macroeconomic data is able to better differentiate between decades. Figure 4.2 shows that standard PCA is unable to handle the non-linearities in the data—exhibiting a parabolic shape while the latent variables in the BGPLVM model are more uniformly distributed. The BGPLVM is also appears to be able to make a distinction between the turbulent economic years and more stable years whereas the PCA representation did not. Furthermore, due to the BGPLVM's ability to handle non-linearities in the data, the BGPLVM is able to somewhat account for the severity in recessions between decades. For instance, the BGPLVM positions 1991, a relatively mild recession, lower in relation to the Great Recession of 2009 and the economic crisis during the COVID pandemic.

Figure 4.3 shows the level of uncertainty in the two-dimensional latent points from the Macro Model. The BGPLVM reconciles uncertainty based on how different the data for a latent point is to other data points. Therefore, if the data for a latent point is very inconsistent with the majority of latent points, the relative uncertainty of that latent point will be higher than the rest. This feature of the BGPLVM, and hence the Macro Model, can serve as a proxy for uncertainty in the economic behaviour of a time period due to accounting for irregularities in macroeconomic data.

It appears as though the Macro Model does not change uncertainty estimates along the proposed time axis; an accurate representation of reality since change in time should not inherently affect uncertainty and values in the data. In contrast, the Macro Model adjusts uncertainty levels along the year specific performance axis. The change in uncertainty along the year specific performance axis is consistent with reality: the uncertainty level increases in the latent space surrounding the recessionary years, years during which the data is irregular and similar data is scarce.

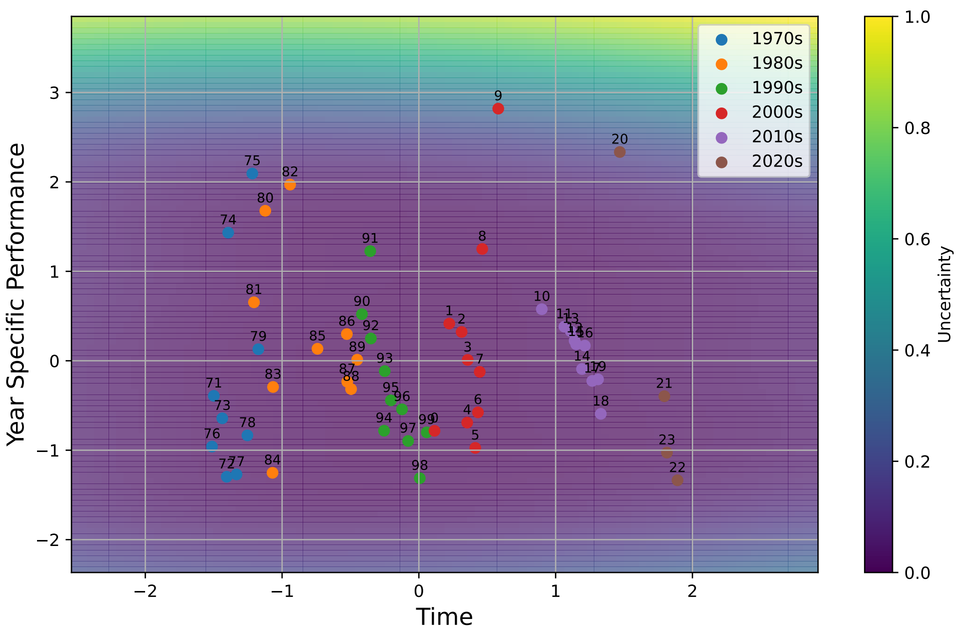

Figure 4.3: Bayesian GPLVM (Two Dimensional Economic Representation of Years)

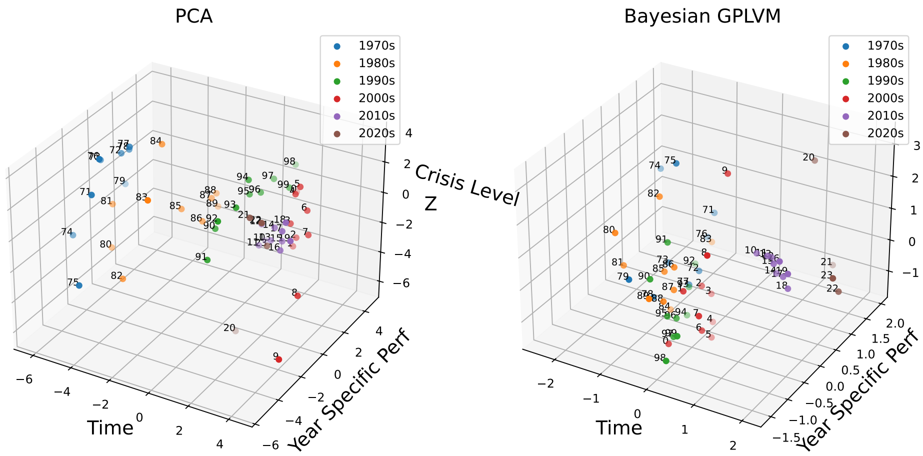

Figure 4.4: PCA VS Bayesian GPLVM (Three Dimensional Economic Representation of Years)

Figure 4.4 further demonstrates the superiority of the BGPLVM in relation to PCA. When plotting for three latent dimensions, the BGPLVM appears to be able to differentiate between crises level more effectively than before due to the addition of the third latent dimension. For example, the year 1991 is placed low on the proposed crisis level axis in relation to years like 2009 and 2020—both of which represent more extreme recessions, while still accurately being placed as a tough year within its decade. Meanwhile, the Z-axis in the PCA plot does not extract as much meaningful information from the data , as most (if not all) recession years are placed on roughly the same level.

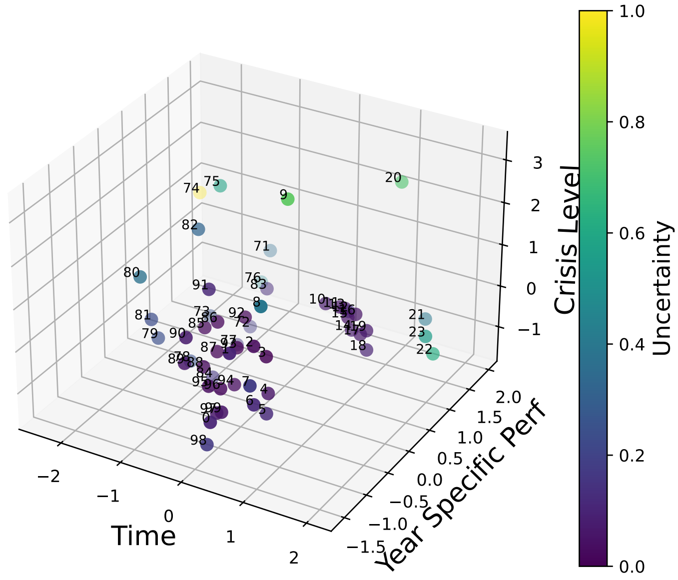

Figure 4.5: BGPLVM: Three Dimensional Economic Representation of Years

When accounting for uncertainty in the latent variables, a curious disparity emerges between the two-dimensional and three-dimensional representations of the latent points from the BGPLVM. The two-dimensional representation (figure 4.3) suggests that the data for 2009 exhibits more uncertainty than that of 1974, while the three-dimensional representation of the data suggests otherwise. Given that the data is three-dimensional according to the ARD mechanism, the third dimension could be adding nuance, which suggests that 1974 was indeed more uncertain of a year than 2009. This may run counter to common opinion, because 2009 was an incredibly difficult and unpredictable year for millions of people across the globe, but recency bias may be in effect. It is important to remember that several significant issues were being faced in 1974. There was an oil embargo that lead to lines at the petrol stations across the United States. The U.S. government was also dealing with massive deficits from the Vietnam War and stagflation was prominent. On the political front, President Nixon faced an insurmountable political battle that eventually lead to his resignation later that year. Furthermore, 1973 was when the Breton-Woods system (gold standard) was fully dissolved (International Monetary Fund, 2023) (Hammes and Wills, 2005), which lead to more unpredictability in the exchange rates of currencies. Therefore, the Macro Model giving 1974 the highest level of relative uncertainty is reasonable, since the data for 1974 is very irregular in comparison to the majority of other years.

4.2.2 Comparing the Squared-Exponential and Additive Kernel

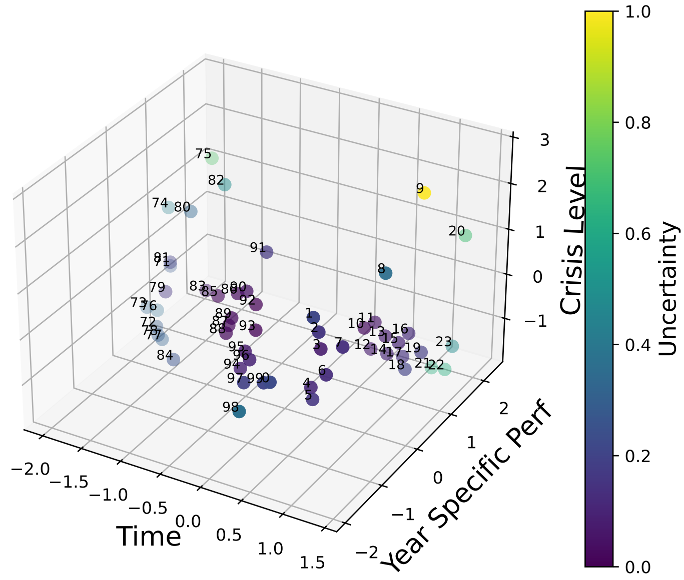

Figure 4.6: BGPLVM with Additive Kernel (Three Dimensional Economic Representation of Years)

The Macro Model better when the additive kernel comprised of a linear kernel (see Formula 3.2) and the squared exponential kernel (see Formula 3.1)is used. In comparing Figures 4.5 and 4.6, the additive kernel appears to separate decades more effectively and uniformly. Furthermore, the addition of the linear component seems to assist the crisis level latent dimension in making meaningful separations. For instance, the Macro Model without the additive kernel, represented by Figure 4.5, is unable to differentiate in crisis level between 1991 and non-recessionary years. Meanwhile, the Macro Model with the additive kernel, represented by figure 4.6, makes a meaningful distinction between non-recessionary years and 1991.

Furthermore, the Macro Model with the additive kernel now presents 2009 as the year with the most relative uncertainty instead of 1974. This is also logically plausible, because data similar to the Great Recession is almost non-existent. What made 2009 stand out in comparison to other periods in time was the unprecedented increase in unemployment and sharp decrease in housing prices—the main source of wealth for most people.

4.3 The Human Experience Model

Although the Human Experience Model will account for less features, the interpretations and outputs of the model are likely to be more nuanced. The aim is to acknowledge variables that capture human experiences and not just solely focus on macroeconomic results. Therefore, relying on the historical economic analysis of previous years in terms of recessionary behaviour will not prove to be as fruitful, which making analysis more difficult. However, that is precisely the point of this exercise: to break out of the one-dimensional way of thinking about the economy.

4.3.1 PCA Versus Squared Exponential BGPLVM

Similarly to the Macro Model, the inherent dimensionality of the data for the Human Experience Model is three-dimensional.

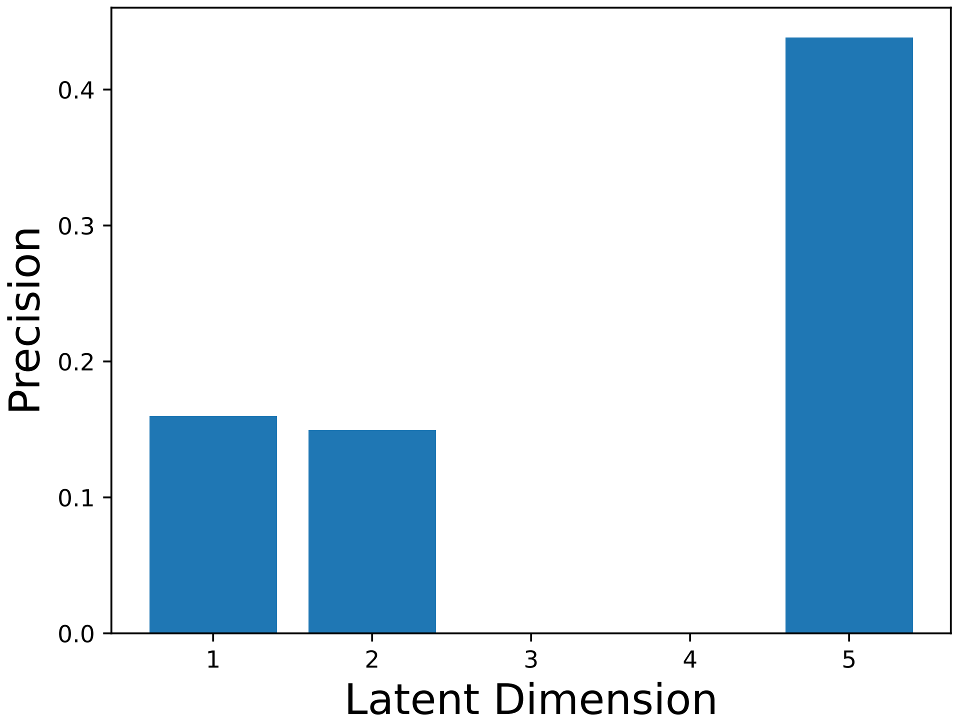

Figure 4.7: Precisions of Latent Dimensions for the Human Experience Model

Figure 4.7, indicates that the fifth latent dimension captures the most amount of variation in the data, along with the first and second latent dimensions both of which capture a significant amount of variation.

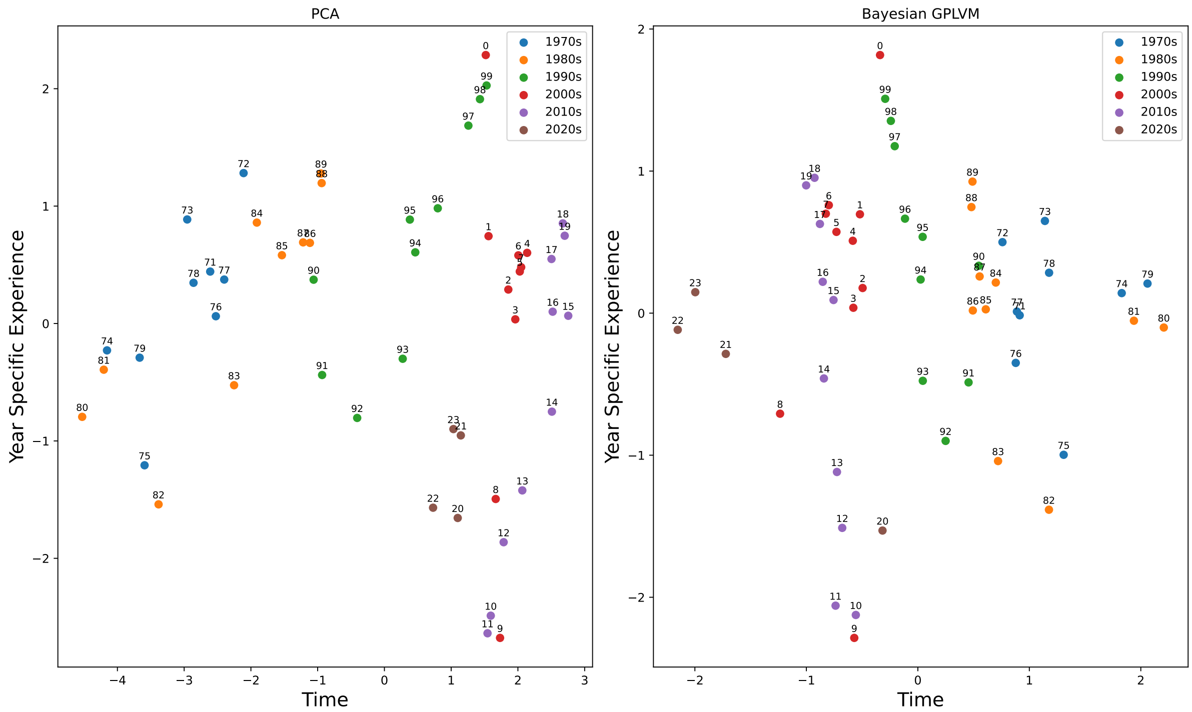

Figure 4.8: PCA VS Bayesian GPLVM(Two Dimensional Economic Representation of Years—the Human Experience Model)

The nuance of the Human Experience Model in comparison to the Macro Model is evident when comparing Figures 4.8 and 4.2. Although the BGPLVM for the Human Experience Model performs better than its PCA counter-part in accounting for variations in the latent space due to time, the BGPLVM for the Human Experience Model does not account for time as accurately as in the Macro Model counterpart. Nevertheless, the Human Experience Model offers potential insights. It is assumed, given the variables that the Human Experience Model is trained on, that the Y-axis should somewhat be able to account for human experiences. Evidence suggests that this may be the case.

The Human Experience Model is able to accurately account for the tough experiences that humanity faced in the 2008-2009 Great Recession and 2020 COVID pandemic along the y-axis. The model is able to uncover an observable improvement in human experience following the 2009 Great Recession, with a linear trend emerging from 2009 through 2019. However, the Human Experience Model captures two interesting characteristics. First, the Human Experience Model seems to place years behind or ahead in time accurately based on the human experience of the time. For instance, the latent points for 1980 and 1981 are placed back in time as if they had lost the better part of a decade which might have been anecdotally true considering the recessionary experiences during this time. Also, the latent point for 2008 appeared to have jumped forward in time almost as if capturing the bubble-like behavior that lead to the Great Recession. Second, although the COVID pandemic was a catastrophic experience for humanity, the latent points for the years after 2020 appear to rebound according to the Human Experience Model which may be due to the decrease in CO_2 emissions during this period, a sharp increase in consumer sentiment, and general recovery in the economy.

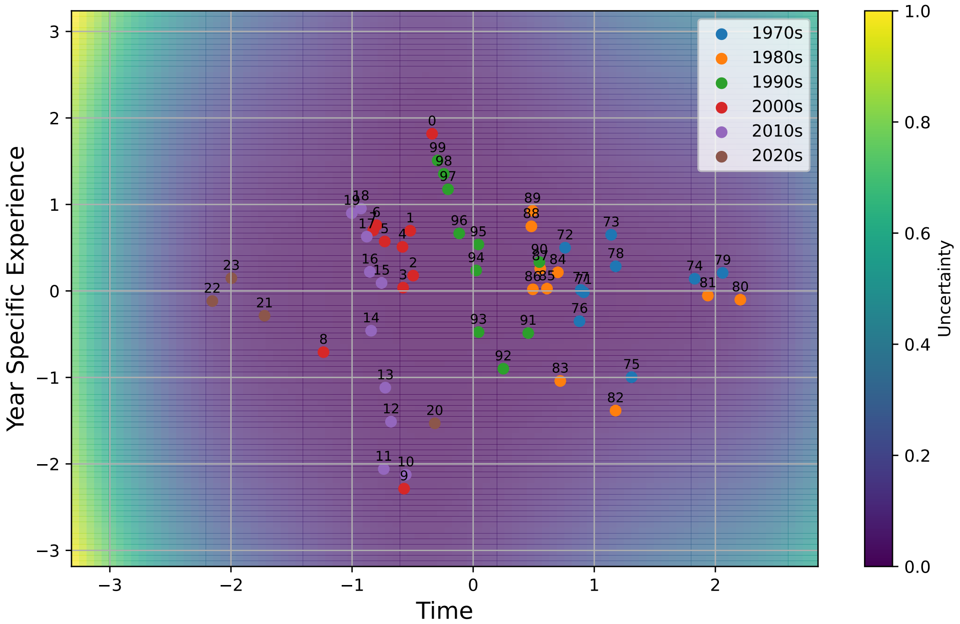

Figure 4.9: Bayesian GPLVM (Two Dimensional Economic Representation of Years—the Human Experience Model)

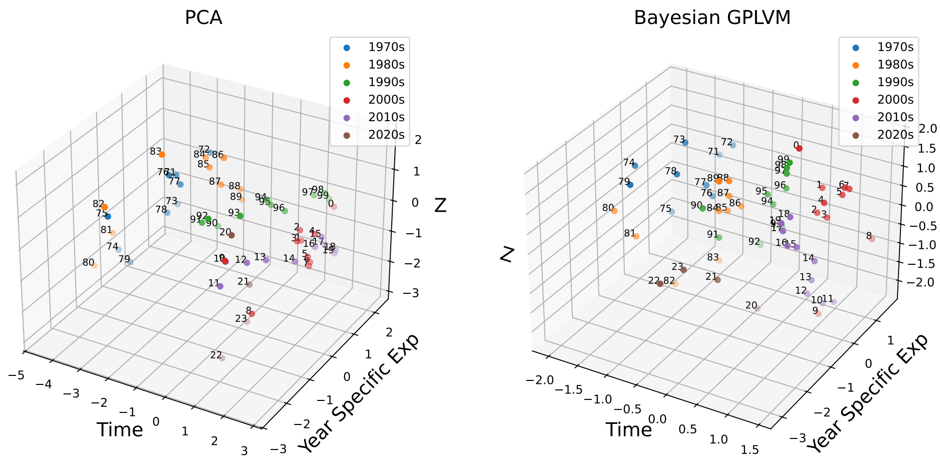

Figure 4.10: PCA VS Bayesian GPLVM (Three Dimensional Economic Representation of Years—the Human Experience Model)

In contrast to Figure 4.3, the Human Experience Model seems to be affected by both time- and year-specific events, as seen in Figure 4.9. In particular, figure 4.9 places the COVID years in the direction of most uncertainty, which makes sense, since the data collected during the 2020 COVID pandemic is incredibly irregular in contrast to other years.

Figure 4.10 once again demonstrates that Human Model is more nuanced than the Macro Model. Remarkably, the three-dimensional Human Experience Model places the COVID years in a similar position to the 1980s, which almost indicates that the Human Experience level regressed four decades. Moreover, at first glance, it seems as though the *z*-axis for Figure 4.10 accounts for a crisis level; however, this is not the case. Although 1974 is placed correctly within its decade in terms of year-specific performance, it is placed on the opposite side of the *z*-axis in comparison to other recessionary years. Therefore, although the recessions during 1974, 1991 and 2020 were very painful for humanity, clearly there was a difference in terms of human experience. This further emphasises the need for economists and policy-makers to look at data in a holistic approach as apposed to relying on a few macroeconomic variables.

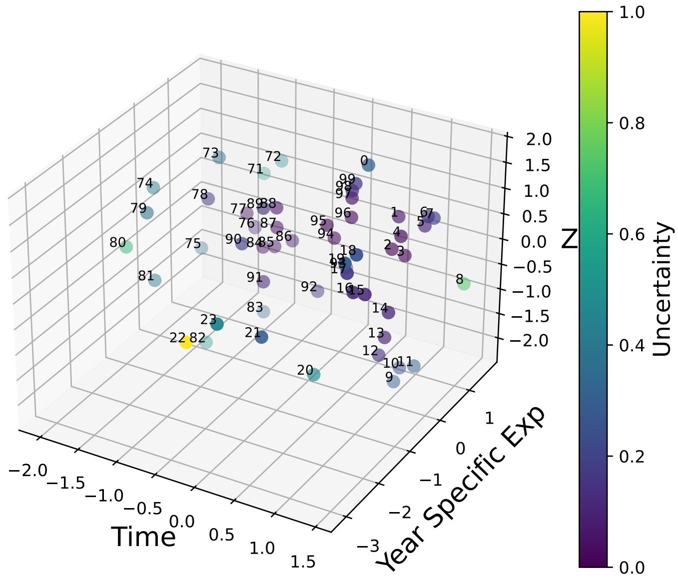

Figure 4.11: BGPLVM: (Three Dimensional Economic Representation of Years—the Human Experience Model)

As seen in Figure 4.11, the three-dimensional Human Experience Model accurately places higher levels of relative uncertainty around recessionary years. However, the relative uncertainty of 2022, in the context of the Human Experience model, is the highest as opposed to a recessionary year. Some possible explanations for 2022 having the highest level of relative uncertainty is that the economy was still experiencing a rapid recovery from 2020. Furthermore, consumer sentiment was the lowest it had ever been in the metric's recorded history, while interest rates were raised to the highest levels since the turn of the century, and unemployment was at its lowest level since 1969. The combination of unemployment being historically low while consumer sentiment being historically low is paradoxical (and certainly very irregular) based on the other observations. Therefore, it is unsurprising that 2022 was the year with the highest level of relative uncertainty.

4.4 Comparing the Squared-Exponential and Additive Kernel

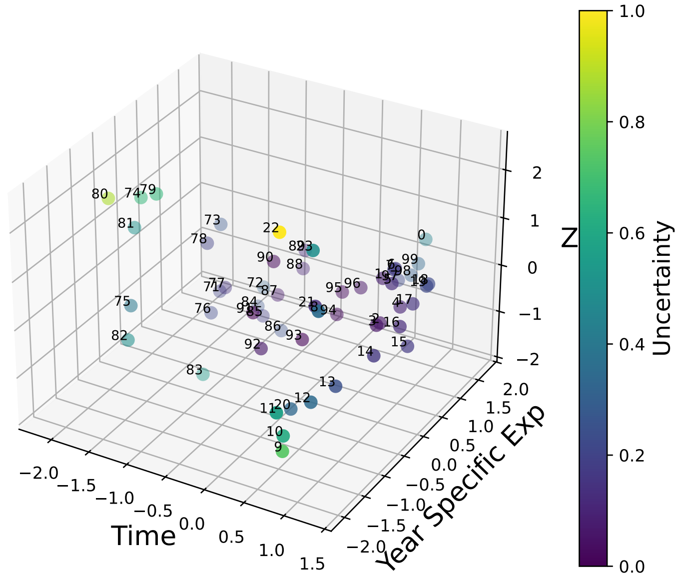

Figure 4.12: BGPLVM with Additive Kernel (Three-Dimensional Economic Representation of Years—the Human Experience Model)

The use of the additive kernel, as seen in Figure 4.12, provided better separation and uniform placements between the latent variables. The addition of the linear kernel enabled the Human Experience model to show more distinctly the linear progression in human experience after the Great Recession in 2008-2009 as the economy recovered. This feature in the Human Experience Model, with the additive kernel, is not easily detected in the Human Experience Model with the squared-exponential kernel (Figure 4.11). Furthermore, in the Macro Model (figure 4.6, this feature is not detected at all.

The most glaring difference between the Human Experience Model with the squared-exponential kernel (Figure 4.11) and the additive kernel (Figure 4.12) is the opposite positioning of 2022 and 2023 latent points along the year specific experience axis. According to the squared-exponential kernel, 2022 and 2023 were tough years for the human experience. On the other hand, the additive kernel positions the latent points for 2022 and 2023 relatively close to good years for human experience. Furthermore, the Human Experience Model with the additive kernel exhibited the highest relative uncertainty for the latent points of 2022 and 2023. This finding aligns with expectations, given that numerous metrics during these years demonstrated atypical patterns, likely due to the COVID pandemic recovery phase.

Chapter 5

Discussions and Conclusions

5.1 Critical Evaluation

The Macro Model and the Human Experience both provide meaningful economic representations of years in their respective contexts. The Macro Model is able to account approximately for time, year-specific economic performance, and the level of crisis in the latent space. Furthermore, the Macro Model gives meaningful relative estimates of uncertainty in the latent points. Likewise, the Human Experience Model is able to paint effectively a different picture of the economy from the average person's point of view while reconciling relative uncertainty estimates with reality. Therefore, the BGPLVM can be used as a tool by economists to diagnose economic activity with more nuance and from multiple perspectives based on the initial features with which the model is trained with; this could enable economists to focus on developing specific policies for specific scenarios. A systematic way in which economists can utilize BGPLVM models is by standardising the variables of interest used in training the models and studying the latent positioning of years and what they usually represent. In essence, a quadrant-like system can be developed for economic diagnosis. Furthermore, the components produced by the BGPLVM can be utilized, along with set thresholds, to classify economic time periods based on what the latent dimension accounts for.

However, the use of the BGPLVM also has potential drawbacks. First, it is computationally expensive because of the need to invert covariance matrices. Thankfully, in the context of economic activity, the number of observations is usually defined by periods in time. Therefore, the number of observations is comparably smaller to other contexts such as in the analysis of cells. Furthermore, the BGPLVM model utilizes inducing variables to reduce the time in training the models while accurately summarising the data. However, the inducing variables present another challenge; the BGPLVM can be sensitive to hyper-parameters. For instance, the number of inducing variables chosen for training the model can have a great effect on the accuracy of the model, and might even lead to the computer's kernel crashing during training. The results of the BGPLVM are also highly dependent on the choice of the Kernel that is used, as seen in Figures 4.11 and 4.12. Moreover, while BGPLVM provides a flexible latent space representation, it can be difficult to directly identify how specific features contribute to these components. In contrast, PCA offers a more interpretable model, where the loading scores of each principal component reveal the relative importance of different features in shaping the latent representation.

5.2 Future Work

The data for this paper is annual, but in reality, there is often a large variation in economic activity within years. Hence, experimenting with smaller time periods would be useful. Also, experimenting with other kernels and kernel structures could yield interesting results. For instance, instead of using an additive kernel, a product kernel could be used to account for different structures in the data. In addition, the use of a periodic kernel (see Figure 3.3 could be promising, because many macroeconomic variables are periodic in nature due to the business cycle (Abel, Bernanke and Croushore, 2011). The incorporation of the Matern kernel might also garner interesting outcomes, because the Matern Kernel allows for control over the level of "smoothness" in the mappings from the latent points to the data (Williams and Rasmussen, 2006). The choice of initial feature selection is also a subject of interest. As previously seen in the contrasting results from the Human Experience and Macro Models, the choice of features in the data causes a major difference in the position of the latent points. Therefore, accounting for different variables in the models would be beneficial, especially when looking at the economy from different perspectives. Accounting for features that represent the experience of people, instead of features of concern to corporations and governments, such as health, happiness, and quality of life are of pressing concern and should be studied in the future and the Bayesian GPLVM can help in this pursuit.

Word Count: 9883

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. and Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems [Online]. Software available from tensorflow.org. Available from: `https://www.tensorflow.org/`.

Abel, A.B., Bernanke, B.S. and Croushore, D., 2011. Macroeconomics, 7th.

Ames, Brown, D. and Izquierdo, 2020. *Macroeconomic policies and poverty reduction* [Online]. International Monetary Fund. Available from: `https://www.imf.org/external/pubs/ft/exrp/macropol/eng/`.

Athey, S., Blei, D., Donnelly, R., Ruiz, F. and Schmidt, T., 2018. Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data. *Aea papers and proceedings*. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 108, pp.64–67.

Barbarino, A. and Bura, E., 2015. *Forecasting with sufficient dimension reductions* [Online]. (Finance and Economics Discussion Series, 2015-074). Board of Governors of the Federal Reserve System. Available from: `https://doi.org/10.17016/FEDS.2015.074`.

Berge, T.J. and Jordà, Ò., 2011. Evaluating the classification of economic activity into recessions and expansions. *American economic journal: Macroeconomics*, 3(2), pp.246–277.

Blei, D.M., Kucukelbir, A. and McAuliffe, J.D., 2017. Variational inference: A review for statisticians. *Journal of the american statistical association*, 112(518), pp.859–877.

Blinder, A.S., 1982. 12 the anatomy of double-digit inflation in the 1970s. *Inflation: Causes and effects*, 261.

Duvenaud, D., 2014. Automatic model construction with gaussian processes.

Federal Reserve Bank of St. Louis, 2024. FRED, Federal Reserve Economic Data. `https://fred.stlouisfed.org/`. `https://fred.stlouisfed.org/`.

Geiger, A., Urtasun, R. and Darrell, T., 2009. Rank priors for continuous non-linear dimensionality reduction. *2009 ieee conference on computer vision and pattern recognition*. IEEE, pp.880–887.

Geweke, J., Koop, G. and Dijk, H.K. van, 2011. *The oxford handbook of bayesian econometrics*. Oxford University Press, USA.

Ghahramani, Z., 2003. Unsupervised learning. *Summer school on machine learning*. Springer, pp.72–112.

Goulet Coulombe, P., Leroux, M., Stevanovic, D. and Surprenant, S., 2022. How is machine learning useful for macroeconomic forecasting? *Journal of applied econometrics*, 37(5), pp.920–964.

Guo, Y., 2023. *Three essays in macroeconomic forecasting using dimensionality reduction methods*. Ph.D. thesis. University of Glasgow.

Hammes, D. and Wills, D., 2005. Black gold: The end of bretton woods and the oil-price shocks of the 1970s. *The independent review*, 9(4), pp.501–511.

Harris, C.R., Millman, K.J., Walt, S.J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M.H. van, Brett, M., Haldane, A., Río, J.F. del, Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. and Oliphant, T.E., 2020. Array programming with NumPy. *Nature* [Online], 585(7825), pp.357–362. Available from: `https://doi.org/10.1038/s41586-020-2649-2`.

Hauzenberger, N., Huber, F., Marcellino, M. and Petz, N., 2024. Gaussian process vector autoregressions and macroeconomic uncertainty. *Journal of business & economic statistics*, (just-accepted), pp.1–27.

Hoberg, G. and Phillips, G., 2016. Text-based network industries and endogenous product differentiation. *Journal of political economy*, 124(5), pp.1423–1465.

Hunter, J.D., 2007. Matplotlib: A 2d graphics environment. *Computing in science & engineering* [Online], 9(3), pp.90–95. Available from: `https://doi.org/10.1109/MCSE.2007.55`.

International Monetary Fund, 2023. End of bretton woods system [Online]. Accessed: 2024-08-17. Available from: `https://www.imf.org/external/about/histend.htm#:~:text=of%20the%20dollar.-,End%20of%20Bretton%20Woods%20system,the%20dollar's%20convertibility%20into%20gold`.

Issah, M. and Antwi, S., 2017. Role of macroeconomic variables on firms' performance: Evidence from the uk. *Cogent economics & finance*, 5(1), p.1405581.

Iyetomi, H., Aoyama, H., Fujiwara, Y., Souma, W., Vodenska, I. and Yoshikawa, H., 2020. Relationship between macroeconomic indicators and economic cycles in us. *Scientific reports*, 10(1), p.8420.

Jackson, T., 2009. *Prosperity without growth?: The transition to a sustainable economy*. Sustainable Development Commission.

Jia, W., Sun, M., Lian, J. and Hou, S., 2022. Feature dimensionality reduction: a review. *Complex & intelligent systems*, 8(3), pp.2663–2693.

John, C., Ekpenyong, E.J. and Nworu, C.C., 2019. Imputation of missing values in economic and financial time series data using five principal component analysis approaches. *Cbn journal of applied statistics (jas)*, 10(1), p.3.

Jolliffe, I.T. et al., 1986a. Principal component analysis. *Principal component analysis*, 372.

Jolliffe, I.T. et al., 1986b. Principal component analysis and factor analysis. *Principal component analysis*, 372, pp.115–128.

Jordan, M.I., Ghahramani, Z., Jaakkola, T.S. and Saul, L.K., 1999. An introduction to variational methods for graphical models. *Machine learning*, 37, pp.183–233.

King, R.G., Plosser, C.I. and Rebelo, S.T., 1988. Production, growth and business cycles: I. the basic neoclassical model. *Journal of monetary economics*, 21(2-3), pp.195–232.

Krugman, P. and Wells, R., 2020. *Microeconomics*. 5th ed. New York, NY: Worth Publishers.

Lawley, D.N. and Maxwell, A.E., 1962. Factor analysis as a statistical method. *Journal of the royal statistical society. series d (the statistician)*, 12(3), pp.209–229.

Lawrence, N., 2003. Gaussian process latent variable models for visualisation of high dimensional data. *Advances in neural information processing systems*, 16.

Lawrence, N., 2005. Probabilistic non-linear principal component analysis with gaussian process latent variable models. *Journal of machine learning research* [Online], 6(60), pp.1783–1816. Available from: `http://jmlr.org/papers/v6/lawrence05a.html`.

Lawrence, N., 2014a. Deep gps. Lecture presented at the GPRS in Pereira, Colombia.

Lawrence, N., 2014b. Unsupervised learning with gaussian processes. Lecture presented at GPWS, University of Sheffield.

Matthews, A.G.d.G., van der Wilk, M., Nickson, T., Fujii, K., Boukouvalas, A., León-Villagrá, P., Ghahramani, Z. and Hensman, J., 2017. GPflow: A Gaussian process library using TensorFlow. *Journal of machine learning research* [Online], 18(40), pp.1–6. Available from: `http://jmlr.org/papers/v18/16-537.html`.

Nirwan, R.S., 2020. *Bayesian machine learning for financial modeling*. Ph.D. thesis. Universitätsbibliothek Johann Christian Senckenberg.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12, pp.2825–2830.

Ruiz, F.J., Athey, S. and Blei, D.M., 2020. Shopper. *The annals of applied statistics*, 14(1), pp.1–27.

Spearman, C., 1904. The proof and measurement of association between two things. *The american journal of psychology*, 15(1), pp.72–101.

Stiglitz, J.E., Sen, A. and Fitoussi, J.P., 2009. *Report by the commission on the measurement of economic performance and social progress*. Commission on the Measurement of Economic Performance and Social Progress.

Storm, H., Baylis, K. and Heckelei, T., 2020. Machine learning in agricultural and applied economics. *European review of agricultural economics*, 47(3), pp.849–892.

Taylor, J. et al., 1992. The great inflation, the great disinflation, and policies for future price stability| conference–1992. *Rba annual conference*.

Team, T.P.D., 2020. *Pandas: Powerful data structures for data analysis, time series, and statistics*. Available from: `https://pandas.pydata.org/`.

The White House, 1998. The clinton presidency: A historic era of progress and prosperity. `https://clintonwhitehouse4.archives.gov/textonly/WH/Work/073198.html`. Accessed: 2024-08-28.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the royal statistical society series b: Statistical methodology*, 58(1), pp.267–288.

Tipping, M.E. and Bishop, C.M., 1999. Probabilistic principal component analysis. *Journal of the royal statistical society series b: Statistical methodology*, 61(3), pp.611–622.

Titsias, M., 2009. Variational learning of inducing variables in sparse gaussian processes. *Artificial intelligence and statistics*. PMLR, pp.567–574.

Titsias, M. and Lawrence, N.D., 2010. Bayesian gaussian process latent variable model. *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, pp.844–851.

University at Buffalo, 2024. What is economics? [Online]. Accessed on August 3, 2024. Available from: `https://arts-sciences.buffalo.edu/economics/about/what-is-economics.html` [Accessed 2024-08-03].

University of Michigan, 2024. Michigan consumer sentiment index. Available from `https://data.sca.isr.umich.edu/`.

Van Der Maaten, L., Postma, E., Herik, J. Van den et al., 2009. Dimensionality reduction: a comparative. *J mach learn res*, 10(66-71).

Venna, J. and Kaski, S., 2001. Neighborhood preservation in nonlinear projection methods: An experimental study. *International conference on artificial neural networks*. Springer, pp.485–491.

Verleysen, M. and François, D., 2005. The curse of dimensionality in data mining and time series prediction. *International work-conference on artificial neural networks*. Springer, pp.758–770.

Walsh, C.E., 1993. What caused the 1990-1991 recession? *Federal reserve bank of san francisco economic review*, 1993(2), pp.33–48.

Williams, C. and Rasmussen, C., 1995. Gaussian processes for regression. *Advances in neural information processing systems*, 8.

Williams, C.K. and Rasmussen, C.E., 2006. *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA.