FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

TREND AND VARIABLE-PHASE SEASONALITY ESTIMATION FROM FUNCTIONAL DATA

By

LIANG-HSUAN TAI

A Dissertation submitted to the
Department of Mathematics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2017

Liang-Hsuan Tai defended this dissertation on April 11, 2017.
The members of the supervisory committee were:

Kyle A. Gallivan

Professor Co-Directing Dissertation

Anuj Srivastava

Professor Co-Directing Dissertation

Wei Wu

University Representative

Eric P. Klassen

Committee Member

Giray Ökten

Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.

This work is dedicated to my mother and father, for their constant encouragement.

ACKNOWLEDGMENTS

I would like to first express the deepest gratitude to my advisers, Dr. Kyle Gallivan and Dr. Anuj Srivastava, for all their continual support and help to make this dissertation possible. Dr. Gallivan has been a wonderful mentor. His knowledge and humbleness affects me deeply. The training from his computational mathematics, numerical linear algebra, and Riemannian optimization classes were the foundations of my research. His valuable advisement was essential to the completion of this dissertation.

Meanwhile, Dr. Srivastava is an amazing teacher. His knowledge and hard working attitude inspire me. He taught me functional data analysis, which later on became my major research area. My thesis topic, trend and seasonality estimation, was given to me by him three years ago. In the beginning, I was completely unsure about how to approach this topic since there were so many open questions. Every time I met with Dr. Srivastava, he guided me and suggested possible directions I could take. By the time I completed my dissertation, many of the questions were answered and my research helped expand this topic.

I also want to give sincere thanks to my dissertation committee members, Dr. Eric Klassen, Dr. Wei Wu and Dr. Giray Ökten. Dr. Klassen is one of the greatest mathematicians I have ever met. He always asks thoughtful questions during my talk, waits for my answer, and corrects errors in a very polite way. The computational statistics class taught by Dr. Wu was extremely useful. I applied all the tools I learned to my research. I would also like to thank Dr. Ökten for providing helpful comments regarding the application of my dissertation.

I would like to acknowledge the following professors from my undergraduate study who were so kind to write letters of recommendation for me: Dr. Jhi-Shen Tsay, Dr. Huei-Mei Liang, and Dr. Hongkun Xu. Without their support, I would not have been enrolled at Florida State University.

Many of my friends helped me went through the difficult years on my graduate study. Their friendship and support encouraged me to continue my graduate studies and complete my dissertation.

Finally and most importantly, none of this dissertation would have been possible without the support of my mother and father. Their constant encouragement made my dream of pursuing a Ph.D. in mathematics possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

ABSTRACT

The problem of estimating trend and seasonality has been studied over several decades, although mostly using single time series setup. This dissertation studies the problem of estimating these components from a functional data point of view, i.e. multiple curves, in situations where seasonal effects exhibit arbitrary time warpings or phase variability across different observations. Rather than ignoring the phase variability, or using an off-the-shelf alignment method to remove phase, we take a model-based approach and seek Maximum Likelihood Estimators (MLEs) of the trend and the seasonal effects, while performing alignments over the seasonal effects at the same time. The MLEs of trend, seasonality, and phase are computed using a coordinate descent based optimization method. We use bootstrap replication for computing confidence bands and for testing hypotheses about the estimated components. We also utilize log-likelihood for selecting the trend subspace, and for comparisons with other candidate models. This framework is demonstrated using experiments involving synthetic data and three real datasets (Berkeley growth velocity, U.S. electricity price, and USD exchange fluctuation). Our framework is further applied to significance analysis of gene sets for time-course gene expression data and outperforms the state-of-the-art method.

CHAPTER 1

INTRODUCTION

1.1 The Problem

We investigate the problem of estimating trend and seasonality using multiple curves, as opposed to a classical single curve setting. In particular, we tackle a difficult problem where the seasonality exhibits arbitrary time warping, or phase variability (see [MRS+15] for the notion of phase variation), in each observation.

This situation arises often in practical situations where the seasonal effect displays cyclostationary behavior, but is seldom aligned perfectly in the observed data. To illustrate this idea, suppose there are one dimensional signals $\{f_i(t)\}$ as shown in Figure 1.1.

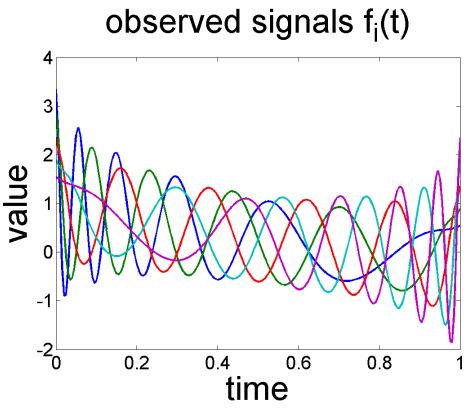


Figure 1.1: Motivation of trend and seasonality estimation.

We observe that for each signal, there is a periodic behavior and all signals have the same number of peaks and valleys. The forms of $\{f_i(t)\}$ lead to the hypothesis that there may be a underlying periodic structure $g(t)$ which is called **seasonality**, **seasonal effect**, or **main structure**. Note that all the peaks and valleys are not exactly aligned in time. For example, the blue signal oscillates from fast to slow whereas the purple signal does the opposite. This time shifting phenomenon is modeled as a **warping function (or phase shift)** $\gamma_i(t)$. The combination of $g(t)$ and $\gamma_i(t)$ is called the variable-phase seasonality in the observation $f_i(t)$.

Notice that the signal envelope is between between -1 and 3 when time $t = 0$. When time moves to $t = 0.5$, the signal envelope is bounded by -1 and 1.5. Until the end at $t = 1$, the signal envelope is between -2 and 2. Because of this behavior, an underlying **trend** $h(t)$ that is decreasing is hypothesized.

Given observed signals $\{f_i(t)\}$, the problem of interest is to recover the seasonality $g(t)$, warping functions $\{\gamma_i(t)\}$, and the underlying trend $h(t)$. This problem hereafter is named as **trend and variable-phase seasonality estimation** , or simply **trend and seasonality estimation**.

The simplest method to analyze the data is to take the cross-sectional mean of all the observed signals $\{f_i(t)\}$, that is

$$\bar{f}(t) = \frac{1}{n}\sum_{i=1}^{n} f_i(t).$$

Figure 1.2 shows the cross-sectional mean $\bar{f}(t)$.

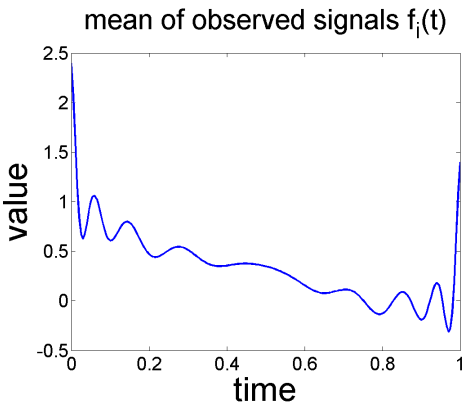
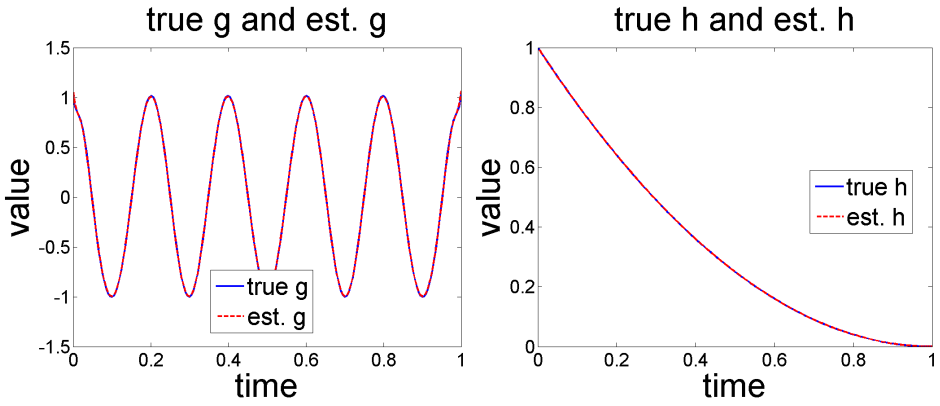


Figure 1.2: Cross-sectional mean of the data in Figure 1.1.

Although the function in Figure 1.2 displays a decreasing trend, it also contains some artifacts that result mainly from the misalignment of seasonal components across individual observations. Therefore, an alignment is needed when estimating components in such data. Without considering all the details, numerical results of trend and seasonality are given in Figure 1.3. The seasonality $g(t)$ is a periodic cosine function and the trend $h(t)$ is monotone decreasing from 1 to 0.



(a) seasonality (b) trend

Figure 1.3: Estimation results from the data in Figure 1.1.

Results for three additional estimations are given in Figures 1.4, 1.5, and 1.6.

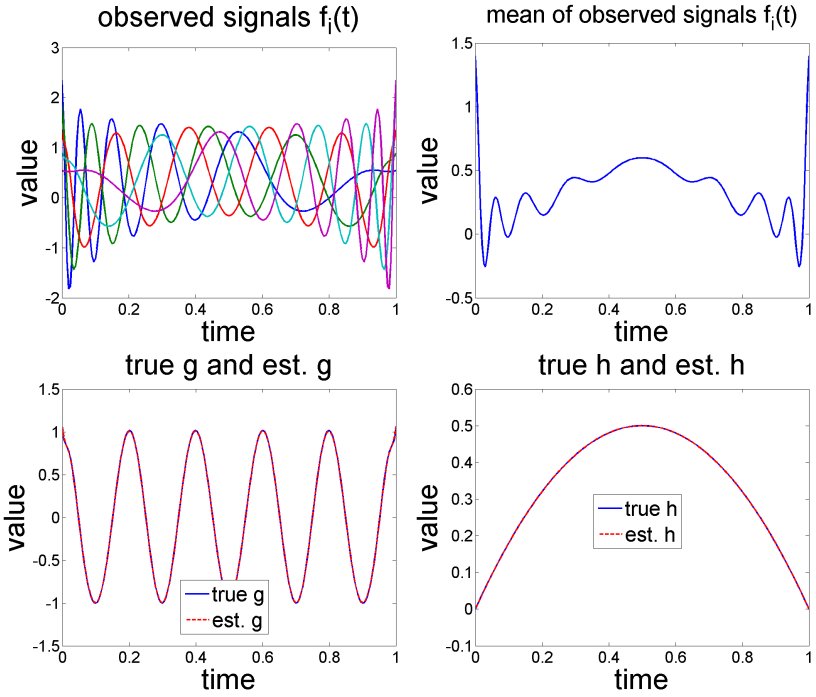


Figure 1.4: Illustration of trend and seasonality estimation. Top left: observations. Top right: cross-sectional mean of observations. Bottom left: recovered seasonality. Bottom right: recovered trend.

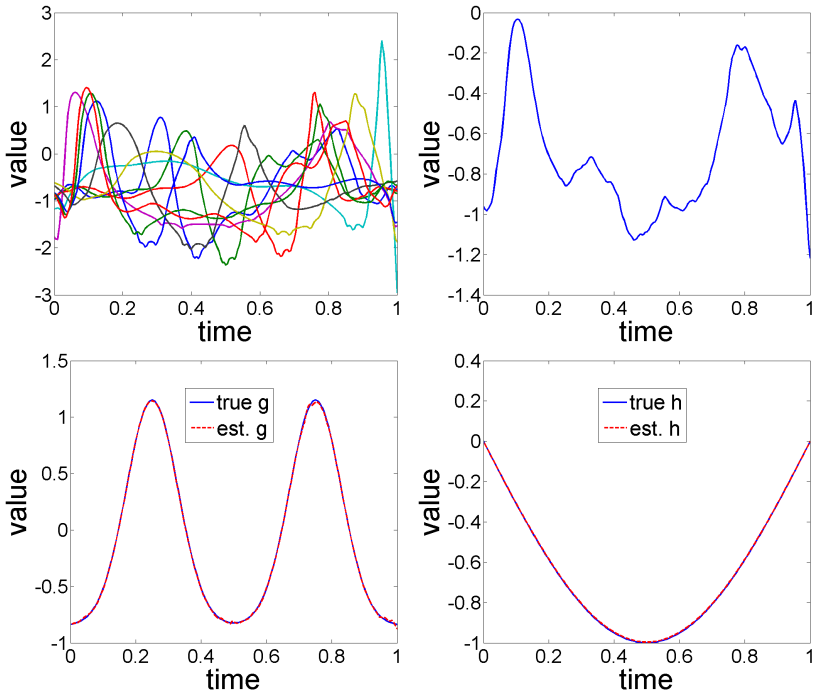


Figure 1.5: Description is same as Figure 1.4

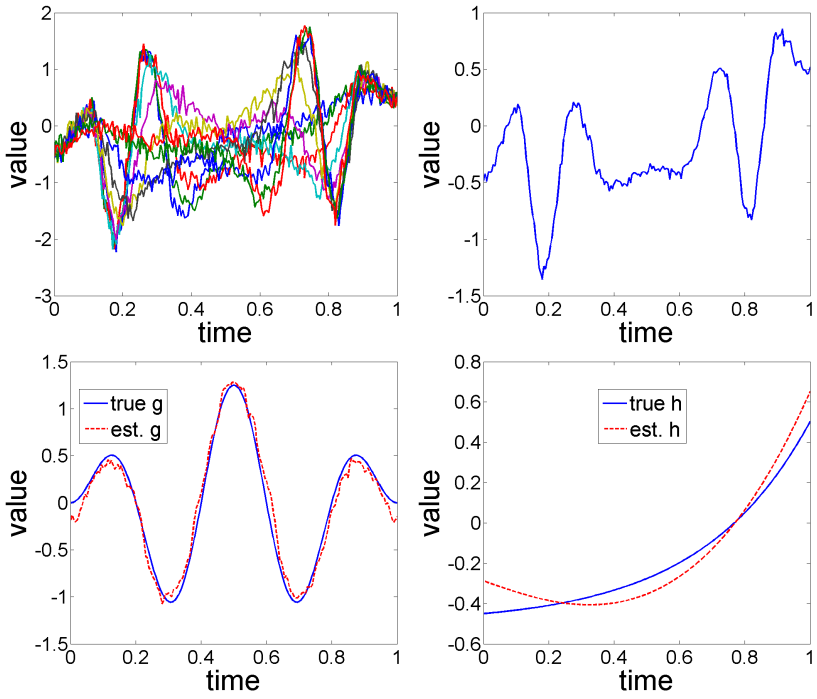


Figure 1.6: Description is same as Figure 1.4

1.2 Overview of the Dissertation

This dissertation develops a statistical model for trend and variable-phase seasonality estimation. An algorithm that, given observed signals $\{f_i\}$, recovers estimates of seasonality, trend, and warping functions is derived and evaluated. The dissertation is organized as follows.

- Chapter 2 starts with a literature review of related work, such as trend and seasonality estimation of a single curve, estimation of a mean function from multiple curves, and registration of curves. A brief discussion of the signal separation problem is given. Finally, several assumptions and the statistical model for estimating the trend and variable-phase seasonality are presented.
- In Chapter 3, we convert the estimation problem into a constrained functional optimization problem and derive a method for its solution.
- In Chapter 4, we describe methods of bootstrap analysis to provide confidence bands around estimates of the trend and the seasonal effect. We also illustrate the strength of the bootstrap framework in formulating hypothesis tests associated with the estimated trends and seasonal effects.
- In Chapter 5, we develop methods for choosing effective parameters in our model.
- Chapter 6 discusses several related models, their assumptions, and their limitations.
- In Chapter 7, experiments with synthetic data are used to study the effect of noise and to demonstrate the robustness of the estimates. Also, results of bootstrap analysis and comparisons of all models are presented.
- In Chapter 8, we study trend and seasonality estimation on three real datasets (Berkeley male growth rates, US electricity prices, and USD currency exchange fluctuations).
- In Chapter 9, we consider significance analysis of gene sets of time-course gene expression data. We utilize the bootstrap analysis in Chapter 4 to select the significant gene sets and show that the statistical power of our method outperforms the state-of-the-art method.
- Chapter 10 summarizes the contribution of this dissertation, and discusses topis for future consideration.

CHAPTER 2

PROBLEM FORMULATION

2.1 Literature Review

Before presenting our model-based solution for trend and seasonality estimation, we summarize the main ideas present in the literature, and point out their limitations and shortcomings. The summary divides relevant literature into four categories, each representing a sub-model of our trend and variable-phase seasonality model.

2.1.1 Trend and Seasonality Estimation from a Single Observation

The problem of estimating trend and seasonality from a single curve originated in economics ([Ner64] and [GK64]), followed by more formal developments in statistics, see [GN70], [CT76], [BHT78], [HT82], [HT83]. Please refer to the review paper by [ABD+12] on this subject, and to U.S. Census Bureau's website[1] for a larger list of papers on this topic. The structure time-series model (also termed the classical decomposition model by [BD06]) was formalized in [HT83] as:

$$f(t) = h(t) + g(t) + \epsilon(t) \ .$$

The goal is to recover the trend h and the seasonality g, from a single observation f. Further special cases of this model result from assuming either h or g to be zero. Common approaches for estimating trend include parametric least squares ([BD06]), moving averages ([BD06]), and local linear smoothing ([FHT01]). On the other hand, the seasonality estimation is often handled by finding cycles (intervals of cyclostationarity) and averaging over those cycles.

For estimating both trend and seasonality, popular approaches include autoregressive-integrated-moving average ([HT82], [HT83]), STL filtering ([CCMT90]), small trend method ([BD06]), moving average estimation ([BD06]), and differencing at lag period ([BD06]). While these methods are based on the assumption of equally-spaced observations, [Eck12] studied the case of unevenly-spaced observations.

[1] http://www.census.gov/srd/www/sapaper/

Note that none of these models address the issue of temporal misalignment of seasonal effects across cycles.

2.1.2 Estimating Mean Function from Multiple Curves

In the case where multiple observations are available, one can use techniques from functional data analysis. For instance, one can pose a model of the type:

$$f_i(t) = h(t) + \epsilon_i(t),\ i = 1, 2, \ldots, n.$$

Under the zero-mean assumption of $\epsilon_i(t)$, an unconstrained estimator of h is the cross-sectional mean, $\hat{h}(t) = \frac{1}{n}\sum_{i=1}^{n} f_i(t)$. If h is assumed to belong to a certain subspace, then there are several techniques available to estimate h: B-splines ([BCF97]), smoothing splines ([BR98]), basis functions ([Ram06]), least squares ([Ram06]), roughness penalty ([Ram06]), and local polynomial kernel ([ZC$^+$07]). This model will perform badly in situations where the data contains some seasonal effects. We mention in passing that although the model used in [Gho01],

$$f_i(t) = h(t) + g_i(t) + \epsilon_i(t),$$

seems different from the one stated above, it is effectively the same given the authors' assumption that $\sum g_i(t) = 0$.

2.1.3 Alignment of Multiple Curves

The third related area is registration or alignment of functions. Here, the model does not have a trend component and only considers time warpings of g according to:

$$f_i(t) = g(\gamma_i(t)) + \epsilon_i(t),\ i = 1, ..., n\ .$$

where $g(\gamma_i(t))$ denotes the misalignment of the seasonality (see more details in Section 2.2). Given several observations $\{f_i\}$, try to remove the effect of $\{\gamma_i\}$ and estimate g.

The simplest case is aligning two signals which was first studied by Sakoe and Chiba [SC78] in 1978 in the signal processing literature. Ten years later, the problem of aligning multiple signals gained substantial interest in the statistics community as the "Registration/Alignment Problem". Table 2.1 summarizes highly cited curve registration papers from 1978 to 2015.

Table 2.1: History review of curve registration methods.

Year	Method	Authors
1978	Dynamic Time Warping (two curves) [SC78]	Sakoe and Chiba;
1992	Landmark Registration [KG92]	Kneip and Gasser
1997	Dynamic Time Warping (multiple curves) [WG97]	Wang and Gasser
1998	Procrustes Fitting Criterion [RL98]	Ramsay and Li
2000	Registration by Local Regression[KLMR00]	Kneip, Li, MacGibbon, Ramsay
2001	Nonparametric MLE [Ron01]	Ronn
2004	Area Under the Curve [LM04]	Liu and Muller
2004	Self-Modeling Registration [GG04]	Gervini and Gasser
2005	Minimum Second Eigenvalue [Ram06]	J. O. Ramsay and X. Li
2007	Moment-Based Method [Jam07]	James
2008	Principal Analysis by Conditional Expectation [TM08]	Tang and Muller
2010	k-means Alignment [SSVV10b, SSVV10a]	Sangalli, Secchi, Vantini , and Vitelli
2011	Fisher-Rao Method [SWK$^+$11, KSW11]	Srivastava, Wu, Kurtek, Klassen, and Marron
2014	Simultaneous Likelihood Inference [RSM14]	Rakêt, Sommer, and Markussen
2015	Bayesian registration [CDH$^+$15]	Cheng, Dryden, and Huang

The alignment properties of the six most common alignment techniques from Table 2.1 are summarized in Table 2.2. Note the Fisher-Rao method is the only method in which the metric is symmetric and isometry to warpings. Moreover, the computational time of the Fisher-Rao method [SWK$^+$11] is less than PACE [TM08] and MBM [Jam07] but higher than AUTC [LM04] and SMR [GG04].

Table 2.2: Comparison of alignment methods.

Methods	alignment performance (visually) [SWK$^+$11]	alignment performance (numerically) [SWK$^+$11]	consistent theorem provided	metric is symmetric	isometry to warping
FTC [RL98]	Low	Low	No	No	No
AUTC [LM04]	Low	Low	Yes	Yes	No
SMR [GG04]	Low	Low	Yes	No	No
MBM [Jam07]	Median	Median	Yes	No	No
PACE [TM08]	Low	Low	Yes	No	No
F-R [SWK$^+$11]	High	High	Yes	Yes	Yes

The majority of the techniques in Table 2.1 formulate the alignment problem using the standard $\mathbb{L}^2$ metric. As pointed out in [MRS$^+$15], this leads to degeneracy in the form of the **pinching effect**,

and also **asymmetry** in the solution. Srivastava et al. [SWK$^+$11] presented a natural solution that extends the Fisher-Rao metric to general function spaces and uses a square-root velocity function (SRVF) representation of curves for alignment (see also [SK16]). This transformation is supported by a fundamental result that the Fisher-Rao Riemannian metric, with its nice invariance properties, transforms to the $\mathbb{L}^2$ inner-product under the SRVF transformation.

Once the registration problem is solved, a natural extension is to estimate the seasonality $g(t)$. This can be accomplished by taking the cross-sectional mean of these aligned SRVFs, see [KSW11] and [CWS16].

Our work differs from the work of Srivastava et al. [SWK$^+$11], Kurtek et al. [KSW11], and Cleveland et al. [CWS16] due to the presence of the trend.

2.1.4 Decomposition of Multiple Signals

Determining the seasonality and the underlying trend can also be regarded as a nonlinear signal separation problem. A standard linear signal separation problem from signal processing is the blind source separation (BSS) [CiA02, HKO01]. The model is

$$\vec{x}(t) = A\vec{s}(t) + \varepsilon(t)$$

where $\vec{x}(t)$ is a vector of observed signals, $\vec{s}(t)$ is a vector of source signals, A is a mixing matrix and $\varepsilon(t)$ is Gaussian noise. Given the observed signals $\vec{x}(t)$, blind source separation attempts to recover the source signal $\vec{s}(t)$ and the mixing matrix A.

Our problem is different from blind source separation due to nonlinear mixing. Also, the trend and seasonality estimation model handles only two signals while blind source separation typically considers mixing many signals.

2.2 Problem Formulation

2.2.1 The Model

To describe the mixing of the seasonality $g(t)$ and the underlying trend $h(t)$, we use an additive model $g(t)+h(t)$. However, the phase/time shifting phenomenon of the seasonality that we discussed in Figure 1.1 is not yet included. We make the warping of the seasonality $g(t)$ to be $g(\gamma_i(t))\sqrt{\dot{\gamma}_i(t)}$ where $\dot{\gamma}$ means the derivative with respect to the time t. Combining all above, the model for the

trend and variable-phase seasonality estimation is

$$f_i(t) = h(t) + g(\gamma_i(t))\sqrt{\dot{\gamma}_i(t)} + \epsilon_i(t),\ i = 1, ..., n. \tag{2.1}$$

On the right, the different terms are:

- **Time**, denoted by t. Without loss of generosity, the time is set to be $[0, 1]$.
- **Observations**, denoted by $f_i \in \mathbb{L}^2([0, 1], \mathbb{R})$, $i = 1, ..., n$.
- **Trend**, denoted by $h \in \mathbb{L}^2([0, 1], \mathbb{R})$, captures the long-term evolution of the data. Generally, we are interested in h being either a lower order polynomial representing a null, constant, linear, quadratic shapes, or slowly varying sinusoid.
- **Warping function**, denoted by $\gamma_i(t)$. We define $\gamma_i : [0, 1] \to [0, 1]$ be a boundary-preserving diffeomorphism. In other words, the warping functions $\{\gamma_i\}$ are assumed to be elements of the set:

  $$\Gamma = \{\gamma : [0, 1] \to [0, 1] \,|\, \gamma(0) = 0,\ \gamma(1) = 1,\ \gamma \text{ is a diffeomorphism}\}.$$

  This set has been studied extensively for function and curve representation in shape and functional data analysis [SKJJ11, SWK$^+$11, SK16]. Notice that Γ is a group with composition as group operation and the identity element $\gamma_{id}(t) = t$. For each $\gamma \in \Gamma$, there exists a unique element γ^{-1} such that $\gamma \circ \gamma^{-1} = \gamma^{-1} \circ \gamma = \gamma_{id}$.
- **Seasonal effect**, denoted by $g \in \mathbb{L}^2([0, 1], \mathbb{R})$, captures seasonal or period effects in the data, Instead of assuming the seasonal effect to be fixed, or perfectly aligned across observations, we make the model more general by including a temporally-misaligned version of g. That is, we utilize the term $g(\gamma_i(t))\sqrt{\dot{\gamma}_i(t)}$, instead of $g(t)$, which represents a time warping of g by a function γ_i, i.e. a variable-phase seasonality g.
- **Observation noise**, denoted by $\epsilon_i : [0, 1] \to \mathbb{R}$ with the assumption that ϵ_i are *i. i. d.* with $E[\epsilon_i(t)] = 0$ for all i and t.

Remark 1. We assume the time warping to be $g \mapsto (g(\gamma))\sqrt{\dot{\gamma}}$, rather than the traditional $g \mapsto g(\gamma)$, as suggested for SRVFs in [SWK$^+$11] and [SK16]. In other words, the model is posed in the SRVF space, rather than the original function space. This warping is norm-preserving, i.e. $\|g\| = \|(g(\gamma))\sqrt{\dot{\gamma}}\|$, with $\|\cdot\|$ denoting the $\mathbb{L}^2$ norm, for any time warping function γ, and thus has fundamentally better mathematical and computational properties. Other possibilities of warpings of g are discussed in Chapter 6.

We replace the term $g(\gamma_i(t))\sqrt{\dot{\gamma}_i(t)}$ by $(g, \gamma_i)(t)$ denoting the right group action of a group Γ on g (see [Fra02]). Thus, Equation 2.1 is simplified to

$$f_i(t) = h(t) + (g, \gamma_i)(t) + \epsilon_i(t),\ i = 1, ..., n. \tag{2.2}$$

To ensure identifiability of trend and seasonality components using the model, Equation 2.2, additional assumptions must be made.

2.2.2 Direct Sum Assumption

Consider a simpler case where $\gamma_i = \gamma_{id} = t$, for all i and the model (Equation 2.2) reduces to

$$f_i(t) = h(t) + g(t) + \epsilon_i(t).$$

In order to identify h and g, we require two subspaces $\mathcal{H}, \mathcal{G} \subset \mathbb{L}^2$, such that $\mathcal{H} \perp \mathcal{G}$, and then we restrict to $h \in \mathcal{H}$ and $g \in \mathcal{G}$. Assuming, as earlier, that $E[\epsilon_i(t)] = 0$, for all i and t, the estimates of h and g are given by

$$\hat{h} = \Pi_{\mathcal{H}}(\frac{1}{n}\sum_{i=1}^{n} f_i)$$

and

$$\hat{g} = \Pi_{\mathcal{G}}(\frac{1}{n}\sum_{i=1}^{n} f_i),$$

where $\Pi_{\mathcal{H}}$ and $\Pi_{\mathcal{G}}$ are the projections onto $\mathcal{H}$ and $\mathcal{G}$, respectively. This, of course, requires the knowledge of $\mathcal{H}$ and $\mathcal{G}$ beforehand. We further assume that $\mathcal{G} = \mathcal{H}^{\perp}$ which is the same as the **direct sum theory** from Functional Analysis [Kre89], that is

$$\mathbb{L}^2 = \mathcal{G} \oplus \mathcal{H} = \mathcal{H}^{\perp} \oplus \mathcal{H}.$$

Now consider the case warpings $\{\gamma_i\}$ other than identity. In this case, the functions (g, γ_i) are no longer guaranteed to be in the subspace $\mathcal{G}$, i.e. (g, γ_i) may have nonzero components in $\mathcal{H}$. This underlines the main challenge in the trend and variable-phase seasonality estimation. If (g, γ_i) had remained orthogonal to $\mathcal{H}$, then we could solve the problem using orthogonal projections (or some related smoothing methods). Note that the warping by itself can be handled using the alignment procedures developed in [SWK$^+$11] and [SK16].

Figure 2.1 illustrates the difficulty when $(g, \gamma_i)(t)$ is no longer orthogonal to $h(t)$. Hence, a more complicated approach is required to perform the separation and characterize its limitations.

To measure how γ acting on g destroys the direct sum assumption, we use the following distortion percentage

$$\frac{\|\Pi_{\mathcal{H}}(g, \gamma)\|_{\mathbb{L}^2}}{\|(g, \gamma)\|_{\mathbb{L}^2}} \times 100\% \tag{2.3}$$

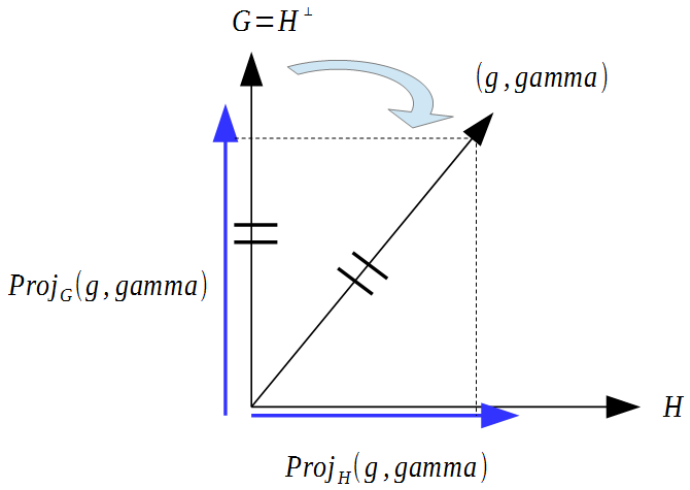


Figure 2.1: Illustration of the difficulty of trend and variable-phase seasonality estimation

as is illustrated by the following synthetic example.

Let $t \in [0,1]$, $g = \cos(11\pi t)$, $\mathcal{H}$= $span\,\{1, \cos(\pi t), \cos(2\pi t), ..., \cos(10\pi t)\}$, and $\gamma = \frac{(0.5)^t - 1}{-0.5}$. The corresponding simulated data is given in Figure 2.2. From Table 2.3, the distortion percentage (Formula 2.3) increases as the dimension of $\mathcal{H}$ increases, as expected. And the corresponding trend and seasonality separation becomes more difficult.

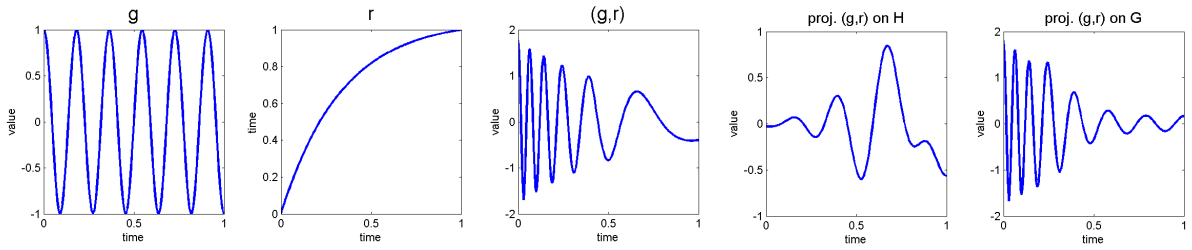


Figure 2.2: Simulated data for the illustration of the distortion percentage.

Table 2.3: Distortion percentage of different dimensions of space $\mathcal{H}$.

$h = \sum_{k=0}^{l_2} c_k \cos(k\pi t)$	$l_2 = 0$	$l_2 = 1$	$l_2 = 2$	$l_2 = 3$	$l_2 = 4$	$l_2 = 5$	$l_2 = 6$	$l_2 = 7$	$l_2 = 8$	$l_2 = 9$
$\frac{\|\Pi_{\mathcal{H}}(g,\gamma)\|}{\|(g,\gamma)\|} \times 100\%$	1.8%	4.9%	12.1%	23.1%	31.5%	31.8%	38.2%	42.0%	45.3%	47.9%

2.2.3 Karcher Mean Assumption

Another difficulty in identifying g arises from the lack of a normalizing assumption. Since

$$\begin{aligned} f_i(t) &= h(t) + (g, \gamma_i)(t) + \epsilon_i(t) \\ &= h(t) + (g, \gamma_{id} \circ \gamma_i)(t) + \epsilon_i(t) \\ &= h(t) + (g, \tilde{\gamma} \circ \tilde{\gamma}^{-1} \circ \gamma_i)(t) + \epsilon_i(t) \\ &= h(t) + \left((g, \tilde{\gamma}), (\tilde{\gamma}^{-1} \circ \gamma_i)\right)(t) + \epsilon_i(t) \\ &= h(t) + (\tilde{g}, \tilde{\gamma}_i) + \epsilon_i(t) \end{aligned}$$

for any $\gamma_0 \in \Gamma$, there is a problem in representing g uniquely. This problem can be avoided by assuming an additional constraint on $\{\gamma_i\}$. Srivastava et al. [SWK$^+$11] suggested forcing the Karcher mean of the inverse warping functions to be the identity, or

$$KM\left\{\gamma_i^{-1}\right\} = \gamma_{id}.$$

(The concept of Karcher mean on Γ is discussed later in Section 3.3.) The similar idea of constraining the mean of warping functions to be γ_{id} was used in [TM08].

2.2.4 The Constrained Functional Optimization

We will use the maximum-likelihood approach to recover the seasonality g, warping functions $\{\gamma_i\}$, and the trend h in Equation 2.2 Since using Maximum Likelihood directly for the infinite dimensional problem is unclear, we must discretize the problem. For notational simplicity, the infinite dimensional model is

$$f_i(t) = f_i^{\Omega}(t) + \epsilon_i(t),\ i = 1, ..., n$$

where $f_i^{\Omega}(t) = h(t) + (g, \gamma_i)(t)$ denotes parameter terms. The finite discretized model is

$$f_i(t_j) = f_i^{\Omega}(t_j) + \epsilon_i(t_j),\ i = 1, ..., n,\ j = 1, ..., m$$

where m is the number of time samples in the interval $[0, 1]$. Assuming $\epsilon_i(t_j) \sim \mathcal{N}(0, \sigma^2)$ for each i and j, the conditional distribution of $f_i(t_j)$ given $f_i^{\Omega}(t_j)$ is

$$f_i(t_j) \sim N(f_{ij}^{\Omega}, \sigma^2).$$

For each observation i, we have random variables $f_i(t_1), ..., f_i(t_m)$ and the likelihood function is

$$\begin{aligned} L(f_i^\Omega(t_1), ..., f_i^\Omega(t_m) | f_i(t_1), ..., f_i(t_m)) &= pdf(f_i(t_1)) \times ... \times pdf(f_i(t_m)) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^m \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{m}\left(f_i(t_j) - f_i^\Omega(t_j)\right)^2\right]. \end{aligned}$$

Taking the average log likelihood gives

$$\frac{1}{m}\log L\left(f_i^\Omega(t_1), ...f_i^\Omega(t_m) |\, f_i(t_1), ..., f_i(t_m)\right) = \log\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2m\sigma^2}\sum_{j=1}^{m}\left(f_i(t_j) - f_i^\Omega(t_j)\right)^2.$$

Taking the average for observation i yields

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\log L\left(f_i^\Omega(t_1), ...f_i^\Omega(t_m) |\, f_i(t_1), ..., f_i(t_m)\right) = \frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2m\sigma^2}\sum_{j=1}^{m}\left(f_i(t_j) - f_i^\Omega(t_j)\right)^2\right]. \tag{2.4}$$

Maximizing Equation 2.4 is the same as minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{m}\sum_{j=1}^{m}\left(f_i(t_j) - f_i^\Omega(t_j)\right)^2\right].$$

In the limit, $m \to \infty$, we have

$$\begin{aligned} \lim_{m\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{m}\sum_{j=1}^{m}\left(f_i(t_j) - f_i^\Omega(t_j)\right)^2\right] &= \frac{1}{n}\sum_{i=1}^{n}\int_0^1\left(f_i(t) - f_i^\Omega(t)\right)^2 dt \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\|f_i(t) - f_i^\Omega(t)\right\|_{\mathbb{L}^2}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\|f_i(t) - h(t) - (g,\gamma_i)(t)\right\|_{\mathbb{L}^2}^2. \end{aligned}$$

Thus, finding the Maximum Likelihood Estimator (MLE) becomes a problem in constrained functional minimization

$$(\hat{g}, \{\hat{\gamma}_i\}, \hat{h}) = \arg\min_{g,\{\gamma_i\},h} C(g, \{\gamma_i\}, h) \tag{2.5}$$

where $C : \mathcal{G} \times \Gamma^n \times \mathcal{H} \to \mathbb{R}$ is given by the cost function

$$C(g, \{\gamma_i\}, h) = \frac{1}{n}\sum_{i=1}^{n}\left\|f_i - h - (g,\gamma_i)\right\|^2 \tag{2.6}$$

with two constrains: direct sum $\mathbb{L}^2 = \mathcal{G} \oplus \mathcal{H}$ and the Karcher mean $KM\{\gamma_i^{-1}\} = \gamma_{id}$.

CHAPTER 3

OPTIMIZATION APPROACH

There are several ways to solve a functional optimization problem but each of them addresses a problem with a specific form. For example, if the cost functional involves integrals, perhaps in an integral equality constraint, one may resort to Variational Calculus [CH66, Lue69]. If the cost functional is no longer an integral form, there are several gradient methods such as Generalized gradient [CMP+06], Sobolev gradient [JCS+09], Fréchet derivative [AB06], and Gâteaux derivative[AB06]. If the argument of the cost functional has a certain shape and is not defined on a vector space, then the problem leads to topology optimization [BS13] or shape optimization [SZ92].

The difficulties of our problem are the arguments $\gamma_i(t)$, which have a particular structure, and the Karcher mean constraint. A coordinate descent method is used because its convergence and efficiency in the finite dimensional case are understood [LT92, TM01, Nes12]. However, little work has been done in the infinite dimensional case.

The coordinate descent method optimizes the cost function along one direction/variable at a time and iterates until a stationary point with a sufficiently small cost function value is reached. To apply coordinate descent method to our problem, we must derive the minimizer in each direction. i.e.,

1. Given current estimates of $g(t)$ and $\{\gamma_i\}$, what is the minimizer $\hat{h}(t)$ to the cost functional?
2. Given current estimates of $\{\gamma_i\}$ and $h(t)$, what is the minimizer $\hat{g}(t)$ to the cost functional?
3. Given current estimates of $g(t)$ and $h(t)$, what is the minimizer $\{\hat{\gamma}_i\}$ to the cost functional?

Recall from Equation 2.6, we have the cost functional

$$C(g(t), \gamma_i(t), h(t)) = \frac{1}{n}\sum_{i=1}^{n} \| f_i(t) - h(t) - (g, \gamma_i)(t)\|^2_{\mathbb{L}^2} \tag{3.1}$$

with direct sum assumption $f_i \in \mathbb{L}^2 = \mathcal{G} \oplus \mathcal{H}$ and the Karcher mean condition $KM\{\gamma_i^{-1}\} = \gamma_{id}$.

3.1 Updating the Estimate of h

Suppose $\hat{g} \in \mathcal{G}$ and $\hat{\gamma}_i \in \Gamma$ are current estimates, the corresponding optimization to update estimate for $h(t)$ is from Equation 2.6

$$\hat{h} = \arg\min_{h \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \|f_i - h(t) - (\hat{g}, \hat{\gamma}_i)\|^2 . \tag{3.2}$$

From

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} \|f_i - (\hat{g}, \hat{\gamma}_i) - h\|^2 &= \frac{1}{n} \sum_{i=1}^{n} \left\| \tilde{f}_i - h \right\|^2 = \frac{1}{n} \sum_{i=1}^{n} \left\| \tilde{f}_i^{\mathcal{H}} + \tilde{f}_i^{\mathcal{H}^\perp} - h \right\|^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} \left\| \tilde{f}_i^{\mathcal{H}} - h \right\|^2 + \frac{1}{n} \sum_{i=1}^{n} \left\| \tilde{f}_i^{\mathcal{H}^\perp} \right\| , \end{aligned}$$

and the cost functional for estimating h is

$$C(h) = \frac{1}{n} \sum_{i=1}^{n} \left\| \tilde{f}_i^{\mathcal{H}} - h \right\|^2 = \frac{1}{n} \sum_{i=1}^{n} \left\langle \tilde{f}_i^{\mathcal{H}}, \tilde{f}_i^{\mathcal{H}} \right\rangle_{\mathbb{L}^2} - 2 \left\langle \frac{1}{n} \sum_{i=1}^{n} \tilde{f}_i^{\mathcal{H}}, h \right\rangle_{\mathbb{L}^2} + \langle h, h \rangle_{\mathbb{L}^2} \tag{3.3}$$

where $\tilde{f}_i = f_i - (\hat{g}, \hat{\gamma}_i)$ and $\tilde{f}_i^{\mathcal{H}}(t)$ is the projection of $\tilde{f}_i$ onto the the space $\mathcal{H}$. Let $\Phi = \{\phi_k, k = 1, \ldots, \infty\}$ denote a complete orthonormal basis of $\mathbb{L}^2$. We assume $\{\phi_k\}$ are ordered so that $\mathcal{H} = \text{span}\{\phi_k | k = 1, \ldots, l\}$ for some integer $1 \leq l < \infty$. In practice, this assumption is motivated by the standard ordering of the basis elements from slowly to rapidly varying. Therefore the choice of $\mathcal{H}$ can be made by considering an appropriate value of l. The minimizer occurs when the middle term in Equation 3.3 is maximized at

$$\hat{h} = \frac{1}{n} \sum_{i=1}^{n} \tilde{f}_i^{\mathcal{H}} = \frac{1}{n} \sum_{i=1}^{n} \Pi_{\mathcal{H}} \left[f_i - (\hat{g}, \hat{\gamma}_i) \right] = \Pi_{\mathcal{H}} \left[\frac{1}{n} \sum_{i=1}^{n} \left(f_i - (\hat{g}, \hat{\gamma}_i) \right) \right] = \sum_{k=1}^{l} \left\langle \frac{1}{n} \sum_{i=1}^{n} \left(f_i - (\hat{g}, \hat{\gamma}_i) \right), \phi_k \right\rangle \phi_k$$

where $\{\phi_1(t), ..., \phi_l(t)\}$ is an orthogonal basis of $\mathcal{H}$ and $\langle \cdot, \cdot \rangle$ denotes the standard $\mathbb{L}^2$ inner product.

We consider the following bases: sine basis $\left\{ \sqrt{2} \sin(n\pi t), \, n = 1, 2, 3, ... \right\}$, cosine basis $\left\{ 1, \sqrt{2} \cos(n\pi t), \, n = 1, 2, 3, ... \right\}$, Fourier basis $\left\{ 1, \sqrt{2} \sin(2n\pi t), \sqrt{2} \cos(2n\pi t), \, n = 1, 2, 3, ... \right\}$, and shifted Legendre basis (see [Kre89])

$$\left\{ \phi_k(t) = \frac{1}{2k-1} (-1)^{k-1} \sum_{j=0}^{k-1} \binom{k-1}{j} \binom{k+j-1}{j} (-t)^j, k = 1, 2, 3... \right\} .$$

3.2 Updating the Estimate of g

Suppose $\hat{\gamma}_i \in \Gamma$ and $\hat{h} \in \mathcal{H}$ are the current estimates, the functional optimization problem for updating the estimator g is

$$\hat{g} = \underset{g \in \mathcal{H}^\perp}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \left\| f_i - \hat{h} - (g, \hat{\gamma}_i) \right\|^2 . \tag{3.4}$$

Lemma 2 in [SWK+11] states that

$$\|f_1 - f_2\|_{\mathbb{L}^2} = \|(f_1, \gamma) - (f_2, \gamma)\|_{\mathbb{L}^2} ,$$

from which we have

$$\frac{1}{n} \sum_{i=1}^{n} \left\| f_i - (g, \hat{\gamma}_i) - \hat{h} \right\|^2 = \frac{1}{n} \sum_{i=1}^{n} \left\| \left((f_i - \hat{h}), \hat{\gamma}_i^{-1} \right) - g \right\|^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left\| \tilde{f}_i - g \right\|^2 = \frac{1}{n} \sum_{i=1}^{n} \left\| \tilde{f}_i{}^{\mathcal{H}} + \tilde{f}_i{}^{\mathcal{H}^\perp} - g \right\|^2 = \frac{1}{n} \sum_{i=1}^{n} \left\| \tilde{f}_i{}^{\mathcal{H}} - g \right\|^2 + \frac{1}{n} \sum_{i=1}^{n} \left\| \tilde{f}_i{}^{\mathcal{H}^\perp} \right\| .$$

The cost function for g is then

$$C(g) = \frac{1}{n} \sum_{i=1}^{n} \left\| \tilde{f}_i{}^{\mathcal{H}^\perp} - g \right\|^2$$

where $\tilde{f}_i = \left((f_i - \hat{h}), \hat{\gamma}_i^{-1} \right)$ and $\tilde{f}_i{}^{H^\perp}$ denotes the projection of $\tilde{f}_i$ onto the space $\mathcal{H}^\perp$. The cost function can be written as

$$\begin{aligned} C(g) &= \frac{1}{n} \sum_{i=1}^{n} \left\langle \tilde{f}_i^{\mathcal{H}^\perp} - g, \tilde{f}_i^{\mathcal{H}^\perp} - g \right\rangle_{\mathbb{L}^2} \\ &= \frac{1}{n} \sum_{i=1}^{n} \left(\left\langle \tilde{f}_i^{\mathcal{H}^\perp}, \tilde{f}_i^{\mathcal{H}^\perp} \right\rangle_{\mathbb{L}^2} - 2 \left\langle \tilde{f}_i^{\mathcal{H}^\perp}, g \right\rangle_{\mathbb{L}^2} + \langle g, g \rangle_{\mathbb{L}^2} \right) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left\langle \tilde{f}_i^{\mathcal{H}^\perp}, \tilde{f}_i^{\mathcal{H}^\perp} \right\rangle_{\mathbb{L}^2} - 2 \left\langle \frac{1}{n} \sum_{i=1}^{n} \tilde{f}_i^{\mathcal{H}^\perp}, g \right\rangle_{\mathbb{L}^2} + \langle g, g \rangle_{\mathbb{L}^2} . \end{aligned} \tag{3.5}$$

The minimizer of $C(g)$ occurs when the middle term in Equation 3.5 is maximized at

$$\hat{g} = \frac{1}{n} \sum_{i=1}^{n} \tilde{f}_i^{\mathcal{H}^\perp} = \frac{1}{n} \sum_{i=1}^{n} \Pi_{\mathcal{H}^\perp} \tilde{f}_i = \Pi_{\mathcal{H}^\perp} \frac{1}{n} \sum_{i=1}^{n} \tilde{f}_i .$$

The projection $\Pi_{\mathcal{H}^\perp}$ can be expressed in terms of the projector Π_{calH} using the orthonormal basis $\{\phi_1, ..., \phi_l\}$ to yield the minimizer from Equation 3.2 in the computationally convenient form

$$\begin{aligned}\hat{g} &= \frac{1}{n}\sum_{i=1}^{n}\tilde{f}_i - \Pi_{\mathcal{H}}\frac{1}{n}\sum_{i=1}^{n}\tilde{f}_i \\ &= \frac{1}{n}\sum_{i=1}^{n}\left((f_i - \hat{h}), \hat{\gamma}_i^{-1}\right) - \Pi_{\mathcal{H}}\frac{1}{n}\sum_{i=1}^{n}\left((f_i - \hat{h}), \hat{\gamma}_i^{-1}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left((f_i - \hat{h}), \hat{\gamma}_i^{-1}\right) - \sum_{k=1}^{l}\left\langle \frac{1}{n}\sum_{i=1}^{n}\left((f_i - \hat{h}), \hat{\gamma}_i^{-1}\right), \phi_k \right\rangle_{\mathbb{L}^2}.\end{aligned}$$

3.3 Updating the Estimates of $\{\gamma_i\}$

Given current estimates of $\hat{g}$ and $\hat{h}$, the estimates of γ_i are updated to

$$\{\hat{\gamma}_1, ..., \hat{\gamma}_n\} = \underset{\{\gamma_1,...,\gamma_n\}\in\Gamma^n}{\arg\min}\frac{1}{n}\sum_{i=1}^{n}\left\| f_i(t) - (\hat{g}, \gamma_i)(t) - \hat{h}(t)\right\|^2 \tag{3.6}$$

with the constraint $KM\left\{\gamma_i^{-1}\right\} = \gamma_{id}$. These minimizers are approximated by first solving the n optimization problem $1 \leq i \leq n$

$$\{\tilde{\gamma}_i\} = \underset{\gamma_i\in\Gamma}{\arg\min}\left\| f_i - \hat{h} - (\hat{g}, \gamma_i)\right\|^2 \tag{3.7}$$

and then imposing the Karcher mean constraint on the resulting warping functions. The optimal $\{\gamma_i\}$ can be found using the technique in [SWK$^+$11].

The Karcher mean condition, $KM\left\{\gamma_i^{-1}\right\} = \gamma_{id}$, is imposed on the solutions of the unconstrained problem, Equation 3.7, in two steps (details are in Appendix). First, the Karcher mean of $\gamma_1^{-1}, ..., \gamma_n^{-1}$ is computed using Algorithm 1 from [SWK$^+$11].

Second, given the Karcher mean of $\gamma_1^{-1}, ..., \gamma_n^{-1}$, the condition $KM\left\{\gamma_i^{-1}\right\} = \gamma_{id}$ can be imposed using the following Lemma (Lemma 4 in [SWK$^+$11]).

Lemma. *Let Γ be a diffeomorphism group on $[0, 1]$. Let $\eta_0 \in \Gamma$ and η_{KM} be the Karcher mean of $\eta_1, ..., \eta_n \in \Gamma$, then $\eta_{KM} \circ \eta_0$ is the Karcher mean of $\eta_1 \circ \eta_0, ..., \eta_n \circ \eta_0$.*

So, if γ_{KM}^{-1} is the Karcher mean of $\gamma_1^{-1}, ..., \gamma_n^{-1}$, then $KM\left\{\gamma_i^{-1}\right\} = \gamma_{id}$ is imposed by the replacement

$$\gamma_i^{-1} \leftarrow \gamma_i^{-1} \circ \left(\gamma_{KM}^{-1}\right)^{-1}, \; 1 \leq i \leq n$$

Algorithm 1: Karcher mean of warping functions $\eta_1, ..., \eta_n$.

Data: warping functions $\eta_1, ..., \eta_n$

Result: Karcher mean of warping functions η_{KM}

Compute $\psi_i = \sqrt{\dot{\eta}_i}$ for $i = 1, ..., n$;

Initialize $\psi^{(0)} = \frac{1}{n}\sum_{i=1}^{n}\psi_i$;

for $k = 1$ *until convergence* **do**

  for $i = 1, ..., n$ **do**

    Compute $v_i = \exp^{-1}_{\psi^{(k)}}(\psi_i) = \frac{\theta}{\sin\theta}\left(\psi_i - \cos(\theta)\psi^{(k)}\right)$ where $\theta = \cos^{-1}\left(\langle\psi, \psi_i\rangle_{\mathbb{L}^2}\right)$;

  end

  Compute the average direction, $\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$;

  Pick up a small ε and times to the average direction $\bar{v}$;

  Back to the unit sphere by doing retraction

$$\psi^{(k+1)} = \exp_{\psi^{(k)}}(\varepsilon\bar{v}) = \cos\left(\|\varepsilon\bar{v}\|\right)\psi^{(k)} + \sin\left(\|\varepsilon\bar{v}\|\right)\frac{\varepsilon\bar{v}}{\|\varepsilon\bar{v}\|}$$

  If $\left\|\psi^{(k+1)} - \psi^{(k)}\right\|$ is very small, then end the **for** loop ;

end

Let $\psi_{KM} = \psi^{(\text{final } k)}$;

Output η_{KM} by computing $\eta_{KM} = \int_0^1 \psi_{KM}^2 dt$;

which is equivalent to

$$\gamma_i \leftarrow \left[\gamma_i^{-1} \circ \left(\gamma_{KM}^{-1}\right)^{-1}\right]^{-1} = \left(\gamma_{KM}^{-1}\right) \circ \gamma_i,\ 1 \le i \le n.$$

The algorithm for approximating the solution of the constrained functional optimization is described in Algorithm 2 and the algorithm to solve the trend and variable-phase seasonality is given in Algorithm 3.

Algorithm 2: Optimization over $\gamma_1, ..., \gamma_n$.

Data: observations $\{f_i\}$, seasonality $g^{(j-1)}$ and trend $h^{(j-1)}$. Requires Algorithm 1.

Result: warping functions $\gamma_1^{(j)}, ..., \gamma_n^{(j)}$ whose $KM\{\gamma_i^{-1}\} = \gamma_{id}$.

for $i = 1, ..., n$ **do**

  Solve

$$\tilde{\gamma}_i^{(j)} = \arg\min_{\gamma_i \in \Gamma} \left\| \left[f_i - h^{(j-1)} \right] - (g^{(j-1)}, \gamma_i) \right\|^2$$

  Compute $\left(\tilde{\gamma}_i^{(j)}\right)^{-1}$;

end

Use Algorithm 1 to compute Karcher mean of $\left(\tilde{\gamma}_1^{(j)}\right)^{-1}, ..., \left(\tilde{\gamma}_n^{(j)}\right)^{-1}$, denoted by $\left(\tilde{\gamma}_{KM}^{(j)}\right)^{-1}$;

for $i = 1, ..., n$ **do**

  Update $\gamma_i^{(j)} = \left(\tilde{\gamma}_{KM}^{(j)}\right)^{-1} \circ \tilde{\gamma}_i^{(j)}$;

end

Algorithm 3: A complete algorithm for solving trend and variable-phase seasonality estimation.

Data: given observations $\{f_i\}$. Requires **Algorithm** 2 (which requires **Algorithm** 1)

Result: warping functions $\{\hat{\gamma}_i\}$, seasonality $\hat{g}$, and trend $\hat{h}$

Initialization $g^{(0)} = f_{idx}$ where $idx = \arg\min_i \left\| f_i - \bar{f} \right\|$, $h^{(0)} = 0$, $\gamma_i^{(0)} = \gamma_{id}$;

for $j = 1, ...,$ *until convergence* **do**

  Update $\gamma_1^{(j)}, ..., \gamma_n^{(j)}$ with Karcher mean condition using Algorithm 2 ;

  Compute $\bar{f}^{(j)} = \frac{1}{n} \sum_{i=1}^{n} \left(\left(f_i - h^{(j-1)} \right), \left(\gamma_i^{(j)} \right)^{-1} \right)$;

  Update $g^{(j)} = \bar{f}^{(j)} - \Pi_{\mathcal{H}} \bar{f}^{(j)}$;

  Compute $\check{f}^{(j)} = \frac{1}{n} \sum_{i=1}^{n} \left[f_i - (g^{(j)}, \gamma_i^{(j)}) \right]$;

  Update $h^{(j)} = \Pi_{\mathcal{H}} \check{f}^{(j)}$;

  If the value $|C(\{\gamma_i^{(j)}\}, h^{(j)}, g^{(j)}) - C(\{\gamma_i^{(j-1)}\}, h^{(j-1)}, g^{(j-1)})|$ is very small, then end the **for** loop;

end

Output $\{\hat{\gamma}_i\} = \{\gamma_i^{(max)}\}$, $\hat{g} = g^{(max)}$, and $\hat{h} = h^{(max)}$

CHAPTER 4

BOOTSTRAP ANALYSIS

Given the complexity of the data model, and the subsequent estimators, it is difficult to derive analytical expressions for asymptotic distributions of the estimated quantities. In this chapter we use a bootstrap approach to compute estimator statistics using random replication, and use these statistics to test hypotheses about trend and seasonality. We present details of computing bootstrap estimates of standard deviations of certain test statistics. These statistics, in turn, can be used to test important hypotheses, such as the presence or absence of a trend or a seasonality component in the observed data.

4.1 Review of Bootstrap Method

In statistics, the most common task is to estimate the mean of a population X. Taking the average of samples $\{x_1, ..., x_n\}$ drawn from the population is the best estimator for this problem. However, one can question the accuracy of the sample mean $\bar{x}$. The idea of a confidence interval which was first introduced by Jerzy Neyman [Ney37]. Assuming the standard deviation of the population is unknown, the confidence interval becomes

$$\left[\bar{x} - t^* \frac{s}{\sqrt{n}}, \bar{x} + t^* \frac{s}{\sqrt{n}}\right]$$

where s is the sample deviation, n is the number of samples, and t^* is the t-score. The quantity $s/\sqrt{n}$ is also called the standard error. Since the mean of $x_1, ..., x_n$ is normally distributed, the standard error is given by an expression directly from the given samples $x_1, ..., x_n$, and the accuracy of the sample mean can be estimated. The jackknife technique was invented by Quenouille [Que49] in 1949 and further developed by Tukey [Tuk58] in 1958 to handle this situation when the standard error is also not known. Given an estimator $\hat{\theta}$, let $\hat{\theta}_{(i)}$ be the estimate obtained from the samples $\{x_1, ..., x_n\}$ with x_i removed. The jackknife estimator $\bar{\theta}$ is

$$\bar{\theta} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{(i)}$$

and the variance of the $\bar{\theta}$ is

$$Var(\bar{\theta}) = \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{(i)} - \hat{\theta}\right).$$

We can find the standard error of $\hat{\theta}$ using the variance of the jackknife estimate. The limitation of jackknife is that it only works well for linear statistics such as the mean. Other non-smooth estimators, e.g. median, and nonlinear estimators, e.g. correlation coefficient, are not accurately approximated. To address this limitation, the **bootstrap** method was invented by Efron [E$^+$79, ET94] in 1979. He observed that the jackknife method is a linear approximation of the bootstrap method.

Suppose the estimator θ is approximated by the estimator $\hat{\theta} = s(\mathbf{x})$ where s is a function of given set of samples from the population. The **bootstrap sample** $\mathbf{x}^* = (x_1^*, ..., x_n^*)$ is a random vector of real numbers drawn from the observed samples x uniformly with replacement. The **bootstrap replicate** of $\hat{\theta}$ is the estimator from a bootstrap sample, denoted as $\hat{\theta}^* = s(x^*)$. Let $\mathbf{x}^{*1}, ..., \mathbf{x}^{*B}$ be B independent bootstrap samples of x and let $\hat{\theta}^{*b} = s(\mathbf{x}^{*b})$, $b = 1, ..., B$ be the resulting bootstrap replicates of $\hat{\theta}$. The **bootstrap estimate of standard error** of $\hat{\theta}$ is defined as the standard deviation of the bootstrap replicates of $\hat{\theta}$, that is

$$\hat{se}_B = \left[\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}^{*b} - \bar{\hat{\theta}}^*\right)^2\right]^{\frac{1}{2}}$$

where

$$\bar{\hat{\theta}}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b}.$$

This expression provides a numerical technique for estimating standard error of θ. Algorithm 4 gives the bootstrap procedure for computing standard error of an arbitrary estimator θ.

Once the standard error of an estimator is known, sample hypothesis testing may be done. Suppose the null hypothesis is $H_0 : \theta = \mu_0$, then the test statistic

$$\frac{\theta - \mu_0}{\hat{se}_B}$$

follows Student's t distribution, or standard normal distribution when $B \geq 200$. Given an assumed distribution of the test statistic, associated p-values can be obtained.

Another application of bootstrap analysis is the cross-sectional confidence band. Suppose the goal is to estimate a function $f(t)$, $t \in \mathcal{T}$ with coverage probability $1 - \alpha$. The **cross-sectional**

Algorithm 4: Bootstrap method for estimating standard error of estimator θ.

Data: observed samples: $\mathbf{x} = \{x_1, ..., x_n\}$

Result: standard error of an estimator θ

1 Construct B bootstrap samples $\mathbf{x}^{*1}, ..., \mathbf{x}^{*B}$. Each of them is of size n, generated by resampling samples $\mathbf{x} = \{x_1, ..., x_n\}$ uniformly with replacement;

2 For each bootstrap samples $\mathbf{x}^{*1}, ..., \mathbf{x}^{*B}$, compute the bootstrap replicates $\hat{\theta}^{1*}, ..., \hat{\theta}^{*B}$;

3 Compute the standard error of $\hat{\theta}$ by

$$\hat{se}_B = \left[\frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}^{*b} - \bar{\hat{\theta}}^{*} \right)^2 \right]^{\frac{1}{2}}$$

where $\bar{\hat{\theta}}^{*} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^{*b}$.

confidence band is defined as

$$Prob\left\{ \hat{f}(t) - w(t) \leq f(t) \leq \hat{f}(t) + w(t) \right\} = 1 - \alpha$$

for all $t \in \mathcal{T}$. Note that $w(t)$ is regarded as the width function which makes $\hat{f}(t) + w(t)$ an upper bound and $\hat{f}(t) - w(t)$ a lower bound. In contrast to the cross-sectional confidence band, the **simultaneous confidence band** is defined as

$$Prob\left\{ \hat{f}(t) - w(t) \leq f(t) \leq \hat{f}(t) + w(t), \, for \, all \, t \in \mathcal{T} \right\} = 1 - \alpha. \tag{4.1}$$

Although the bootstrap simultaneous confidence band applied to functional data analysis has been introduced in recent literature [CFF06, Deg11, CVK03, FZ00, GGC13, MYC12], it is only used for the confidence band of the mean function. Computing the simultaneous confidence band a non-mean function is not yet developed, and hence, we use the cross-sectional confidence band.

When considering the pointwise confidence band, we compute the confidence interval for each time point. Let $i = 1, ..., n$ be the index of observed functions and $j = 1, ..., T$ be the index of time grids. For each time t_j, we have observed samples x_{ij} with estimator $\hat{\theta}_j$. We construct bootstrap resamples $\{x_{ij}^{*1}, ..., x_{ij}^{*B}\}$ from observed samples x_{ij}. For each bootstrap resamples, calculate bootstrap replicates to have $\theta^{*1}, ..., \theta^{*B}$. Algorithm 4 is used to obtain standard error $\hat{se}_B$. By selecting significance level α (or confidence level $1 - \alpha$), the confidence interval for for θ_j at each time t_j is

given by

$$Prob\left\{ z^{(\alpha/2)} \leq \frac{\hat{\theta}_j - \theta_j}{\hat{se}_B} \leq z^{(1-\alpha/2)} \right\} = 1 - \alpha, \tag{4.2}$$

where $z^{\alpha/2}$ and $z^{1-\alpha/2}$ indicate, respectively, the $100 \cdot (\alpha/2)$th and $100 \cdot (1-\alpha/2)$th percentile point of a standard normal distribution. Equation 4.2 can be written in the more compact form

$$\theta_j \in [\hat{\theta}_j \pm z^{1-\alpha/2} \cdot \hat{se}_B].$$

4.2 Bootstrap Analysis of Trend and Seasonality Estimation Results

4.2.1 Bootstrap Hypothesis Testing

Given a set of functions $\{f_i, i = 1, ..., n\}$, we can pose the question: Is $h = 0$, or not? This leads to a formal binary hypothesis test, with null hypothesis H_0: $h = 0$ and the alternative hypothesis H_1: $h \neq 0$. In view of the implicit assumptions of continuity of h (due to the condition that $h \in \mathcal{H}$, a subspace of smooth functions), we have that $h = 0$ is equivalent to $\|h\| = 0$. Therefore, we define a test statistic $\rho_{h_0} = \|\hat{h}\|$ and rewrite the hypothesis test as: null hypothesis H_0: $\rho_{h_0} = 0$ and the alternative hypothesis H_1: $\rho_{h_0} > 0$. Let $\hat{h}_b$ denote the bootstrap replicate of the estimator $\hat{h}$, and let $\rho_{h_0,b}$ denote its $\mathbb{L}^2$ norm. Furthermore, let $\hat{se}_B$ be the standard error of $\rho_{h_0,b}$ using B replicates. Then, we can compute the p value of the test statistic assuming a normal distribution $\mathcal{N}(0, \hat{se}_B)$ under the null hypothesis. (Note: this $\hat{se}_B$ can be estimated from the data using Algorithm 4.)

In fact, one can use the bootstrap procedure to test any specific shape pattern of the trend and seasonality functions. For instance, one can test the trend function being constant, linear, or monomial of certain order. As an example, we can test if the trend h is a constant function by modifying the test statistic to be $\rho_{h_c} = \left\| \hat{h} - \int_0^1 \hat{h}\, dt \right\|$.

Property 4.2.1. *$\hat{h}(t) = c$ if and only if $\rho_{h,c} = 0$*

Proof. ($\Rightarrow$) if $\hat{h} = c$, then

$$\rho_{h,c} = \left\| \hat{h} - \int_0^1 \hat{h}\, dt \right\| = \left\| c - \int_0^1 c\, dt \right\| = \|c - c\| = 0.$$

($\Leftarrow$) if we have $\rho_{h_c} = 0$, then

$$\left\|\hat{h} - \int_0^1 \hat{h}\,dt\right\| = 0 \Rightarrow \left[\int_0^1 \left(\hat{h} - \int_0^1 \hat{h}\,dt\right)^2 dt\right]^{\frac{1}{2}} = 0$$
$$\Rightarrow \int_0^1 \left(\hat{h} - \int_0^1 \hat{h}\,dt\right)^2 dt = 0$$
$$\Rightarrow \hat{h} = \int_0^1 \hat{h}\,dt = \text{constant}$$

□

A test statistic for testing the linearity of the trend h is $\rho_{h_l} = \left\|\dot{\hat{h}} - \int_0^1 \dot{\hat{h}}\,dt\right\|$.

Property 4.2.2. *$\hat{h} = at + b$ if and only if $\rho_{h_l} = 0$*

Proof. ($\Rightarrow$) If $\hat{h} = at + b$, then

$$\rho_{h_l} = \left\|\dot{\hat{h}} - \int_0^1 \dot{\hat{h}}\,dt\right\| = \left\|a - \int_0^1 a\,dt\right\| = \|a - a\| = 0.$$

($\Leftarrow$) If we have $.\rho_{h_l} = 0$, then

$$\left\|\dot{\hat{h}} - \int_0^1 \dot{\hat{h}}\,dt\right\| = 0 \Rightarrow \dot{\hat{h}} = \int_0^1 \dot{\hat{h}}\,dt = a \Rightarrow \hat{h} = at + b$$

□

Similarly, the test statistics for testing the presence, constant function, and linear function for the seasonality g are $\rho_{g_0} = \|\hat{g}\|$, $\rho_{g_c} = \left\|\hat{g} - \int_0^1 \hat{g}\,dt\right\|$, and $\rho_{g_l} = \left\|\dot{\hat{g}} - \int_0^1 \dot{\hat{g}}\,dt\right\|$ respectively.

4.2.2 Bootstrap Cross-Sectional Confidence Band

In addition to providing point estimates of h and g in their respective subspaces, one can use bootstrap to provide a confidence region associated with these estimates. The basic idea is to take bootstrap replicates of the estimator and use the $\mathbb{L}^2$ norm to build confidence regions around the estimate, for either h and g. Since under the $\mathbb{L}^2$ metric, the mean of functions corresponds to a cross-sectional mean, this task simplifies to building a confidence interval at each time t from bootstrap replicates. Let $\bar{h}$ and $\bar{g}$ be the bootstrap averages, and $\hat{se}_h$ and $\hat{se}_g$ be respectively the bootstrap estimates of the standard errors of $\hat{h}$ and $\hat{g}$ (obtained from Algorithm 4) as functions of t. For a significance level α, the confidence interval for $\hat{h}(t)$ is simply $[\bar{h}(t) \pm z^{1-\alpha/2} \cdot \hat{se}_h(t)]$, where $z^{1-\alpha/2}$ is the $100 \cdot (1 - \alpha/2)$th percentile point of a standard normal distribution. Similarly, the confidence interval for the estimated seasonal effect $\hat{g}(t)$ is simply $[\bar{g}(t) \pm z^{1-\alpha/2} \cdot \hat{se}_g(t)]$.

CHAPTER 5

TREND SUBSPACE SELECTION

The current framework provides a criterion for choosing between potential candidates for $\mathcal{H} = \text{span}\{\phi_k | k = 1, \ldots, l\}$ which also defines $\mathcal{G} = \mathcal{H}^{\perp}$, by simply maximizing the likelihood under each candidate subspace and selecting the one that results in the highest maximized-likelihood. In other words, given several candidate subspaces, we compare their minimized values of the cost functional, Equation 2.6. Empirical evidence shows the best estimates results from the subspace with the smallest minimized values.

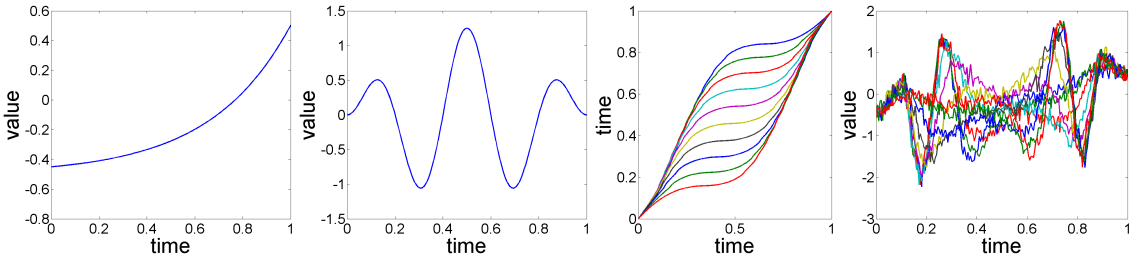
5.1 Basis Range Selection

We assumed that the subspace $\mathcal{H}$ is of the form $\mathcal{H} = \text{span}\{\phi_k | k = 1, \ldots, l\}$ for some positive integer $1 \leq l < \infty$. Thus, choosing $\mathcal{H}$ becomes choosing an appropriate l. With this setting, we can try each potential value of l, up to a certain large value, maximize the likelihood under each choice of l, and select the one with the highest value of the likelihood.

We demonstrate this idea using a synthetic data example, as demonstrated earlier in Figure 1.6. In this experiment, we generate data using
$g = 5\left(0.25 - (t - 0.5)^2\right)\sin(5\pi t)$, $h = 0.05e^{3t} - 0.5$, and $\gamma_i = \int_0^t \check{\gamma}_i dt / \int_0^1 \check{\gamma}_i dt$ where

$$\check{\gamma}_i = \left(3\cos(\pi t - 0.5 + \frac{i}{n})\right)^2 + 0.1$$

for $i = 1, ..., n$, with the additional constraint $KM\{\gamma_i^{-1}\} = \gamma_{id}$ imposed on $\{\gamma_i\}$. We add noise according to $\epsilon_i(t) \sim \mathcal{N}(0, \sigma^2)$, $\sigma = 0.1$. The resulting data are shown in Figure 5.1(d).



(a) true trend (b) true seasonality (c) warping functions (d) observed functions

Figure 5.1: Synthetic ground truth data for range number selection experiment.

The results of applying Algorithm 3 to this data for $l = 1, 2, 3, 4, 5, 8$, and 10, and presented in the Figure 5.3. The ϕs used in this experiment are the shifted Legendre polynomial, and the trend $\hat{h}$ is thus a polynomial of degree l. From a visual perspective, setting $l = 4$ yields the best estimates of g and h. Figure 5.2 displays the minimized negative log-likelihood for $l = 1$ to $l = 10$ and the optimal is obtained when $l = 3$, supporting our approach for selecting l. $\mathcal{H}$ with $l = 3$ and $l = 4$ are predicted to be similar in performance which is reflected in the results of Figure 5.3. This selection rule can be generalized to situations when several potential basis types (polynomial, sine, cosine, Fourier) are given (see Section 5.2).

Table 5.1 summarizes the numerical behavior of the $\mathbb{L}^2$ cost, relative error of g, h, and absolute error of γ_i. From Table 5.1, the lowest $\mathbb{L}^2$ cost not only gives the best recovered $g, h, \{\gamma_i\}$ visually, but also has the smallest $\mathbb{L}^2$ error of g, h, and $\{\gamma_i\}$.

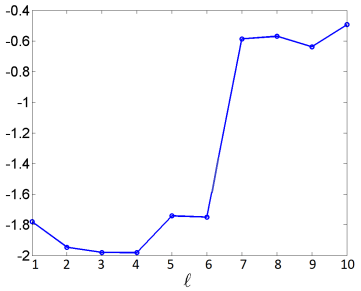


Figure 5.2: Minimized negative log-likelihood for $l = 1, 2, ..., 10$, plotted on a log scale.

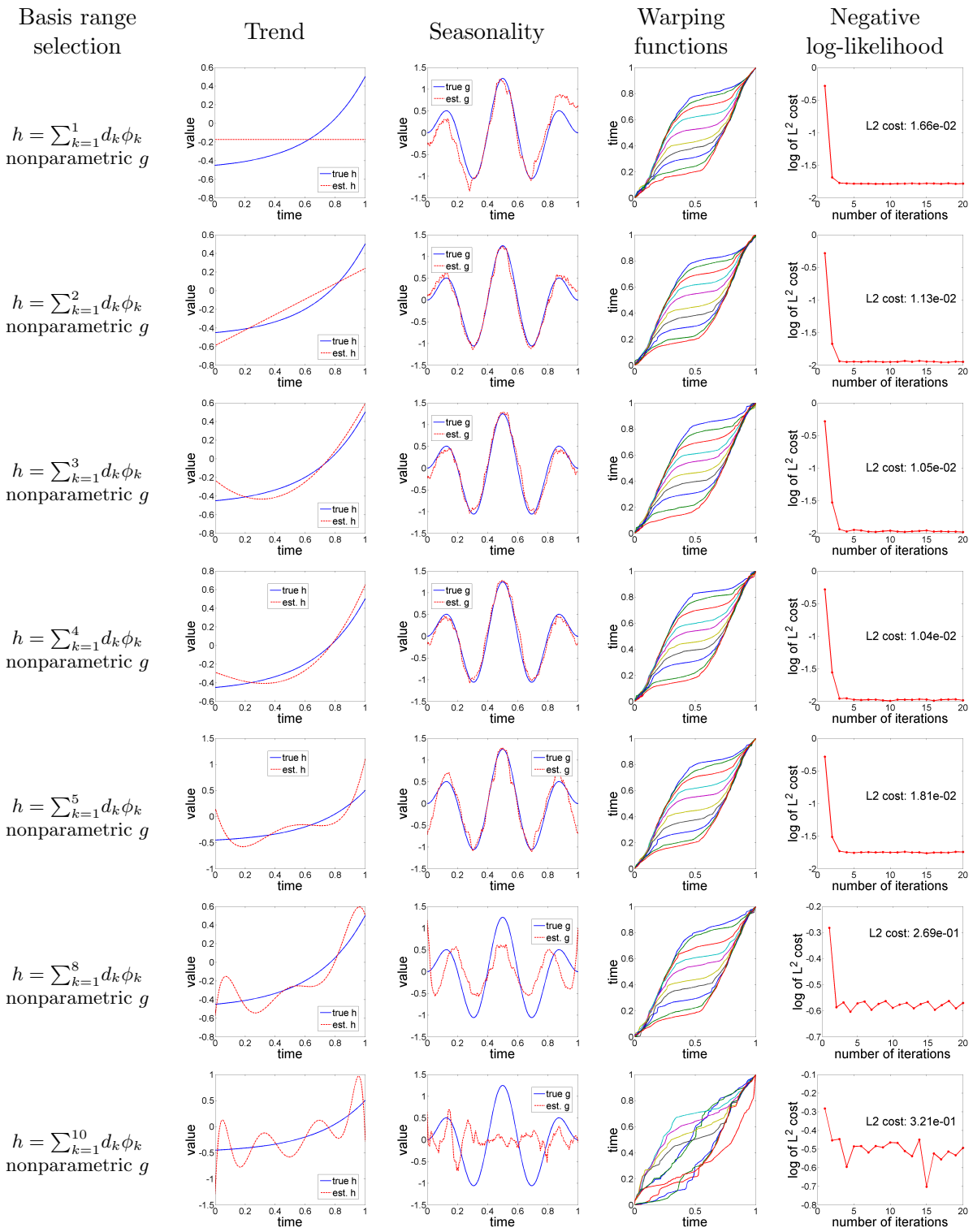


Figure 5.3: Numerical results for basis range selection experiment. Figures in the column of negative negative log-likelihood are plotted on a log scale and the number inside the figures are the minimized negative log-likelihood at the 20th iteration .

Table 5.1: $\mathbb{L}^2$ Cost comparison of basis range selection experiment. Columns from left to right are basis range selection, $\mathbb{L}^2$ cost, relative error of the trend, relative error of the seasonality, average absolute error of warping functions, and orthogonality of seasonality and trend.

Basis range selection	$\mathbb{L}^2$ cost $\frac{1}{n}\sum_{i=1}^{n}\|lfs - rhs\|^2$	h relative error $\left\|\hat{h}-h\right\| / \|h\|$	g relative error $\|\hat{g}-g\| / \|g\|$	γ_i abs. error $\frac{1}{n}\sum_{i=1}^{n}\|\gamma_i - \hat{\gamma}_i\|$	Orthogonality $\langle g(t), h(t)\rangle_{\mathbb{L}^2} = 4.29336e-03$
$h=\sum_{k=1}^{1} d_k\phi_k$ nonparametric g	3.92545e-02	6.29738e-01	1.20758e+00	5.97033e-02	7.06842e-17
$h=\sum_{k=1}^{2} d_k\phi_k$ nonparametric g	7.79053e-03	2.21529e-01	1.33989e-01	1.84620e-03	-2.9023e-17 (**lowest**)
$h=\sum_{k=1}^{3} d_k\phi_k$ nonparametric g	1.27729e-03	6.35327e-02	4.66616e-02	8.35612e-04	-4.9416e-14
$h=\sum_{k=1}^{4} d_k\phi_k$ nonparametric g	7.08971e-04 (**lowest**)	3.63075e-02 (**lowest**)	3.95206e-02 (**lowest**)	6.09604e-04 (**lowest**)	-6.8298e-14
$h=\sum_{k=1}^{5} d_k\phi_k$ nonparametric g	1.03965e-01	3.17688e-01	4.86339e-01	5.39322e-03	-9.5457e-08
$h=\sum_{k=1}^{6} d_k\phi_k$ nonparametric g	1.77772e-02	5.16760e-01	3.01323e-01	9.72263e-03	-1.6256e-10
$h=\sum_{k=1}^{7} d_k\phi_k$ nonparametric g	2.59025e-01	3.54250e-01	7.88289e-01	2.38560e-02	-2.3940e-09
$h=\sum_{k=1}^{8} d_k\phi_k$ nonparametric g	2.69130e-01	4.17224e-01	7.87592e-01	1.40076e-02	-4.9792e-09
$h=\sum_{k=1}^{9} d_k\phi_k$ nonparametric g	2.29605e-01	1.04054e+00	1.19400e+00	9.53862e-02	3.02811e-08
$h=\sum_{k=1}^{10} d_k\phi_k$ nonparametric g	3.20779e-01	8.43377e-01	1.07823e+00	1.25064e-01	-1.2564e-08

5.2 Selection of Basis Family

The second part of subspace $\mathcal{H}$ selection is the choice of the family of functions. We consider Legendre, cosine, sine, and Fourier bases. In Section 5, for a given basis family, the best value of l is the lowest $\mathbb{L}^2$ cost (or highest likelihood). We apply the same idea when determining the basis family. For each family, we find the best l and choose the best family by comparing the $\mathbb{L}^2$ costs for their l values.

We illustrate this idea with a noise-free synthetic data example. More systematic empirical evaluations of the robustness of the estimation in the presence of noise are given in Chapter 7. Let $g = \cos(10\pi t)$, $h = 1.5e^{-3t}$, and warping functions be $\gamma_i(t) = \frac{e^{a_i t}-1}{e^{a_i}-1}$ where $a_i = -3 + i\frac{6}{n}$ for $1 \le i \le n$ with $KM\{\gamma_i^{-1}\} = \gamma_{id}$ imposed. Observations $\{f_i(t)\}$ are generated from

$$f_i(t) = h(t) + (g, \gamma_i)(t) + \epsilon_i(t) = h(t) + (g \circ \gamma_i)(t)\sqrt{\dot{\gamma}_i(t)} + \epsilon_i(t).$$

Synthetic data are summarized in Figure 5.4.

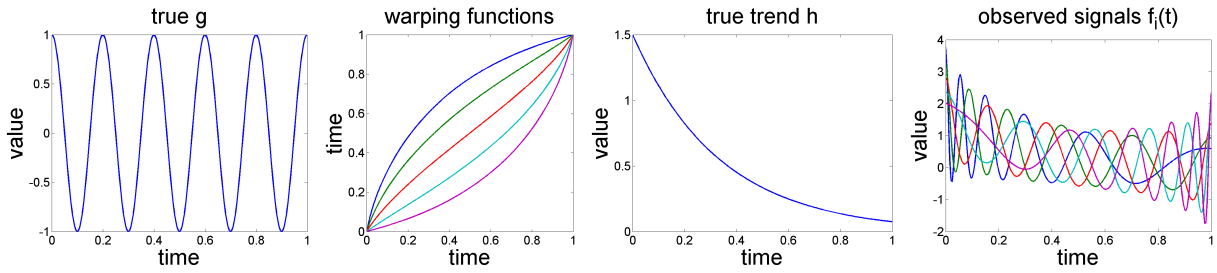


Figure 5.4: Synthetic data of basis family selection experiment

Our trend and seasonality estimation algorithm yields the best l for each basis family: shifted Legendre basis $l = 4$, Fourier basis picks $l = 5$, sine basis $l = 8$, and cosine basis $l = 8$. Figure 5.5 contains estimation results of lowest $\mathbb{L}^2$ cost of each basis family. We see that the the Legendre and cosine bases have similarly good estimates of g, h, γ_i but not Fourier and Sine basis. It is hard to visually determine whether the Legendre or cosine basis is better, so more careful comparison of their numerical behavior is needed.

Table 5.2 quantifies the difference between these basis families. The cosine basis has the smallest $\mathbb{L}^2$ cost with desired side effects - smallest relative error of g and h and we will take cosine basis as the best basis choice for this experiment.

Table 5.2: Numerical comparison of estimation results from different basis families.

Basis family selection	$\mathbb{L}^2$ cost $\frac{1}{n}\sum_{i=1}^{n}\|\|lfh - rhs\|\|^2$	g relative error $\|\|g-\hat{g}\|\| \,/\, \|\|g\|\|$	h relative error $\left\|\left\|h-\hat{h}\right\|\right\| \,/\, \|\|h\|\|$	γ_i absolute error $\frac{1}{n}\sum_{i=1}^{n}\|\|\gamma_i - \hat{\gamma}_i\|\|$	Orthogonality $\langle g(t), h(t)\rangle = 4.293368e-03$
nonpara. g sLegendre h	1.49524e-03	5.13135e-02	3.20385e-02	8.45832e-04 (**lowest**)	-4.95299e-14 (**lowest**)
nonpara. g MFS h	3.52676e-01	1.20394e+00	3.09338e-01	2.13350e-01	-5.95870e-03
nonpara. g sine h	2.56641e-01	1.03715e+00	4.02526e-01	6.70850e-02	-9.37360e-05
nonpara. g cosine h	1.12074e-03 (**lowest**)	3.39472e-02 (**lowest**)	3.19216e-02 (**lowest**)	1.02762e-03	7.35131e-05

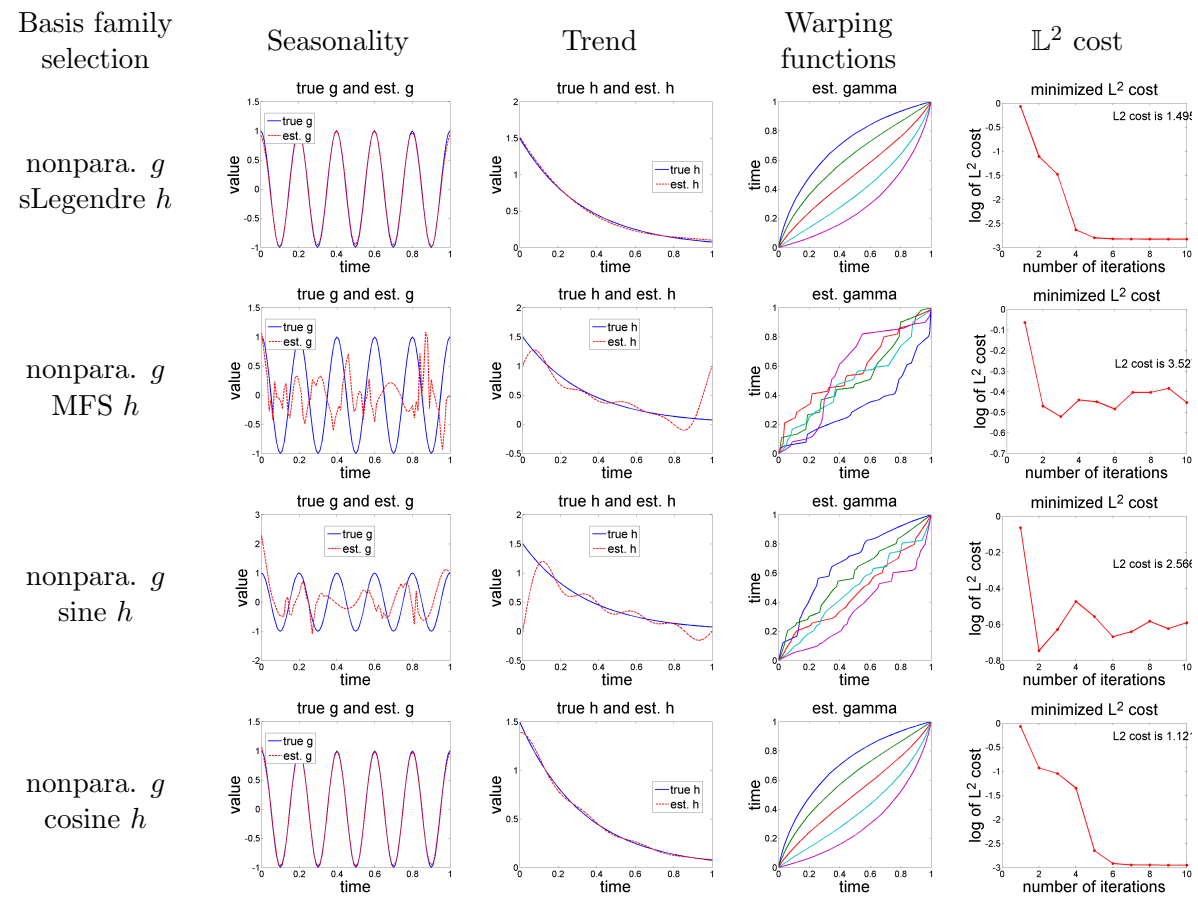


Figure 5.5: Estimation results of different basis families.

CHAPTER 6

RELATED MODELS

In this chapter, we compare our model with several related models. For each model, the motivation, assumptions and method of solution are discussed.

6.1 Models with Trend and Seasonality

6.1.1 Simple Separation Model (Simple Sep.)

The simplest possible model is additive without warping defined as

$$f_i(t) = h(t) + g(t) + \epsilon_i(t) \tag{6.1}$$

where $\epsilon_i(t)$ is Gaussian noise with mean zero for all i and t. A a unique decomposition of $f_i(t)$ into $g(t)$ and $h(t)$ results simply if the direct sum assumption $\mathbb{L}^2 = \mathcal{G} \oplus \mathcal{H}$ is imposed. By the same maximum likelihood reasoning as used in Section 2.2.4, the cost functional of this model is

$$C(g,h) = \frac{1}{n}\sum_{i=1}^{n} \|f_i - h - g\|^2 . \tag{6.2}$$

with direct sum condition $\{f_i\} \in \mathbb{L}^2 = \mathcal{G} \oplus \mathcal{H}$. Due to the direct sum assumption, estimates of h and g are recovered using one step of coordinate descent described in Chapter 3. Specifically, given estimate the $\hat{h}$, the estimate for g is updated to

$$\hat{g} = \frac{1}{n}\sum_{i=1}^{n}\left(f_i - \hat{h}\right) - \Pi_{\mathcal{H}}\frac{1}{n}\sum_{i=1}^{n}\left(f_i - \hat{h}\right), \tag{6.3}$$

and given the estimate $\hat{g}$, the estimate for h is updated to

$$\hat{h} = \Pi_{\mathcal{H}}\left(\frac{1}{n}\sum_{i=1}^{n} f_i\right). \tag{6.4}$$

There are two main disadvantages for this model. The first is its simplicity. Without warping it is unable to model the assumed variable-phase seasonality. The second disadvantage follows from its simplicity and the direct sum assumption. Since optimizing the cost function is done by two independent projections once $\mathcal{H}$ is chosen, the value of the minimized cost functional is the same for all choices of space $\mathcal{H}$. As a result, our method to choose the best subspace $\mathcal{H}$ is not applicable.

6.1.2 Trend and Seasonality Estimation with Alignment Warping Model (Trend-Warp)

The physical and mathematical invariants of the trend and seasonality problems of interest are important considerations when choosing the model of warping and trend interaction. For some problems the constraint that the warping does not change the height of $g(t)$ is reasonable. Along with an additive trend this leads to proposing the model

$$\begin{aligned} f_i(t) &= h(t) + (g \circ \gamma_i)(t) + \epsilon_i(t) \\ &= h(t) + g(\gamma_i(t)) + \epsilon_i(t) \end{aligned}$$

where $\varepsilon_i(t)$ is Gaussian noise with mean zero for all i and t. The direct sum $f_i \in \mathbb{L}^2 = \mathcal{G} \oplus \mathcal{H}$ and the Karcher mean of $\left\{\gamma_i^{-1}\right\} = \gamma_{id}$ are also imposed. Similar to Section 2.2.4, maximum likelihood reasoning can be used to derive the cost functional

$$C(g, h, \{\gamma_i\}) = \frac{1}{n}\sum_{i=1}^{n} \|f_i - h - (g \circ \gamma_i)\|^2 . \tag{6.5}$$

This proposed model has some difficulties. The first concerns the use of the coordinate descent method to optimize the cost functional, Equation 6.5. The $\mathbb{L}^2$-norm is not invariant under warping with composition. As a result, γ_i or its current estimate cannot be moved from $g(\gamma_i(t))$ term to $f_i - h$ term to allow the use of projection to update the estimate of g. Indeed, when the cost function is based on the $\mathbb{L}^2$-norm any optimization algorithm will have difficulties since the γ_i and g cannot be treated as independent variables for the cost function. The second and more important difficulty is that determining the optimal γ_i under $\mathbb{L}^2$-norm may cause the well-known pinching problem [RL98].

The pinching problem is addressed in the literature for the signal model $f_i(t) = (g \circ \gamma_i)(t) + \epsilon_i$ (see [SWK$^+$11]) by transforming the problem to the SRVF space. To exploit the SRVF technique, we update the proposed model to the **Trend-Warp** model defined as

$$f_i(t) = h(t) + (g \circ \gamma_i)(t) + \epsilon_i. \tag{6.6}$$

where the noise is a constant random variable ϵ_i follows $\mathcal{N}(0, \sigma)$.

We cannot simply apply SRVF transformation to Equation 6.6 since the addition in $h + (g \circ \gamma_i)$ is not preserved under SRVF transformation, i.e., we do not have $q_h + (q_g, \gamma_i)$ in the SRVF space.

The trend must be removed from in order to apply the SRVF-based approach. So the first step in the estimation algorithm for the Trend-Warp model is to estimate $h(t)$ by

$$\hat{h}(t) = \Pi_{\mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} f_i(t). \tag{6.7}$$

This is equivalent to assuming that $\mathcal{G} \circ \Gamma = \mathcal{H}^{\perp}$. Of course, as discussed earlier, this is not exactly true so one would expect this model to suffer degradation in estimation accuracy when there are several $(g \circ \gamma_i)$ with significant components not in $\mathcal{H}^{\perp}$.

Once an estimate $\hat{h}$ for the trend h is known, the model (Equation 6.6) simplifies to the alignment estimation problem in [SWK$^+$11]. That is, given observations $\{\tilde{f}_i\}$ of the form

$$\tilde{f}_i(t) = f_i(t) - \hat{h}(t) = (g \circ \gamma_i)(t) + \epsilon_i,$$

we attempt to recover g and $\{\gamma_i\}$. Rather than apply the algorithm of [SWK$^+$11], we optimize the cost functional

$$C(g, \gamma_i) = \frac{1}{n} \sum_{i=1}^{n} d_{FR}^2 \left(\tilde{f}_i, (g \circ \gamma_i) \right) \tag{6.8}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left\| q_{\tilde{f}_i} - q_g(\gamma_i) \sqrt{\dot{\gamma}_i} \right\|^2 \tag{6.9}$$

where $d_{FR}\left(\tilde{f}_i, g \circ \gamma_i \right)$ is the geodesic distance between $\tilde{f}_i$ and $(g \circ \gamma_i)$ under the Fisher-Rao Riemannian metric and Equation 6.9 is derived from Equation 6.8 using the SRVF transformation [JKSJ07, SWK$^+$11] which simplifies d_{FR} into the $\mathbb{L}^2$-norm.

As before, the coordinate descent method is used to estimate g and $\{\gamma_i\}$. Given the SRVF $q_{\hat{g}}$ of the estimate $\hat{g}$, the estimates of the $\{\gamma_i\}$ are updated by minimizing

$$\tilde{\gamma}_i = \underset{\gamma_i \in \Gamma}{\arg\min} \left\| q_{f_i - h} - q_{\hat{g}}(\gamma_i) \sqrt{\dot{\gamma}_i} \right\|^2, \; i = 1, ..., n.$$

The constraint $KM\{\gamma_i^{-1}\} = \gamma_{id}$ is then imposed on the $\{\tilde{\gamma}_i\}$ to finalize the update to $\{\hat{\gamma}_i\}$. Given the estimates $\{\hat{\gamma}_i\}$, the estimate of the SRVF q_g is given by

$$\tilde{q}_g = \frac{1}{n} \sum_{i=1}^{n} \left(q_{f_i - \hat{h}}, \hat{\gamma}_i^{-1} \right).$$

After the iterations of updates to the γ_i and q_g satisfy a convergence criterion, we apply the inverse SRVF transformation to $\tilde{q}_g$,

$$\begin{aligned}\tilde{g}(t) = Q^{-1}(\tilde{q}_g)(t) &= \int_0^t \tilde{q}_g(s)|\tilde{q}_g(s)|\,ds + \hat{g}(0) \\ &\approx \int_0^t \tilde{q}_g(s)|\tilde{q}_g(s)|\,ds + \frac{1}{n}\sum_{i=1}^{n}\left[f_i(0) - \hat{h}(0)\right].\end{aligned}$$

to recover an estimate of the seasonality g. Finally, the direct sum condition $f_i \in \mathbb{L}^2 = \mathcal{G} \oplus \mathcal{H}$ is imposed by the projection

$$\hat{g} = \tilde{g} - \Pi_{\mathcal{H}}\tilde{g}.$$

This model is perhaps the most intuitive for the problem of estimating trend and variable-phase seasonality since the warping function γ_i does not change the height of the seasonality g and is simple reparamterization, i.e. $g(\gamma_i)$ instead of (g, γ_i) in our model. However, the extra assumption $\mathcal{G} \circ \Gamma = \mathcal{H}^{\perp}$ we imposed is, in general, not satisfied and this will affect the expected accuracy of the estimation.

6.1.3 Trend and Seasonality Estimation with Area-Preserving Warping Model (Area-Warp)

Motivated from the Trend-Warp model in which the warping γ_i does not change the height of g, we can develop a model with a trend and warping where the warping preserves other aspects of g. For example, to preserve the area we can exploit the identity $\int g(t)dt = \int g(\gamma_i(t))\dot{\gamma}_i(t)\,dt$ to define the the Area-Warp model as

$$f_i(t) = g(\gamma_i(t))\dot{\gamma}_i(t) + h(t) + \epsilon_i \tag{6.10}$$

where ε_i is Gaussian noise with mean zero. The two conditions, direct sum and $KM\left\{\gamma_i^{-1}\right\} = \gamma_{id}$, are also maintained.

If the cost functional is taken as

$$C(g, \{\gamma_i\}, h) = \frac{1}{n}\sum_{i=1}^{n}\left\|f_i - g(\gamma_i)\dot{\gamma}_i - h\right\|^2,$$

there will be, as before, difficulty using the coordinate descent method or any other optimization approach based on the $\mathbb{L}^2$-norm due to lack of invariance under warping acting on g.

We make an extra condition $(\mathcal{G} \circ \Gamma)\,\Gamma = \mathcal{H}^{\perp}$ to the Area-Wapr model.

Because of this extra assumption, h can be immediately estimated by

$$\hat{h}(t) = \Pi_{\mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} f_i(t). \tag{6.11}$$

Once we have the estimate $\hat{h}$, the Area-Warp model, Equation 6.10, is reduced to estimating g and $\{\gamma_i\}$ in

$$\tilde{f}_i(t) = f_i(t) - \hat{h}(t) = g(\gamma_i(t))\dot{\gamma}_i(t) + \epsilon_i. \tag{6.12}$$

To utilize the warping $(g \circ \gamma)(t)\sqrt{\dot{\gamma}(t)}$ which has $\mathbb{L}^2$-norm invariance, we take the square root to Equation 6.12

$$\begin{aligned} \sqrt{f_i(t) - \hat{h}(t)} &= \sqrt{g(\gamma_i(t))\dot{\gamma}_i(t) + \varepsilon_i} \\ &= ((\sqrt{g} \circ \gamma_i)(t)) \sqrt{\dot{\gamma}_i(t)} + \tilde{\varepsilon}_i, \end{aligned}$$

to get the cost functional

$$C(g, \{\gamma_i\}) = \frac{1}{n} \sum_{i=1}^{n} \left\| \sqrt{f_i - \hat{h}} - (\sqrt{g} \circ \gamma_i) \sqrt{\dot{\gamma}_i} \right\|^2 \tag{6.13}$$

with the direct sum assumption and Karcher mean condition $KM\{\gamma_i^{-1}\} = \gamma_{id}$ constraints.

This cost functional, Equation 6.13, is solved by the coordinate descent method. Given the estimate $\hat{g}$, updating estimates $\{\gamma_i\}$ first solves the unconstrained optimization

$$\tilde{\gamma}_i = \underset{\gamma_i \in \Gamma}{\arg\min} \left\| \sqrt{f_i - \hat{h}} - \left((\sqrt{\hat{g}} \circ \gamma_i) \right) \sqrt{\dot{\gamma}_i} \right\|^2, \; i = 1, ..., n$$

then imposes $KM\{\gamma_i^{-1}\} = \gamma_{id}$ on unconstrained solutions $\{\tilde{\gamma}_i\}$ to have $\{\hat{\gamma}_i\}$. The estimate for g is then updating by the projection

$$\hat{g}(t) = \tilde{f}(t) - \Pi_{\mathcal{H}} \tilde{f}(t)$$

where

$$\tilde{f} = \left[\frac{1}{n} \sum_{i=1}^{n} \left(\sqrt{f_i - \hat{h}}, \hat{\gamma}_i^{-1} \right) \right]^2.$$

We implicitly assume both $\sqrt{f_i - \hat{h}}$ and $\sqrt{\hat{g}}$ are always positive when applying the square root on Equation 6.12. The positive assumption becomes a limitation because the parameter g and $\sqrt{f_i - h}$ may not be positive. Secondly, the extra assumption $(\mathcal{G} \circ \Gamma)\,\Gamma = \mathcal{H}^{\perp}$ does not hold, in general, and imposing it affects the accuracy of the estimates. because the probability of $(\mathcal{G} \circ \Gamma)\,\Gamma$

stays in $\mathcal{G} = \mathcal{H}^{\perp}$ is extremely small. Because of all theoretical and algorithmic limitations, but mainly due to the difficulty in practice of maintaining positive arguments for the square root, we do not implement Area-Warp model and compare with other models.

6.1.4 Separation with same Warping Model (Sep. Warp)

This model is defined as

$$f_i(t) = ((g+h) \circ \gamma_i)(t) + \epsilon_i \tag{6.14}$$

where both the trend h and the seasonality g are affected by composition with the warping function γ_i. The cost functional is

$$C(g, \{\gamma_i\}, h) = \frac{1}{n}\sum_{i=1}^{n} d_{FR}^2(f_i,\, ((g+h) \circ \gamma_i)) \tag{6.15}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \|q_{f_i} - (q_{g+h}, \gamma_i)\|^2 \tag{6.16}$$

with two constraints $KM\{\gamma_i^{-1}\} = \gamma_{id}$ and direct sum $f_i \in \mathbb{L}^2 = \mathcal{G} \oplus \mathcal{H}$. We again apply coordinate descent method to solve the optimization, Equation 6.16. Suppose $\hat{g}$ and $\hat{h}$ are the current estimates, updating estimates of $\{\gamma_i\}$ of the unconstrained optimization first solves

$$\tilde{\gamma}_i = \underset{\gamma_i \in \Gamma}{\arg\min} \left\| q_{f_i} - q_{\hat{g}+\hat{h}}(\gamma_i)\sqrt{\dot{\gamma}_i} \right\|^2. \tag{6.17}$$

and then we imposes $KM\{\gamma_i^{-1}\} = \gamma_{id}$ on the unconstrained solutions $\{\tilde{\gamma}_i\}$ to get estimates $\{\hat{\gamma}_i\}$.

Given $\hat{\gamma}_i$, the updated q_{g+h} is given by

$$q_{\tilde{g}+\tilde{h}} = \frac{1}{n}\sum_{i=1}^{n} \left(q_{f_i}, \hat{\gamma}_i^{-1}\right).$$

We convert $q_{\tilde{g}+\tilde{h}}$ back to the time domain using the inverse SRVF transformation,

$$(\tilde{g}+\tilde{h})(t) = Q^{-1}\left(q_{\tilde{g}+\tilde{h}}\right)(t) = \int_0^t q_{\tilde{g}+\tilde{h}}(s)|q_{\tilde{g}+\tilde{h}}(s)|\, ds + (g(0)+h(0))$$
$$\approx \int_0^t q_{\tilde{g}+\tilde{h}}(s)|q_{\tilde{g}+\tilde{h}}(s)|\, ds + \frac{1}{n}\sum_{i=1}^{n} f_i(0).$$

Finally, to compute a coordinate iteration, projection imposes the direct sum assumption to get estimates

$$\hat{g} = \left(\tilde{g}+\tilde{h}\right) - \Pi_{\mathcal{H}}\left(\tilde{g}+\tilde{h}\right), \tag{6.18}$$

and

$$\hat{h} = \Pi_{\mathcal{H}}\left(\tilde{g} + \tilde{h}\right). \tag{6.19}$$

Similar to the Simple Sep. model, the Sep. Warp model has difficult selecting the best subspace $calH$.

6.2 Models with Seasonality Only

There are two related models that do not have the trend h and focus on estimating the seasonality g. They are the Simple Estimation Model and the Alignment Estimation Model.

6.2.1 Simple Estimation Model (Simple Est.)

The model is

$$f_i(t) = g(t) + \epsilon_i(t) \tag{6.20}$$

where $\epsilon_i(t)$ is Gaussian noise with mean zero for all i and t. We take the cost functional to be

$$C(g) = \frac{1}{n}\sum_{i=1}^{n} \|f_i - g\|^2 \tag{6.21}$$

and the estimate for g is $\hat{g} = \frac{1}{n}\sum_{i=1}^{n} f_i$.

A drawback of the simple estimation model is the absence of the time warping γ_i. In real data, the assumption of all curves are aligned is not true, as pointed out by [Ram06]. Instead, a better estimated seasonality can be obtained by incorporating time warpings $\{\gamma_i\}$ in the model.

6.2.2 Alignment Estimation Model (Align Est.)

In [SWK$^+$11] and [KSW11], the model is defined as

$$f_i(t) = (g \circ \gamma_i)(t) + \epsilon_i \tag{6.22}$$

where ϵ_i is Gaussian noice with mean zero for all i. For identifiability, the Karcher mean condition $KM\left\{\gamma_i^{-1}\right\} = \gamma_{id}$ is imposed. In the papers, the authors estimated seasonality g by computing the Karcher mean of orbits of g. However, we take an optimization approach and define the cost functional to be

$$C(g, \{\gamma_i\}) = \frac{1}{n}\sum_{i=1}^{n} d_{FR}^2(f_i,\, (g \circ \gamma_i)) = \frac{1}{n}\sum_{i=1}^{n} \left\| q_{f_i} - q_g(\gamma_i)\sqrt{\dot{\gamma}_i} \right\|^2 \tag{6.23}$$

where $d_{FR}\left(f_i,\ (g \circ \gamma_i)\right)$ is the geodesic distance between f_i and $(g \circ \gamma_i)$ under the Fisher-Rao Riemannian metric. The optimization problem is simplified from d_{FR} to an $\mathbb{L}^2$ cost functional (Equation 6.23) by taking the SRVF transformation [JKSJ07, SWK$^+$11].

We optimize Equation 6.23 by coordinate descent method. Given the estimate $\hat{g}$, we find the unconstrained solution $\tilde{\gamma}_i$ by finding

$$\tilde{\gamma}_i = \underset{\gamma_i \in \Gamma}{\arg\min} \left\| q_{f_i} - q_{\hat{g}}\left(\gamma_i\right) \sqrt{\dot{\gamma}_i} \right\|^2.$$

The condition $KM\{\gamma_i^{-1}\} = \gamma_{id}$ is then enforced to $\tilde{\gamma}_i$ to have $\hat{\gamma}_i$. When $\{\hat{\gamma}_i\}$ are given estimates, we estimate q_g by

$$q_{\hat{g}} = \frac{1}{n} \sum_{i=1}^{n} q_{f_i}\left(\hat{\gamma}_i^{-1}\right) \sqrt{\dot{\hat{\gamma}}_i^{-1}}.$$

We take the inverse SRVF transformation of q_g in order to go back to time domain, that is

$$\begin{aligned} \hat{g}(t) = Q^{-1}(q_{\hat{g}}) &= \int_0^t q_{\hat{g}}(s)|q_{\hat{g}}(s)|\, ds + g(0) \\ &\approx \int_0^t q_{\hat{g}}(s)|q_{\hat{g}}(s)|\, ds + \frac{1}{n}\sum_{i=1}^{n} f_i(0). \end{aligned}$$

Table 6.1 summarizes assumptions and disadvantages of our and related models.

Table 6.1: Summary of all models.

Number	Estimation	Name	Model	Cost functional	Assumption	Disadvantage
1	g and h	Simple Sep.	Eqn. 6.1	Eqn. 6.2	$\{f_i\} \in \mathbb{L}^2 = \mathcal{G} \oplus \mathcal{H}$	no warping no criterion for choosing $\mathcal{H}$
2	g, h, $\{\gamma_i\}$	Trend-L2	Eqn. 2.2	Eqn. 2.6	$\{f_i\} \in \mathbb{L}^2 = \mathcal{G} \oplus \mathcal{H}$ $KM\{\gamma_i^{-1}\} = \gamma_{id}$	none
3	g, h, $\{\gamma_i\}$	Trend-Warp	Eqn. 6.6	Eqn. 6.9	$\{f_i\} \in \mathbb{L}^2 = \mathcal{G} \oplus \mathcal{H}$ $\mathcal{G} \circ \mathcal{H} = \mathcal{H}^{\perp}$ $KM\{\gamma_i^{-1}\} = \gamma_{id}$	$\mathcal{G} \circ \mathcal{H} = \mathcal{H}^{\perp}$ is inappropriate
4	g, h, $\{\gamma_i\}$	Sep. Warp	Eqn. 6.14	Eqn. 6.16	$\{f_i\} \in \mathbb{L}^2 = \mathcal{G} \oplus \mathcal{H}$ $KM\{\gamma_i^{-1}\} = \gamma_{id}$	no criterion for choosing $\mathcal{H}$
5	g	Simple Est.	Eqn. 6.20	Eqn. 6.21	none	no warping
6	g, $\{\gamma_i\}$	Align. Est.	Eqn. 6.22	Eqn. 6.23	$KM\{\gamma_i^{-1}\} = \gamma_{id}$	none

CHAPTER 7

EXPERIMENTAL RESULTS OF SYNTHETIC DATA

In this chapter, we present experimental results for estimating the trend and seasonality using the MLE algorithm (Algorithm 3) including the selection of the basis family and the subspace dimension l (Chapter 5).

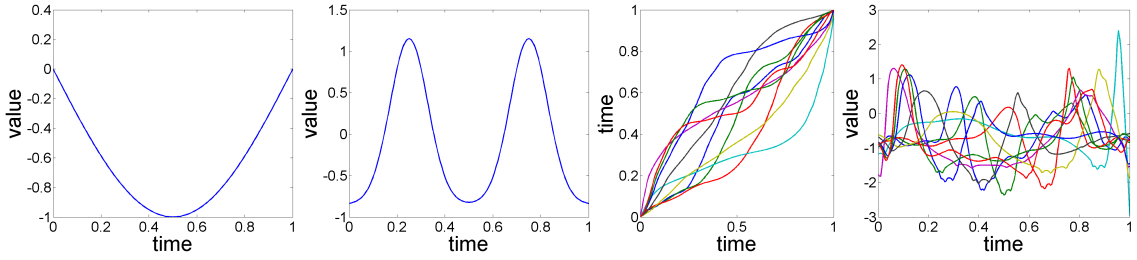
This chapter is organized as follows: Section 7.1 demonstrates estimation results of three components are robust under noise perturbation. In Section 7.2, we illustrate the use of bootstrap for testing hypothesis for different shapes of estimated trend and seasonality. The cross-sectional confidence bands for the estimated trend and seasonality are also given in this section. In Section 7.3, our model is compared with other models discussed in Chapter 6. We show that our model is the best model when estimating three components in the sense of best virtual fit and smallest error to the true data.

7.1 Performance under Different Noise Levels

In this experiment, we select a specific form of the trend and seasonal components, and increase the variance, σ^2, of the additive noise $\epsilon_i(t)$ to study the effect of noise on estimation performance. For synthetic data generation, we used $g = 2\exp\left(-0.8(10t-7.5)^2\right) + 2\exp\left(-0.8(10t-2.5)^2\right)$, $h = \cos(\pi t + \frac{\pi}{2})$, and warping functions $\gamma_i = \int_0^t \breve{\gamma}_i dt / \int_0^1 \breve{\gamma}_i dt$, where $\breve{\gamma}_i = 3\sin(2\pi a_i t) + 2\cos(a_i t), a_i = -2 + i\frac{4}{n}$. For each t, we consider Gaussian noise $\epsilon_i(t) \sim \mathcal{N}(0, \sigma^2)$ where σ is $0, 0.2, 0.4, 0.6, 0.8,$ and 1.6 for the noise level experiments. Each observation f_i is generated from Eqn. 2.2 and the number of time samples is taken to be $T = 200$. Figure 7.1 shows the true trend and the seasonal components, when no noise is added.

Recall a distortion percentage is provided in Section 2.2.2 that measures how γ acting on g destroys the direct sum assumption. We slightly relax the direct sum condition, $\mathcal{G} \perp \mathcal{H}$, by using $\langle g, h \rangle_{\mathbb{L}^2} = -0.0024$. Table 7.1 summarizes the distortion percentage under different basis range and basis families. Without knowing the right basis to be $\{\cos(\pi t + \frac{\pi}{2})\}$, we see there is always a

portion of (g, γ) inside the space $\mathcal{H}$. This portion increases as the dimension of $\mathcal{H}$ increases. The mixing of g and h by γ is nonlinear and significant and there is no simple way of separating them.



(a) true trend (b) true seasonality (c) warping functions (d) observed functions

Figure 7.1: Synthetic truth data of noise perturbation experiment.

Table 7.1: Distortion percentage, Equation 2.3, of different dimensions of space $\mathcal{H}$ in the noise perturbation experiment.

$\{\phi_k\}_{k=1}^{l}$	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=6$	$l=7$	$l=8$	$l=9$	$l=10$
cosine basis	14.59%	20.89%	29.12%	42.32%	57.66%	66.68%	72.82%	74.56%	79.81%	81.59%
sine basis	16.45%	34.28%	52.47%	52.62%	52.62%	68.91%	69.20%	71.59%	72.62%	74.23%
Fourier basis	14.59%	39.10%	55.49%	76.91%	84.03%	92.26%	92.96%	96.43%	97.76%	98.67%
Legendre	14.59%	18.25%	18.74%	38.27%	71.28%	72.30%	73.26%	84.80%	85.26%	86.57%

Table 7.2: Estimation error of noise perturbation experiment.

σ	$\mathbb{L}^2$ cost $\frac{1}{n}\sum_{i=1}^{n}\|lfh - rhs\|^2$	g relative error $\|g-\hat{g}\| / \|g\|$	h relative error $\left\|h-\hat{h}\right\| / \|h\|$	γ_i absolute error $\frac{1}{n}\sum_{i=1}^{n}\|\gamma_i - \hat{\gamma}_i\|$
0	4.741542e-04	1.151167e-02	4.092103e-03	1.037551e-03
0.2	3.736493e-02	1.037331e-01	1.581187e-02	1.189589e-02
0.4	1.495452e-01	1.939097e-01	3.094852e-02	1.916047e-02
0.6	3.780457e-01	3.093084e-01	4.359065e-02	3.863639e-02
0.8	6.509582e-01	3.646527e-01	6.746058e-02	3.478918e-02
1.6	2.834968e+00	1.301447e+00	4.079581e-01	1.284459e-01

As shown in Figure 7.2, the estimation results are very good when the noise level is relatively low ($\sigma \leq 0.4$), with the relative $\mathbb{L}^2$ error 1.04×10^{-1}, 1.58×10^{-2} and 1.19×10^{-2} for h, g, and $\{\gamma_i\}$ respectively. When the noise increases to $\sigma = 0.6$ and $\sigma = 0.8$, the reconstructed trends still have the desired pattern over $[0, 1]$ but the recovered seasonality results are not good. With large noise,

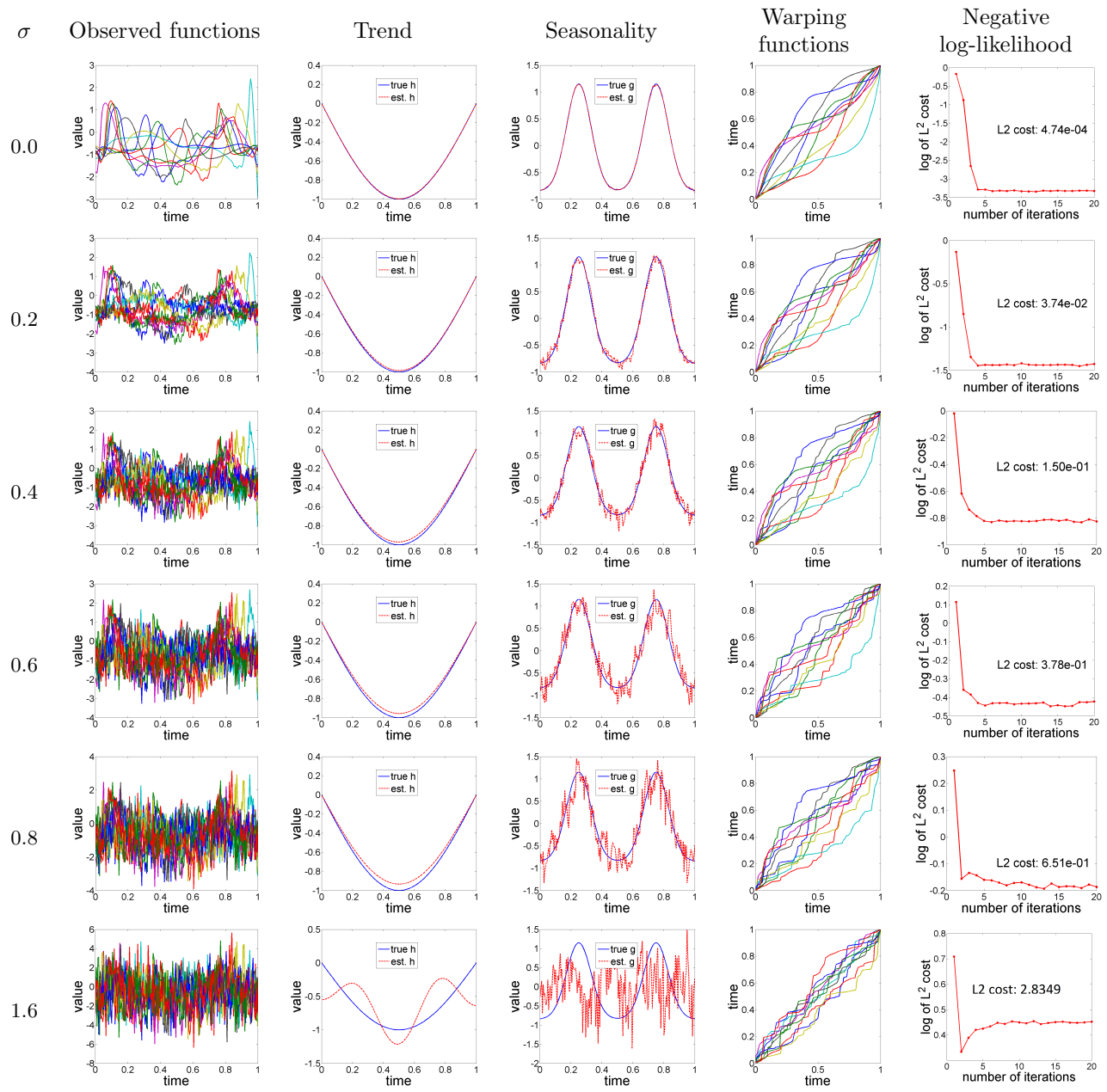


Figure 7.2: Numerical results of noise perturbation experiment. Note that figures in the column of negative log-likelihood are plotted in a log scale. The number inside their panels are minimized negative log-likelihood at the 20th iteration.

$\sigma = 1.6$, all the estimated results are far from their true values. Table 7.2 contains the estimation error of three components: trend, seasonality, and warping functions.

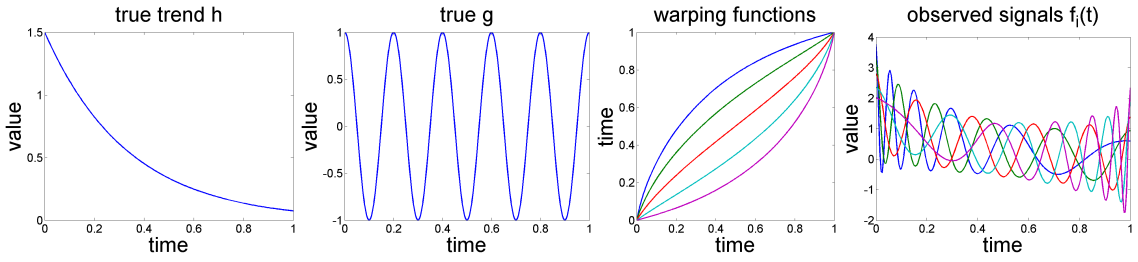
Table 7.2 and Figure 7.2 show that the estimation error is largest in $\hat{g}$ and smallest in $\hat{h}$. One

possible reason is the dimensionality and the property of the spaces themselves. Since the space $\mathcal{H}$ has the most restricted parametrized structure with fewest degree of freedom, it should have the smallest error. Secondly, space Γ is infinite dimensional with smoothness structure (diffeomorphism) and therefore the estimation error is larger then $\mathcal{H}$ but smaller than $\mathcal{G}$. Finally, the space $\mathcal{G}$ is left largely unconstrained and infinite dimensional so it has the largest estimation error.

These experiments provide evidence that the MLE algorithm with subspace selection can recover good estimates of the trend, seasonality and warping functions in the presence of some levels of noise.

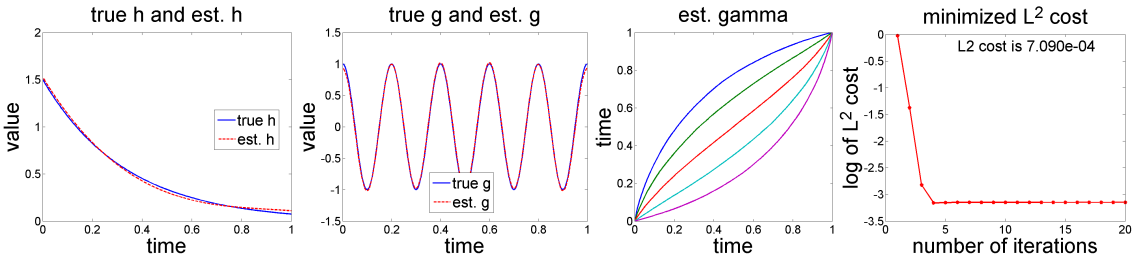
7.2 Bootstrap Analysis

In this experiment, we illustrate the use of a bootstrap technique for testing different hypotheses associated with the shapes of estimated trend and seasonal effect with cross-sectional bands. The general idea was described in Chapter 4. In this data, we use $g(t) = \cos(10\pi t)$, $h(t) = 1.5e^{-3t}$, and $\gamma_i(t) = \frac{e^{a_i t}-1}{e^{a_i}-1}, a_i = -3 + i\frac{6}{n}$ for $i = 1, ..., n$. Additionally, we set $\sigma = 0$. Figure 7.3 shows the synthetic data and the estimated trend and seasonality components are in Figure 7.4.



(a) true trend (b) true seasonality (c) warping functions (d) observed functions

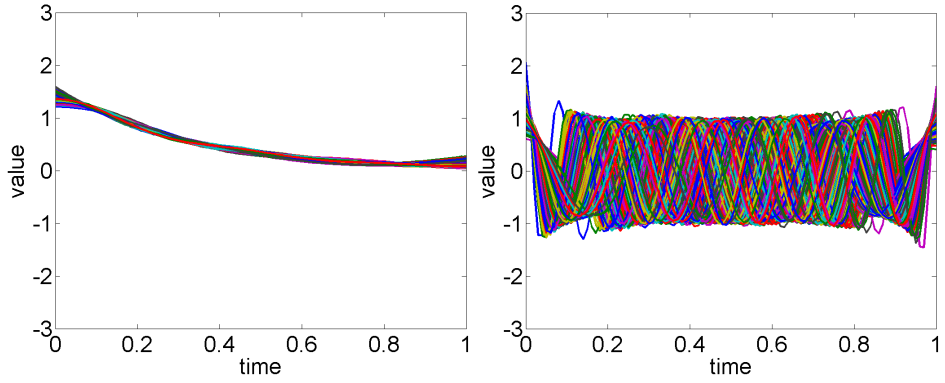
Figure 7.3: Synthetic ground truth of bootstrap experiment



(a) true trend (b) true seasonality (c) warping functions (d) observed functions

Figure 7.4: Estimated results of three components in bootstrap experiment.

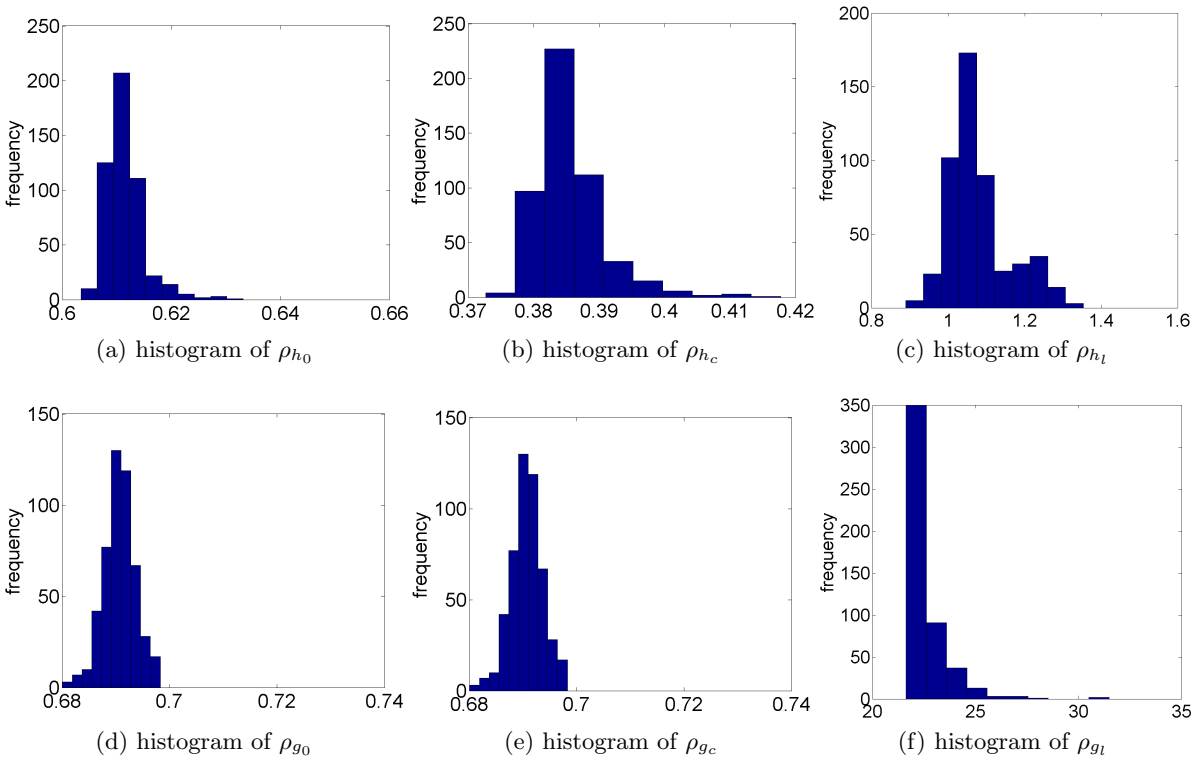
Figure 7.5, (a) and (b), shows the bootstrap replicates $\{\hat{h}_b\}$ and $\{\hat{g}_b\}$ of the trend and seasonality estimates for $B = 500$. As these replicates indicate, there is a significant phase variation in the replicates $\{\hat{g}_b\}$ and relatively small variability in the replicates $\{\hat{h}_b\}$.



(a) bootstrap replicates of $\hat{h}$ (b) bootstrap replicates of $\hat{g}$

Figure 7.5: Five hundred bootstrap replicate.

We presents histograms of test statistics ρ_{h_0},ρ_{h_c},ρ_{h_l},ρ_{g_0},ρ_{g_c} and ρ_{g_l} to study whether they follow the normal distribution $\mathcal{N}(0, \hat{se}_B)$ assumption, see Figure 7.6, (a)-(f). It can be seen that none of these test statistics have a mean zero so it is likely that the null hypothesis will be rejected.



(a) histogram of ρ_{h_0} (b) histogram of ρ_{h_c} (c) histogram of ρ_{h_l}

(d) histogram of ρ_{g_0} (e) histogram of ρ_{g_c} (f) histogram of ρ_{g_l}

Figure 7.6: Histograms of test statistics in testing the shapes of reconstructed results.

Given the bootstrap distribution of test statistics, we calculate the p values of these test statistics under different shape hypothesis:

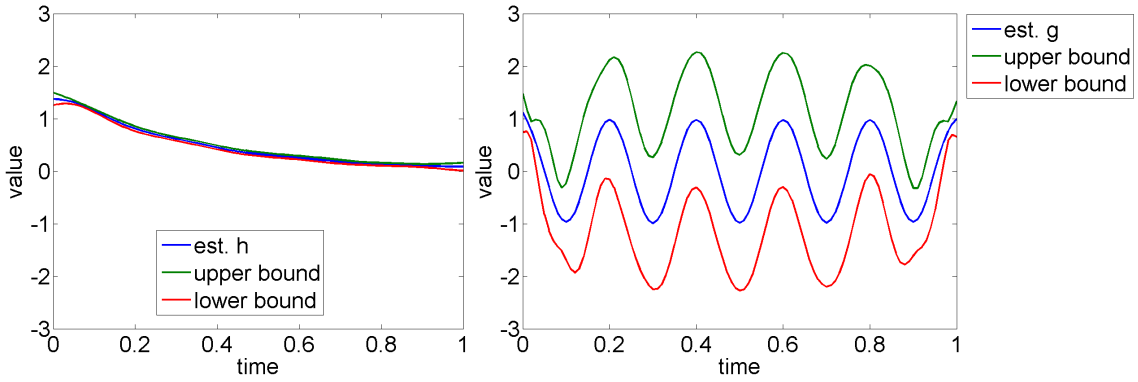
- **Testing presence of trend and seasonality:** For testing $h = 0$, the test statistic using bootstrap is $\rho_{h_0} = 0.61$ with $\hat{se}_B = 3.5 \times 10^{-3}$ and a p value of 0. The null hypothesis $h = 0$ is therefore rejected. For testing $g = 0$, the test statistic is $\rho_{g_0} = 0.38$ with $\hat{se}_B = 5.22 \times 10^{-3}$ and a p value of 0. The null hypothesis $g = 0$ is therefore rejected.
- **Testing constant shape of trend and seasonality:** For testing $h = c$, the test statistic using bootstrap is $\rho_{h_c} = 0.38$ with $\hat{se}_B = 5.2 \times 10^{-3}$ and a p value of 0. Therefore, we reject the null hypothesis: $h = c$. For testing $g = c$, the test statistic using bootstrap is $\rho_{g_c} = 0.69$ with $\hat{se}_B = 3 \times 10^{-3}$ and a p value of 0. Hence, we reject the null hypothesis: $g = c$.
- **Testing linear shape of trend and seasonality:** For the linearity of a trend, we obtain $\rho_{h_l} = 1.05$, $\hat{se}_B = 5.2 \times 10^{-3}$, and a p value of 0. Thus, we reject the null hypothesis: h is a linear function. For the seasonality g being a linear function, we have $\rho_{g_l} = 21.8$, $\hat{se}_B = 1.07$, and a p value of 0. The null hypothesis, g is a linear function, is rejected.

Table 7.3 summarizes conclusions of the trend and the seasonality under different hypothesis tests.

Table 7.3: Conclusions of bootstrap hypothesis testing of different shapes.

	Trend h	Seasonality g
H_0 : zero function	reject	reject
H_0 : constant function	reject	reject
H_0 : linear function	reject	reject

Based on the conclusion that both estimated trend and seasonality are not a lower order polynomial, we develop cross-sectional confidence bands with 95% confidence level in Figure 7.5, (a) and (b), indicating that the bandwidth of estimator $\hat{h}$ is much narrower than the $\hat{g}$.



(a) confidence band of $\hat{h}$ (b) confidence band of $\hat{g}$

Figure 7.7: cross-sectional confidence bands.

7.3 Model Comparison

In this section, our trend and seasonality estimation model is compared with the related models described in Chapter 6. Recall our MLE-based method is abbreviated as Trend-L2, and acronyms of other related models are defined in Table 6.1 in Chapter 6. When comparing across different models, we propose the following criteria for indicating a better model :

1. visual fitting to the true data,
2. lowest $\mathbb{L}^2$ cost (negative log-likelihood),
3. lowest absolute error of estimated trend (if a trend is needed),

4. lowest absolute error of estimated seasonality, and
5. ability to choose the best space $\mathcal{H}$ given several candidates.

The decision of whether a trend is needed is left to users. As mentioned in Table 6.1, we emphasize again that both the Simple Sep. model and Sep. Warp model have fundamental disadvantages because their estimation results are very sensitive and highly depend on the pre-determined space $\mathcal{H}$. This is a theoretical limitation, not an empirical result. Therefore, for the following experiments, these two models will use the best space $\mathcal{H}$ choice from the trend-L2.

Data generation is the same as the bootstrap analysis experiment in Section 7.2, shown in Figure 7.8 and Figure 7.9. Estimation results from applying our and other related models are given in Figure 7.10. Trend-L2 picks the cosine basis of the form $\sum_{k=1}^{8} d_k \phi_h$, and Trend-Warp takes the first element of sine basis. Simple separation and Sep. Warp, by definition, use the same choice $\mathcal{H}$ as Trend-L2.

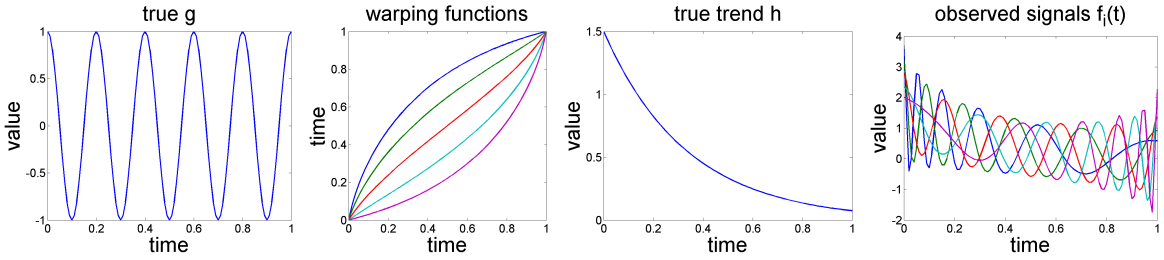


Figure 7.8: Synthetic ground truth data for method comparison experiment.

From Figure 7.10, recovered g, h, and $\{\gamma_i\}$ of Trend-L2 has the visual fitting to the true data while all other models have some off-fitting. Both simple separation and simple estimation models have poor performance because the lack of time warping parameters $\{\gamma_i\}$ when the observed data have time phase shifting. Alignment estimation model yields a decaying sinusoid seasonality g, but, by definition, does not return a trend estimate. The performance of alignment estimation is descent and compatible to our trend and seasonality model. Sep. Warp has a good approximation of seasonality g but the resulting trend h is not a good match to the true trend. Finally, Trend-Warp gives a sinusoid estimated seasonality g but with a concave down approximation of the trend h that is completely different from the true trend h. The performance of Sep. Warp and Trend-Warp are not as good as our Trend-L2 model.

Table 7.4 summarizes the performance of all models. Notice that Trend-L2 has the lowest $\mathbb{L}^2$ cost among all models, together with the lowest error of g, h, and $\{\gamma_i\}$. By the model comparison criteria discussed earlier, we conclude that our MLE-based algorithm (Trend-L2) is the best model since it satisfies the most criteria.

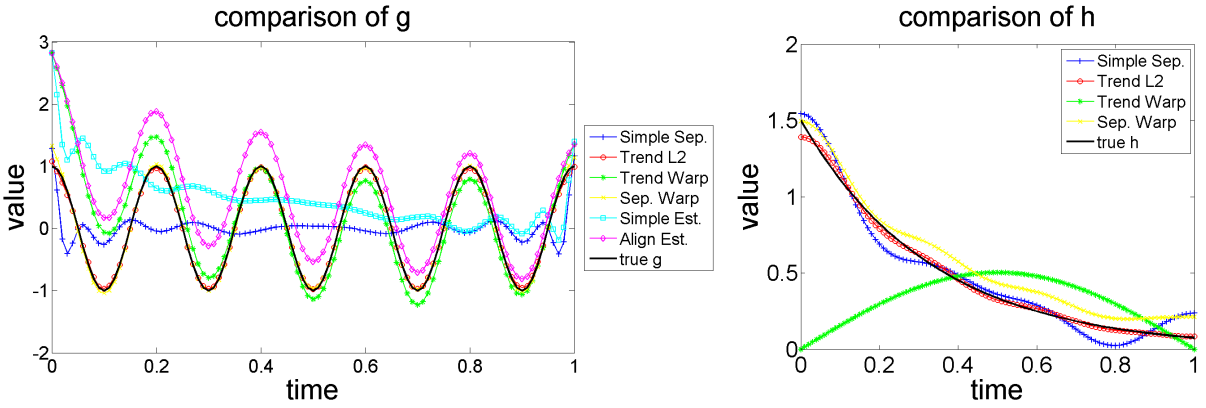


Figure 7.9: Trend (right) and seasonality (left) estimation results among all models.

Table 7.4: $\mathbb{L}^2$ cost comparison of all models of synthetic data.

Models	$\mathbb{L}^2$ cost $\frac{1}{n}\sum_{i=1}^{n}\|lfs - rhs\|^2$	g relative error $\|\hat{g}-g\| \,/\, \|g\|$	h relative error $\left\|\hat{h}-h\right\| / \|h\|$	γ_i absolute error $\frac{1}{n}\sum_{i=1}^{n}\|\gamma_i - \hat{\gamma}_i\|$	Orthogonality $\langle g(t), h(t)\rangle_{\mathbb{L}}^2 = 4.293368e-03$
Simple Separation $f_i(t) = g(t) + h(t) + \varepsilon_i(t)$	4.5233e-01	1.0013e+00	1.3145e-01	none	-4.751e-05
Trend-L2 $f_i(t) = g(\gamma_i(t))\sqrt{\dot{\gamma}_i(t)} + h(t) + \varepsilon_i(t)$	1.1207e-03 (**lowest**)	3.3947e-02 (**lowest**)	3.1921e-02 (**lowest**)	1.0276e-03 (**lowest**)	7.3513e-05
Trend Warp $f_i(t) = (g \circ \gamma_i)(t) + h(t) + \varepsilon_i$	5.7701e-01	7.0043e-01	7.9570e-01	1.7300e-03	-9.293e-07 (**lowest**)
Separation Warp $f_i(t) = ((g+h) \circ \gamma_i)(t) + \varepsilon_i$	9.7076e-02	5.5508e-02	1.6282e-01	1.8129e-03	1.1305e-06
Simple Estimation $f_i(t) = g(t) + \varepsilon_i(t)$	4.5233e-01	1.3342e+00	none	none	none
Alignment Estimation $f_i(t) = (g \circ \gamma_i)(t) + \varepsilon_i$	9.7076e-02	9.6778e-01	none	1.8129e-03	none

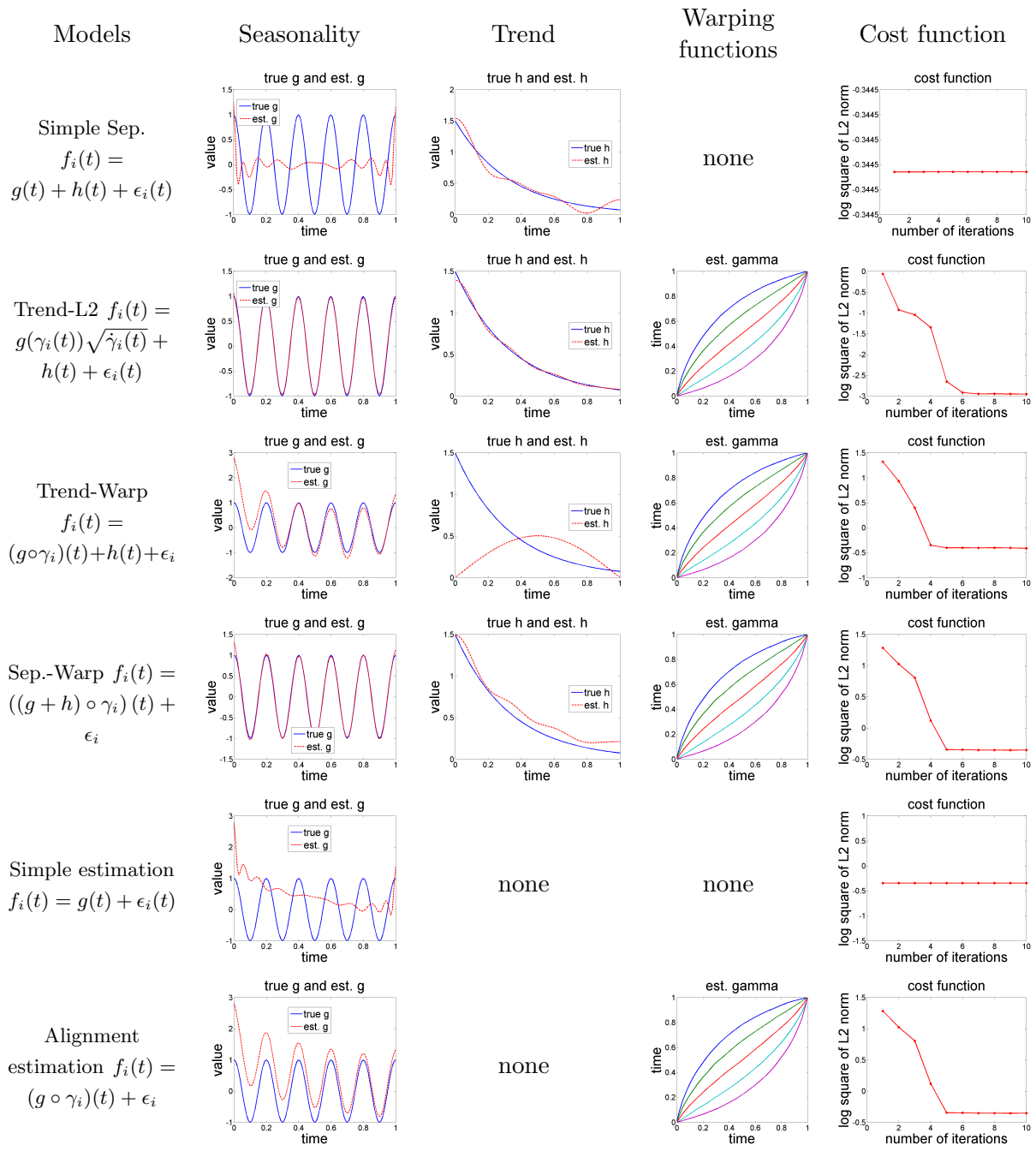


Figure 7.10: Estimation results of three components under different models.

CHAPTER 8

EXPERIMENTAL RESULTS OF REAL DATA

Chapter 8 presents estimation results of Trend-L2 and related models for three real datasets : Berkeley growth velocity (Section 8.1), electricity price (Section 8.2) and currency exchange rates (Section 8.3). Since truth is not available for these real data sets, as with the synthetic data experiments, we propose the following criteria for model comparison in the real data experiments:

1. resulting $\hat{g}$ and $\hat{h}$ fit the physical interpretations,
2. lowest $\mathbb{L}^2$ cost, and
3. when given candidates, the ability to choose the appropriate space $\mathcal{H}$.

8.1 Berkeley Male Growth Velocity

The Berkeley growth study of 54 female and 39 male was performed by Tuddenham and Snyder [TS54] and further discussed by Ramsay et al. [RS02]. As an example, we take 10 out of 39 male growth curve for analyzing. The raw data, height, was measured four times a year from new-born baby to 1 year old. After 1 year old, the height was measured annually from age 1 to age 8. Beyond 8 years old, the height was measured twice a year until 18 years old, shown in the left graph of Figure 8.1. Due to the monotonic nature of the curve, researchers often analyze the time derivative, termed the growth velocity, shown in the right graph of Figure 8.1.

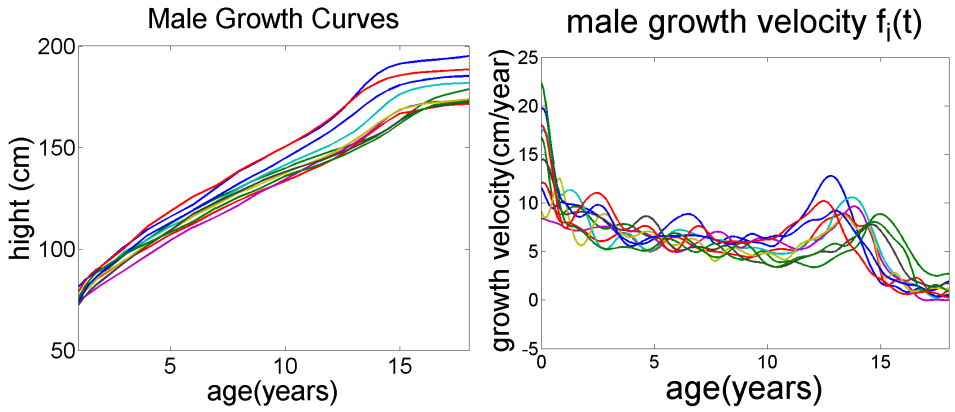


Figure 8.1: Male height (left) and growth velocity (right) data.

Looking at these growth curves, we see children grow the fastest with a speed of more than 10 centimeters per year before two years old. Another observation is the presence of a growth spurt between 10 to 15 years that is different for each child. By visually analyzing the growth velocity data, we expect that

1. the estimated seasonality g contains the fast growth under two years old and the another growth spurt around twelve years old,
2. the estimated trend h is monotone decreasing, and
3. the warping function γ_i reflects the different growth spurt timing of each child.

After applying our algorithm and the other related models, Trend-L2 chooses $\mathcal{H} = span\{1, \cos(\pi t)\}$ and Trend-Warp takes $\mathcal{H} = span\{\sin(\pi t), \sin(2\pi t)\}$. Simple separation and Sep.-Warp do not perform well unless a suitable space $\mathcal{H}$ is given. As explained earlier, for these two models we take $\mathcal{H}$ to be the same as that chosen by Trend-L2. Estimation results are shown in Figure 8.2 and Figure 8.3.

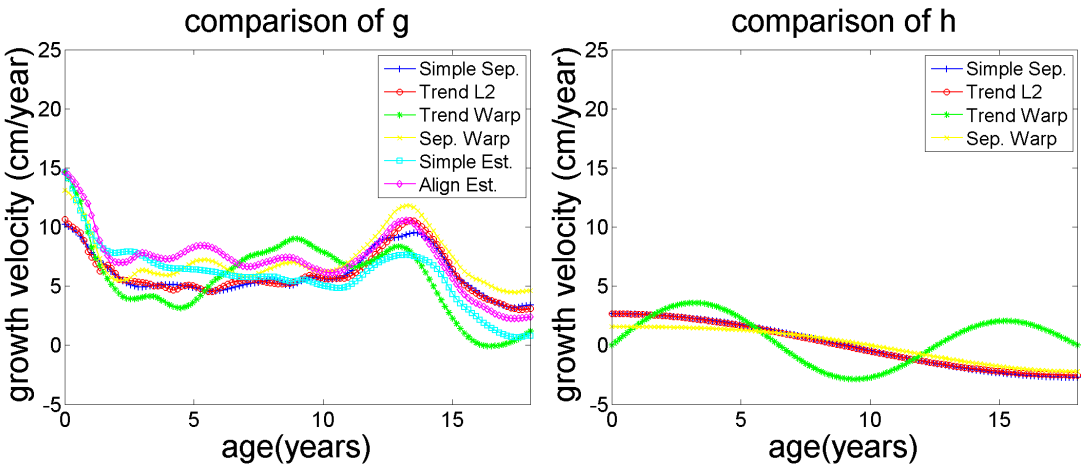


Figure 8.2: Estimation results of seasonality g and trend h among different models from male growth velocity data.

From these graphs, if no trend is sought, alignment estimation has a more accurate growth spurt $\hat{g}$ than simple estimation because alignment estimation includes the warping functions γ_i in the model. On the other hand, if a trend is sought, Trend-Warp model does not yield any useful $\hat{g}$ or $\hat{h}$ approximations, while Trend-L2 not only detects the growth spurt at around age 13, but also has the desired decaying trend. Simple separation and Sep.-Warp are able to produce estimates of

$\hat{g}$ and $\hat{h}$ with accuracy similar to Trend-L2 only when the subspace $\mathcal{H}$ used by Trend-L2 id used for them also.

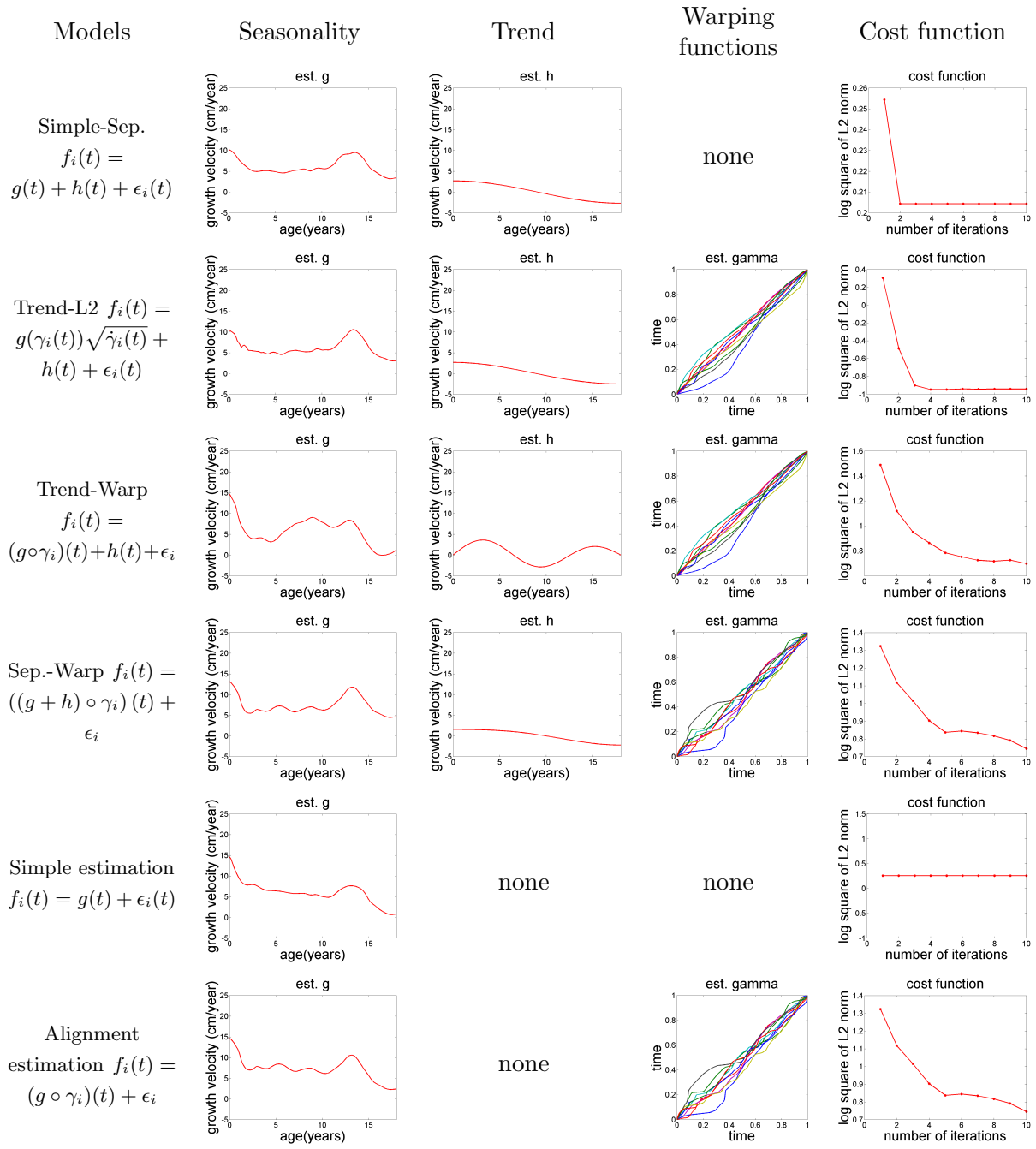


Figure 8.3: Estimation results of three componenst under different models for male growth velocity data.

Table 8.1 summarizes $\mathbb{L}^2$ costs of these models. Trend-L2 has the lowest $\mathbb{L}^2$ cost among all models. Based on the comparison criteria, Trend-L2 is the best model for the male growth velocity data.

Table 8.1: $\mathbb{L}^2$ cost comparison of all models of male growth velocity data.

Models	$\mathbb{L}^2$ cost $\frac{1}{n}\sum_{i=1}^{n}\|l.f.s. - r.h.s.\|^2$	Orthogonality $\langle g(t), h(t)\rangle_{L^2}$
Simple separation $f_i(t) = g(t) + h(t) + \epsilon_i(t)$	1.600453e+00	-1.252192e-06
Trend-L2 $f_i(t) = g(\gamma_i(t))\sqrt{\dot{\gamma}_i(t)} + h(t) + \epsilon_i(t)$	1.146533e-01 **(lowest)**	2.199665e-04
Trend-Warp $f_i(t) = (g \circ \gamma_i)(t) + h(t) + \epsilon_i$	1.780442e+01	-6.663887e-05 **(lowest)**
Sep.-Warp $f_i(t) = ((g + h) \circ \gamma_i)(t) + \epsilon_i$	3.470216e+00	8.908946e-05
Simple Estimation $f_i(t) = g(t) + \epsilon_i(t)$	1.806460e+00	none
Alignment Estimation $f_i(t) = (g \circ \gamma_i)(t) + \epsilon_i$	3.470216e+00	none

In the previous model comparison, our Trend-L2 was demonstrated to be the best model. We follow up with bootstrap analysis to verify the extracted trend and the seasonality are significant. The 500 bootstrap replicates for the estimated seasonality and trend are given in Figure 8.4. We see bootstrap replicates $\{\hat{g}_b\}$ have very much the same shape pattern. Each of them has a very high growth velocity under 2-years old followed by steady growth and a growth spurt at age 12 to 13-years old. On the other hand, although bootstrap replicates $\{\hat{h}_b\}$ have a decaying pattern, their shape variation is much larger than $\{\hat{g}_b\}$.

We present histograms of test statistics ρ_{h_0},ρ_{h_c},ρ_{h_l},ρ_{g_0},ρ_{g_c} and ρ_{g_l} to study whether they follow the normal distribution $\mathcal{N}(0, \hat{se}_B)$ assumption, see Figure 8.5, (a)-(f). It can be seen that all test statistics except ρ_{h_l} have means well away from zero and they do not follow a mean zero normal distribution.

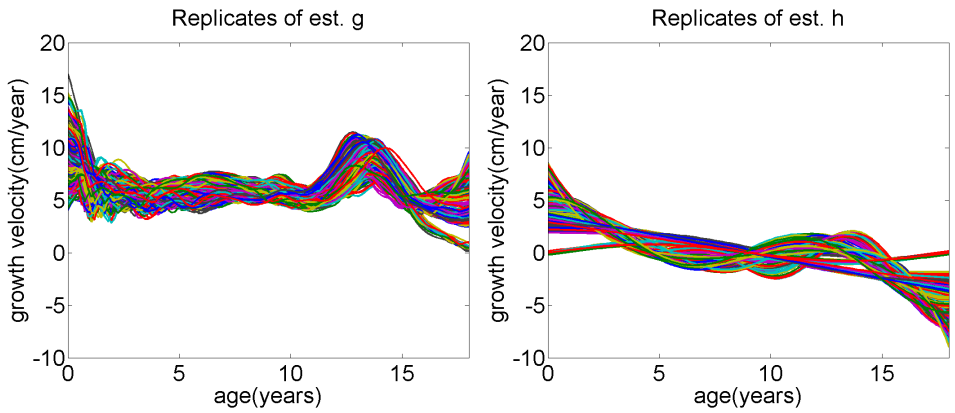
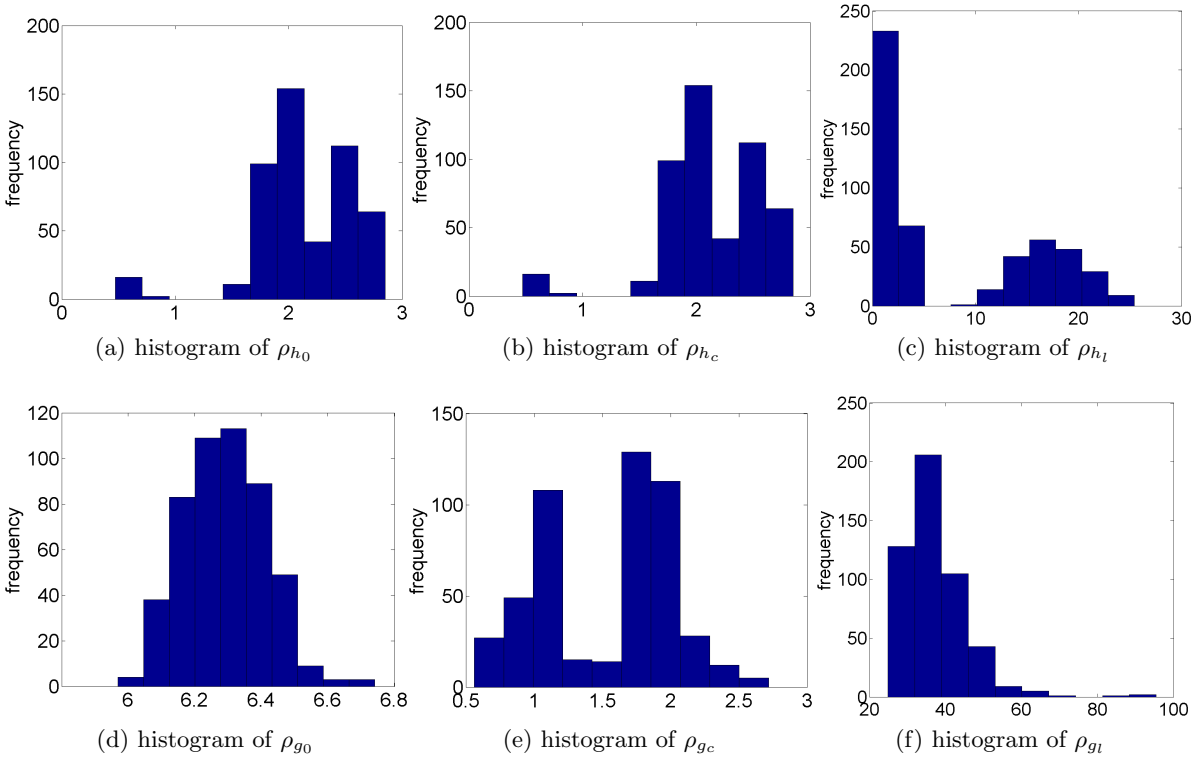


Figure 8.4: Five hundred bootstrap replicates of $\hat{g}$ (left) and $\hat{h}$ (right) for male growth velocity data.



(a) histogram of ρ_{h_0} (b) histogram of ρ_{h_c} (c) histogram of ρ_{h_l}

(d) histogram of ρ_{g_0} (e) histogram of ρ_{g_c} (f) histogram of ρ_{g_l}

Figure 8.5: Histograms of test statistics of Berkeley growth velocity data.

Given the bootstrap distribution of test statistics, we calculate the p values of these test statistics under different shape hypothesis:

- **Testing presence of trend and seasonality:** For testing $h = 0$, the test statistic using bootstrap is $\rho_{h_0} = 1.83$ with $\hat{se}_B = 0.44$ and a p value of 1.53×10^{-5}. The null hypothesis $h = 0$ is therefore rejected. For testing $g = 0$, the test statistic is $\rho_{g_0} = 6.38$ with $\hat{se}_B = 0.12$ and a p value of 0. The null hypothesis $g = 0$ is therefore rejected.
- **Testing constant shape of trend and seasonality:** For testing $h = c$, the test statistic using bootstrap is $\rho_{h_c} = 1.83$ with $\hat{se}_B = 0.44$ and a p value of 1.53×10^{-5}. Therefore, we reject the null hypothesis: $h = c$. For testing $g = c$, the test statistic using bootstrap is $\rho_{g_c} = 1.92$ with $\hat{se}_B = 0.47$ and a p value of 2.11×10^{-4}. Hence, we reject the null hypothesis: $g = c$.
- **Testing linear shape of trend and seasonality:** For the linearity of a trend, we obtain $\rho_{h_l} = 2.52$, $\hat{se}_B = 8.08$, and a p value of 0.37. Thus, we the null hypothesis: h is a linear function can not be rejected. For the seasonality g being a linear function, we have $\rho_{g_l} = 30.41$, $\hat{se}_B = 8.34$, and a p value of 1.34×10^{-4}. The null hypothesis, g is a linear function, is rejected.

Table 8.2 summarizes conclusions of the estimated trend and the seasonality under different hypothesis testings.

Table 8.2: Conclusions of bootstrap hypothesis testing of different shapes.

	Trend h	Seasonality g
H_0 : zero function	reject	reject
H_0 : constant function	reject	reject
H_0 : linear function	cannot reject	reject

Since bootstrap hypothesis testing concludes that both the seasonality $\hat{g}$ and the trend $\hat{h}$ are not constant functions, we presents their corresponding cross-sectional confidence bands with 95% confidence level in Figure 8.6.

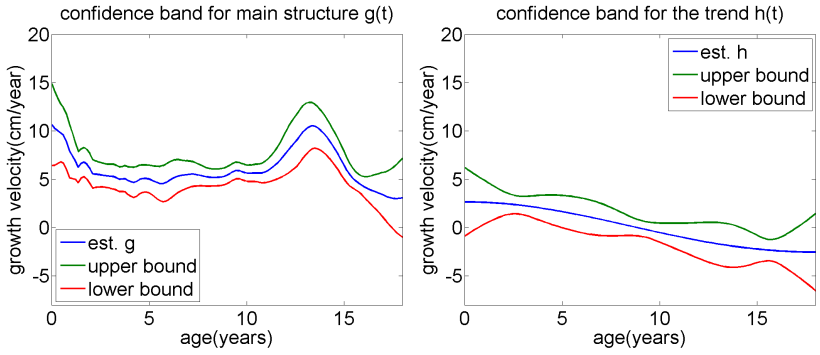


Figure 8.6: Cross-sectional confidence bands for estimated $\hat{g}$ (left) and $\hat{h}$ (right) in male growth velocity data.

8.2 U.S. Electricity Price

This example studies the monthly U.S. electricity price from year 2005 to 2010, with data shown in the Figure 8.7. The x-axis represents months from 2005.1 to 2010.12. The y-axis is unit price of electricity in cents per kilowatt-hour. According to the data source, U.S. Energy Information Administration (`http://www.eig.gov/electricity/data/`), divides U.S. into several regions: Alaska, Hawaii, New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, Pacific Contiguous and Pacific Non-contiguous. We restrict to six regions in the analysis since these regions use the same electricity generation method. The raw data is shown in Figure 8.7.

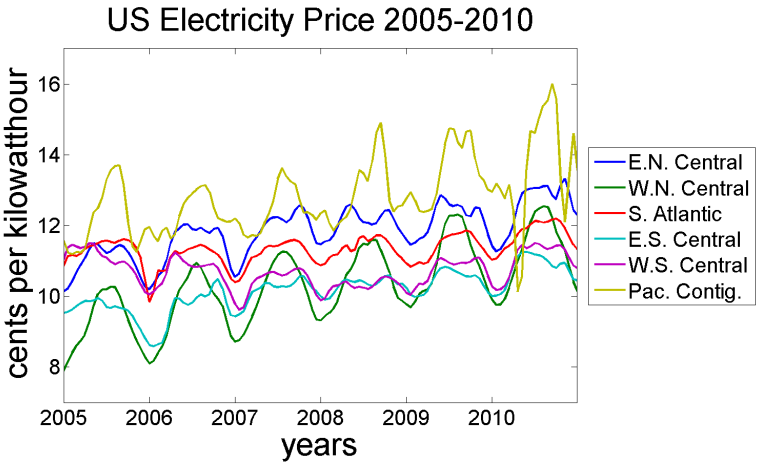


Figure 8.7: US monthly electricity price from 2005 to 2010.

In this data, we clearly see a seasonal effect on a yearly basis. Generally speaking, the electricity prices increase each summer and fall back each winter. This seasonal effect is expected to be our seasonality g. Since there are six annual cycles in the data, we expect six peaks in the estimated seasonality $\hat{g}$. Also, as the cycles are not quite synchronized, there is a potential for phase variability in the seasonal effect that will be characterized by the time warpings $\{\gamma_i\}$. Notice that there is a slowly increasing pattern from 2005 to 2010, pointing to the presence of a trend in the data.

We apply our algorithm to this data, and found the optimal choice of trend subspace is $\mathcal{H} = span\{1, \cos(\pi t), ..., \cos(5\pi t)\}$. Also Trend-Warp found the subspace $\mathcal{H} = span\{\sin(\pi t), ..., \sin(10\pi t)\}$. Estimation results under different models are given in Figure 8.8 and Figure 8.9.

In the scenario that a trend is not needed, it is surprising that the simple estimation behaviors better than alignment estimation. The reason is the warping definition in the alignment estimation model only allow changes in the phase while in data the seasonality g varies mostly in the height. On the other hand, if a trend is needed, Trend-Warp has the poorest estimated $\hat{g}$ and $\hat{h}$. Also, both simple separation and Sep.-Warp are able to produce reasonable estimation results only when the appropriate space $\mathcal{H}$ is given by our Trend-L2 model. Finally, Trend-L2 captures the six peaks in the estimated seasonality $\hat{g}$, and a slow increasing estimated trend $\hat{h}$ very well.

Table 8.3 summarizes $\mathbb{L}^2$ cost of all related models. Simple separation and Trend-L2 have the first and second lowest $\mathbb{L}^2$ cost among all models. Based on the criteria for the real data comparison, Trend-L2 is the best model since chooses the subspace $\mathcal{H}$, yields reasonable $\hat{g}$ and $\hat{h}$ that fit physical interpretation, and produces the second lowest $\mathbb{L}^2$ cost.

Table 8.3: $\mathbb{L}^2$ cost comparison under different models of US electricity price data.

Models	$\mathbb{L}^2$ cost $\frac{1}{n}\sum_{i=1}^{n}\|l.f.s. - r.h.s.\|^2$	Orthogonality $\langle g(t), h(t)\rangle_{L^2}$
Simple Separation $f_i(t) = g(t) + h(t) + \epsilon_i(t)$	5.573749e-01 **(lowest)**	2.212442e-04
Trend-L2 $f_i(t) = g(\gamma_i(t))\sqrt{\dot{\gamma}_i(t)} + h(t) + \epsilon_i(t)$	7.098805e-01	7.449066e-02
Trend-Warp $f_i(t) = (g \circ \gamma_i)(t) + h(t) + \epsilon_i$	4.833387e+02	1.187636e-04
Sep.-Warp $f_i(t) = ((g + h) \circ \gamma_i)(t) + \epsilon_i$	1.186032e+00	9.149722e-05 **(lowest)**
Simple estimation $f_i(t) = g(t) + \epsilon_i(t)$	1.089495e+00	none
Alignment estimation $f_i(t) = (g \circ \gamma_i)(t) + \epsilon_i$	1.186032e+00	none

We use bootstrap analysis to show the estimated trend $\hat{h}$ and seasonality $\hat{g}$ are both significant. Also, their cross-sectional confidence bands are provided in the following. The 500 bootstrap replicates of estimated trend h and seasonality g are shown in Figure 8.10. We see the shape pattern of the trend replicates $\{\hat{h}_b\}$ is mostly slowly increasing.

We present histograms of test statistics ρ_{h_0},ρ_{h_c},ρ_{h_l},ρ_{g_0},ρ_{g_c} and ρ_{g_l} to study whether they follow the normal distribution $\mathcal{N}(0, \hat{se}_B)$ assumption, see Figure 8.11, (a)-(f). Clearly ρ_{h_0} and ρ_{g_l} have means well away from zero and both do not follow a normal distribution $\mathcal{N}(0, \hat{se}_B)$.

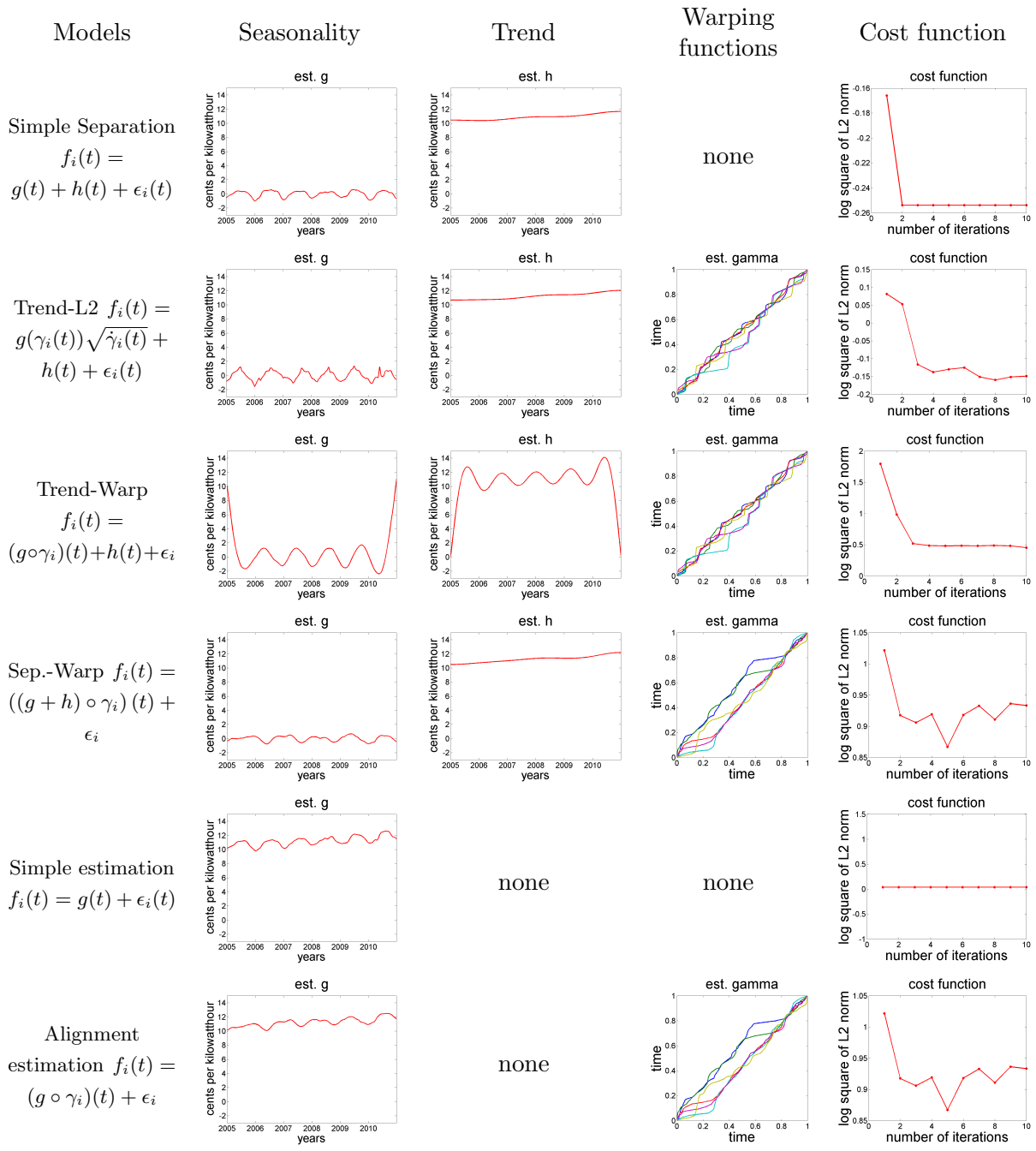


Figure 8.8: Estimation results of three components under different models of US electricity price data.

Given the bootstrap distribution of test statistics, we calculate the p values of these test statistics under different shape hypothesis:

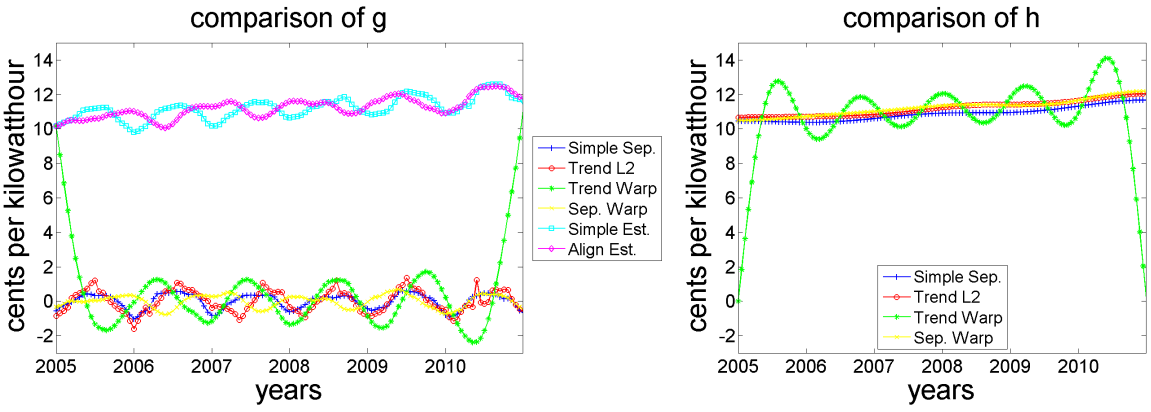


Figure 8.9: Estimation results for trend h (right) and seasonality g (left) under different models in US electricity price data.

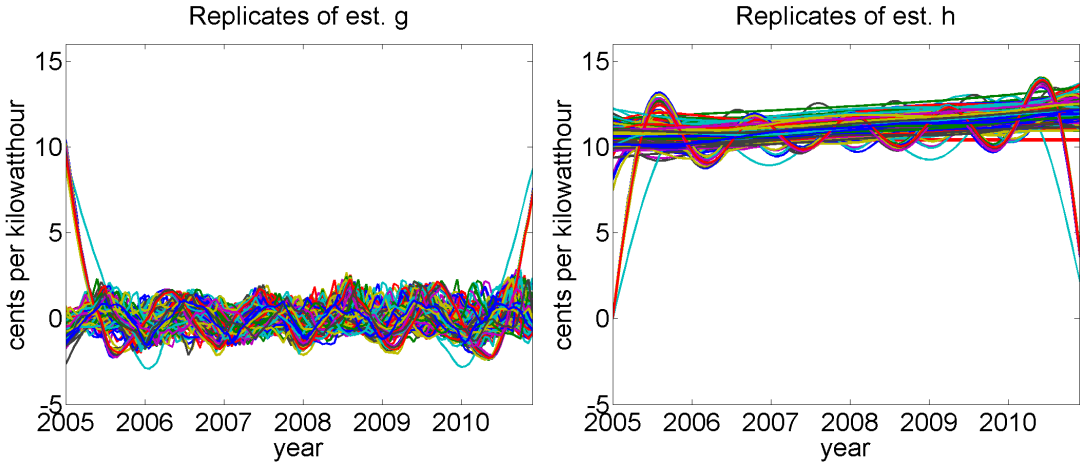


Figure 8.10: Five hundred bootstrap Replicates of $\hat{g}$ (left) and $\hat{h}$ (right) of US electricity price data.

- **Testing presence of trend and seasonality:** For testing $h = 0$, the test statistic using bootstrap is $\rho_{h_0} = 11.21$ with $\hat{se}_B = 0.398$ and a p value of 0. The null hypothesis $h = 0$ is therefore rejected. For testing $g = 0$, the test statistic is $\rho_{g_0} = 0.628$ with $\hat{se}_B = 0.419$ and a p value of 6.73×10^{-2}. The null hypothesis $g = 0$ is at the boundary of rejection.
- **Testing constant shape of trend and seasonality:** For testing $h = c$, the test statistic using bootstrap is $\rho_{h_c} = 0.431$ with $\hat{se}_B = 0.437$ and a p value of 0.162. Therefore, we cannot reject the null hypothesis: $h = c$. For testing $g = c$, the test statistic using bootstrap is $\rho_{g_c} = 0.628$ with $\hat{se}_B = 0.419$ and a p value of 6.73×10^{-2}. Hence, the null hypothesis, $g = c$, is at the boundary of rejection.
- **Testing linear shape of trend and seasonality:** For the linearity of a trend, we obtain $\rho_{h_l} = 0.173$, $\hat{se}_B = 2.864$, and a p value of 0.475. Thus, the null hypothesis: h is a linear

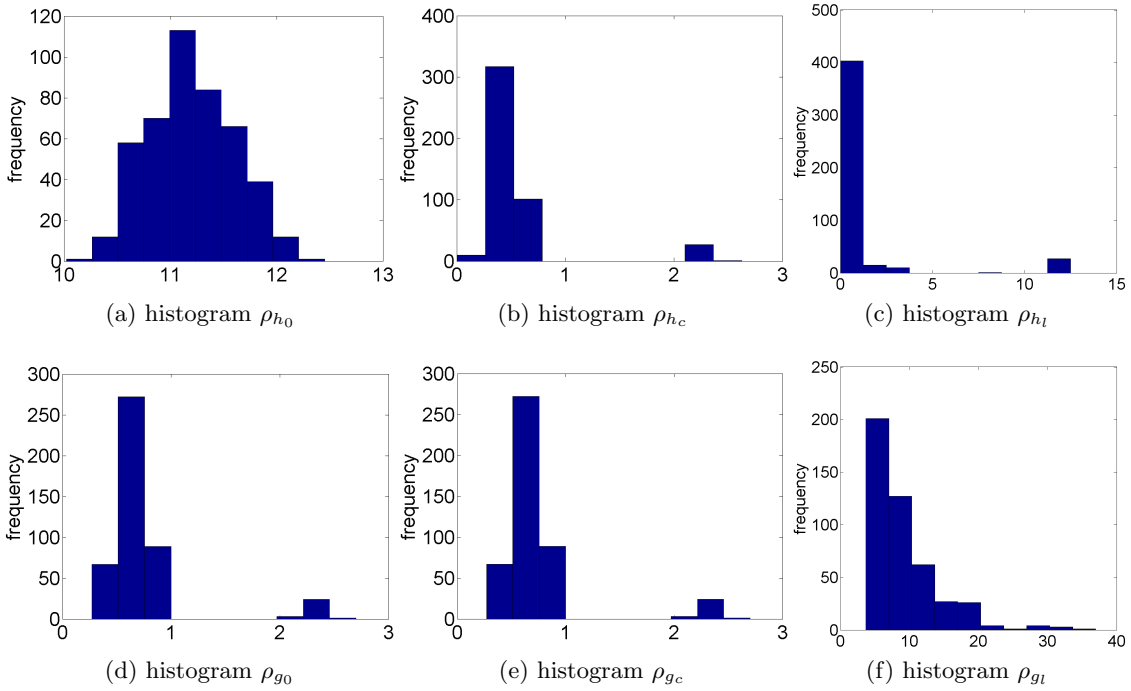


Figure 8.11: Histograms of test statistics of US electricity price data.

function can not be rejected. For the seasonality g being a linear function, we have $\rho_{g_l} = 7.64$, $\hat{se}_B = 5.093$, and a p value of 6.68×10^{-2}. The null hypothesis, g is a linear function, is at the boundary of rejection.

Table 8.4 summarizes the results for of the estimated trend and the seasonality under different hypothesis testings. Since bootstrap hypothesis testing suggests that both $\hat{g}$ and $\hat{h}$ are non-zero functions, their cross-sectional confidence bands with 95% confidence level are given in Figure 8.12.

Table 8.4: Conclusions of bootstrap hypothesis testing of different shapes of US electricity price data.

	Trend h	Seasonality g
H_0 : zero function	reject	reject
H_0 : constant function	cannot reject	reject
H_0 : linear function	cannot reject	reject

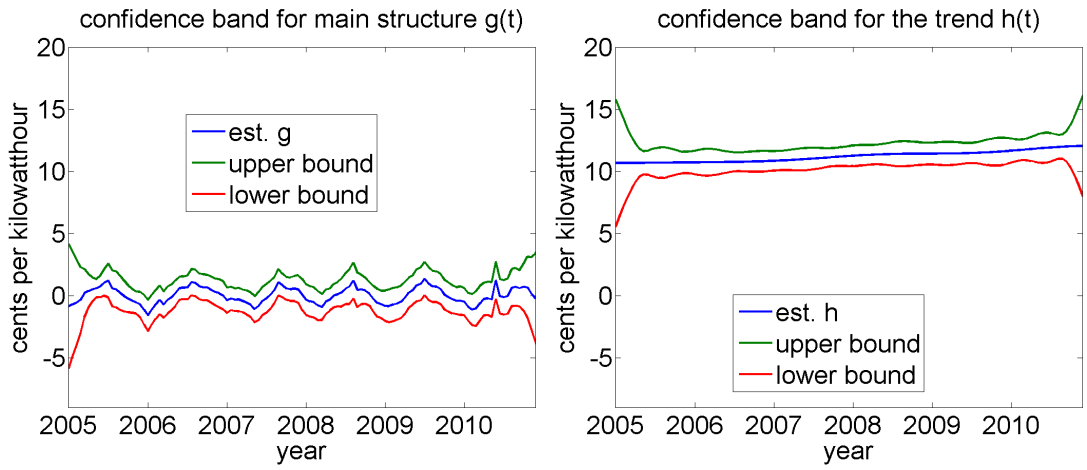


Figure 8.12: Cross-sectional confidence bands for estimates $\hat{g}$ (left) and $\hat{h}$ (right) in electricity price data.

8.3 U.S. Currency Exchange

In this experiment, we consider an application to financial data. The daily US dollar foreign exchange rates from October 2015 to December 2015 can be found from `http://www.usforex.com/` and are shown in the left graph of Figure 8.13. The y value is the worth of one U.S. dollar in other foreign currencies: Euro, British Pound, Swiss Franc, Canadian dollar, Australian dollar and Japanese Yen. Take US dollar to Euro as an example: on October 1, 2015 , the exchange rate value is 0.895, meaning 1 US dollar is worth 0.895 Euro.

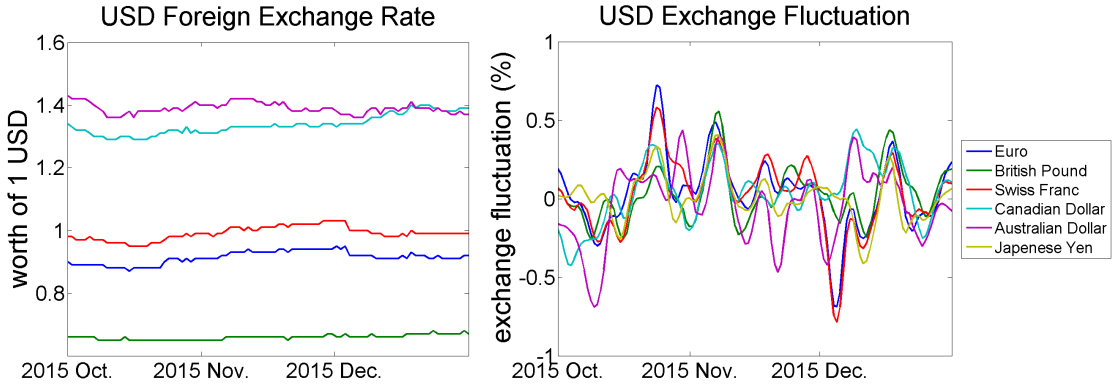


Figure 8.13: USD exchange rates (left) and fluctuation (right) from October 2015 to December 2015.

In finance, the exchange fluctuation is studied rather than using exchange rates. The exchange fluctuation is defined as

$$\tau = \frac{R_1 - R_0}{R_0} \times 100\%$$

where R_1 is the current exchange rate and R_0 is the previous exchange rate. If the number τ is positive, the US dollar is undergoing revaluation and a negative number means devaluation. The right graph in Figure 8.13 displays the USD exchange fluctuation.

The analysis of exchange fluctuation is harder than growth velocity data and electricity price data since there is no physical interpretation. Moreover, the observed currency behavior may be related to less quantifiable considerations such as economic policies and governments. Therefore, we can only look at the shapes of fluctuation data and speculate on the estimated $\hat{g}$ and $\hat{h}$.

From the right graph of Figure 8.13, there are roughly four peaks in the middle which characterize the seasonality g. We expect the trend h to follow the shape of average envelope of the data, i.e., increasing from October to November, decreasing from November to December and increasing until the end of December. The need of warping functions is suggested since the peaks and valleys in the data are not perfectly synchronized. For example, in the beginning of October, U.S. dollar reaches devaluation with the order Canadian dollar, Australian dollar, then British Pound and Swiss Franc.

Our Trend-L2 selects $\mathcal{H} = span\{1, \cos(\pi t), \cos(2\pi t), \cos(3\pi t)\}$ and Trend-Warp chooses $\mathcal{H} = span\{1, \cos(\pi t)\}$. Simple separation and Sep. Warp again use the basis choice from Trend-L2. Figure 8.15 and Figure 8.15 summarize the estimation of three components under different models.

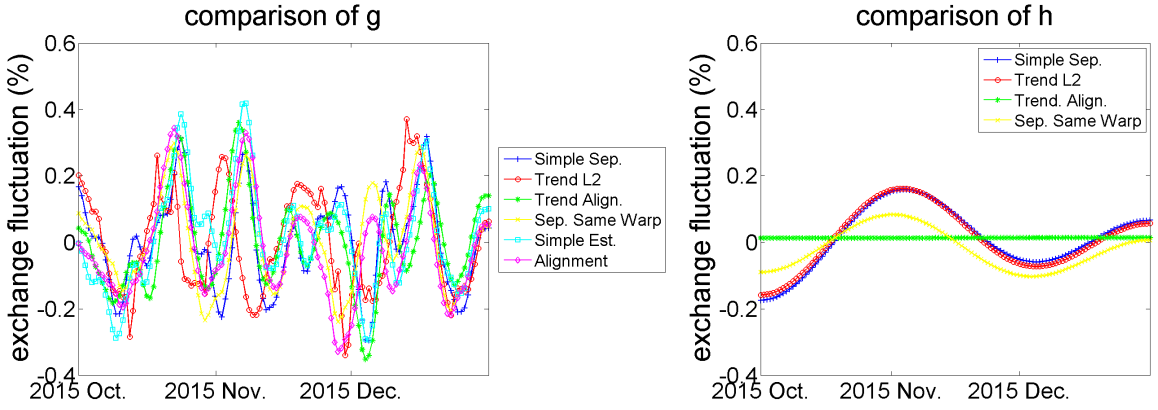


Figure 8.14: Estimation results of the seasonality g (left) and the trend h (right) under different models in USD fluctuation data.

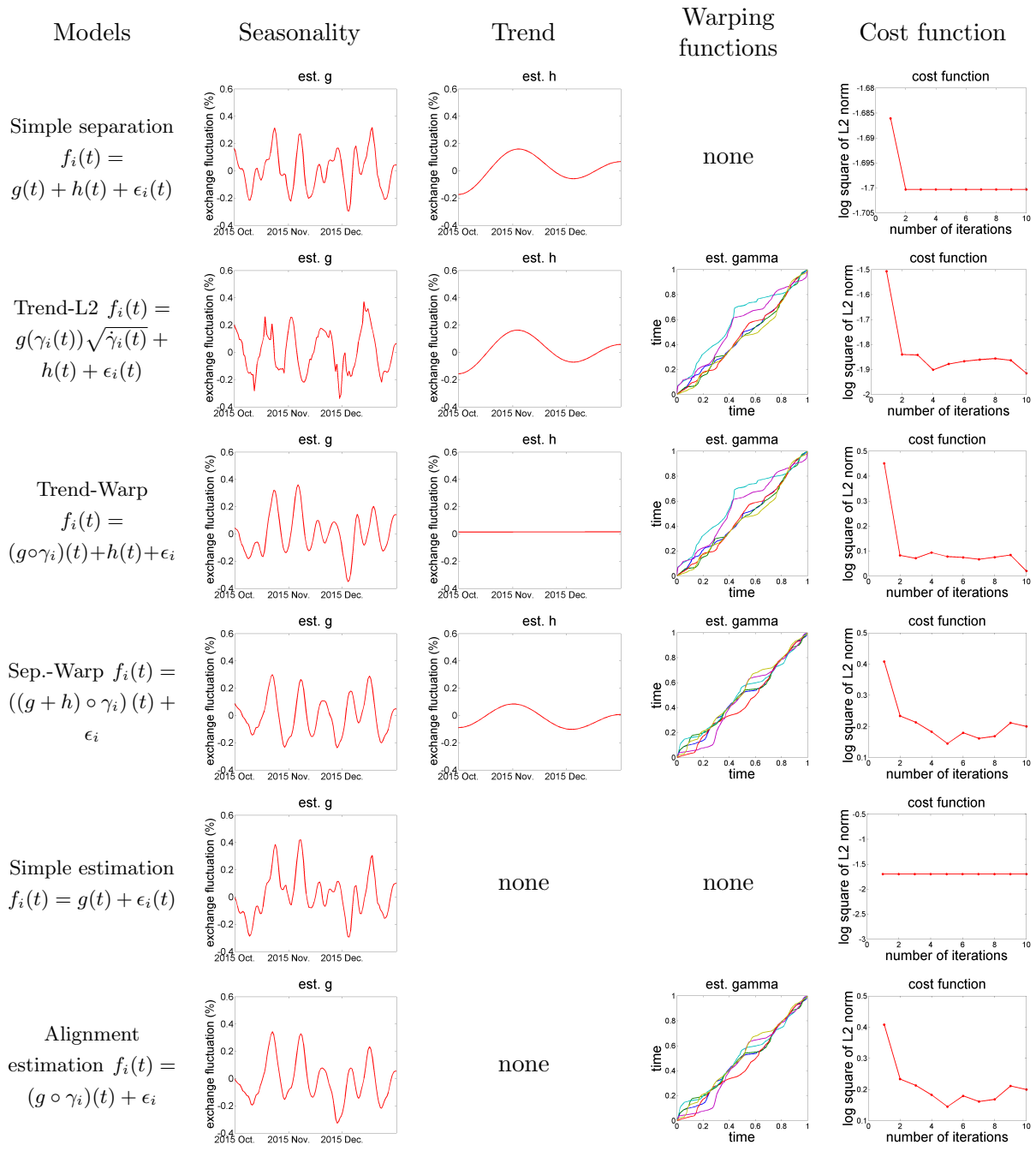


Figure 8.15: Estimation results of three components under different models in USD fluctuation data.

When a trend is not needed, both simple estimation and alignment estimation give reasonable approximations of $\hat{g}$ including the four peaks in the fluctuation data. When a trend is sought, we see

both simple separation and Sep. Warp have the desired peaks structure $\hat{g}$, and an expected trend $\hat{h}$ when using the subspace $\mathcal{H}$ selected by Trend-L2. Notice that Trend-Warp gives an undesirable constant trend which does fit expectations. Trend-L2 has similar estimated $\hat{g}$ and $\hat{h}$ to model simple separation and Sep. Warp.

Table 8.5 summarizes the $\mathbb{L}^2$ costs for all models and the Trend-L2 has the lowest $\mathbb{L}^2$ cost. By the comparison criteria for real data experiment, Trend-L2 satisfies the most criteria and is concluded to be the best model.

Table 8.5: $\mathbb{L}^2$ cost comparison of all models of USD fluctuation data.

Models	$\mathbb{L}^2$ cost $\frac{1}{n}\sum_{i=1}^{n}\|l.f.s. - r.h.s.\|^2$	Orthogonality
Simple Separation $f_i(t) = g(t) + h(t) + \epsilon_i(t)$	1.993696e-02	-2.174559e-07
Trend-L2 $f_i(t) = g(\gamma_i(t))\sqrt{\dot{\gamma}_i(t)} + h(t) + \varepsilon_i(t)$	1.214340e-02 (**lowest**)	-3.750052e-06
Trend-Warp $f_i(t) = (g \circ \gamma_i)(t) + h(t) + \epsilon_i$	2.313540e-02	2.495800e-08 (**lowest**)
Sep.-Warp $f_i(t) = ((g + h) \circ \gamma_i)(t) + \epsilon_i$	2.553436e-02	9.263048e-07
Simple estimation $f_i(t) = g(t) + \epsilon_i(t)$	1.993941e-02	none
Alignment estimation $f_i(t) = (g \circ \gamma_i)(t) + \epsilon_i$	2.553436e-02	none

We apply bootstrap analysis to confirm the estimation results are significant by using hypothesis testing. Figure 8.16 contains five hundred bootstrap replicates $\{\hat{h}_b\}$ and $\{\hat{g}_b\}$ applying Trend-L2 to US dollar fluctuation data. We see the sinusoid structure in $\{\hat{g}_b\}$ and the constant or two-peaks shape in $\{\hat{h}_b\}$.

Figure 8.17 presents the histograms of test statistics ρ_{h_0},ρ_{h_c},ρ_{h_l},ρ_{g_0},ρ_{g_c} and ρ_{g_l} It can be seen that all test statistics except ρ_{h_l} have means well away from the mean zero so the null hypotheses are likely to be rejected.

Given the bootstrap distribution of test statistics, we calculate the p values of these test statistics under different shape hypothesis:

- **Testing presence of trend and seasonality:** For testing $h = 0$, the test statistic using bootstrap is $\rho_{h_0} = 0.0925$ with $\hat{se}_B = 0.03$ and a p value of 1.079×10^{-3}. The null hypothesis

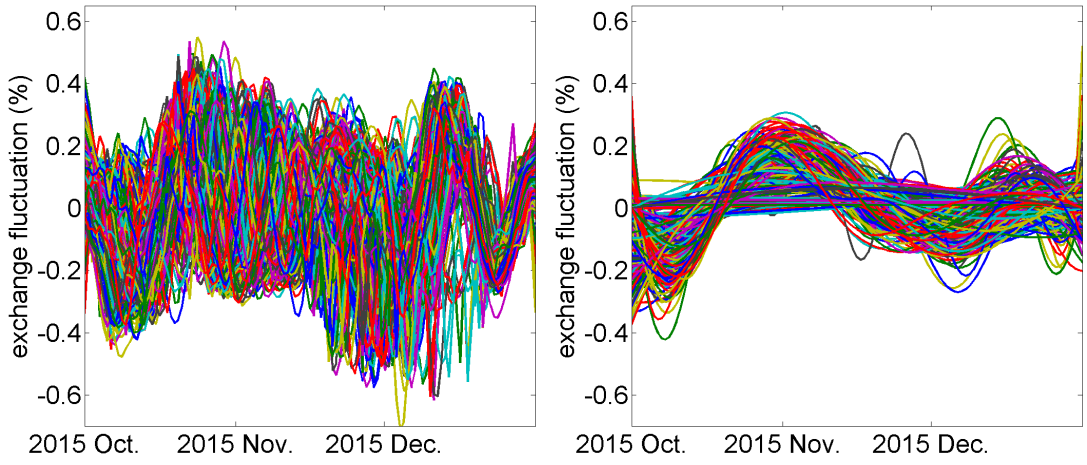
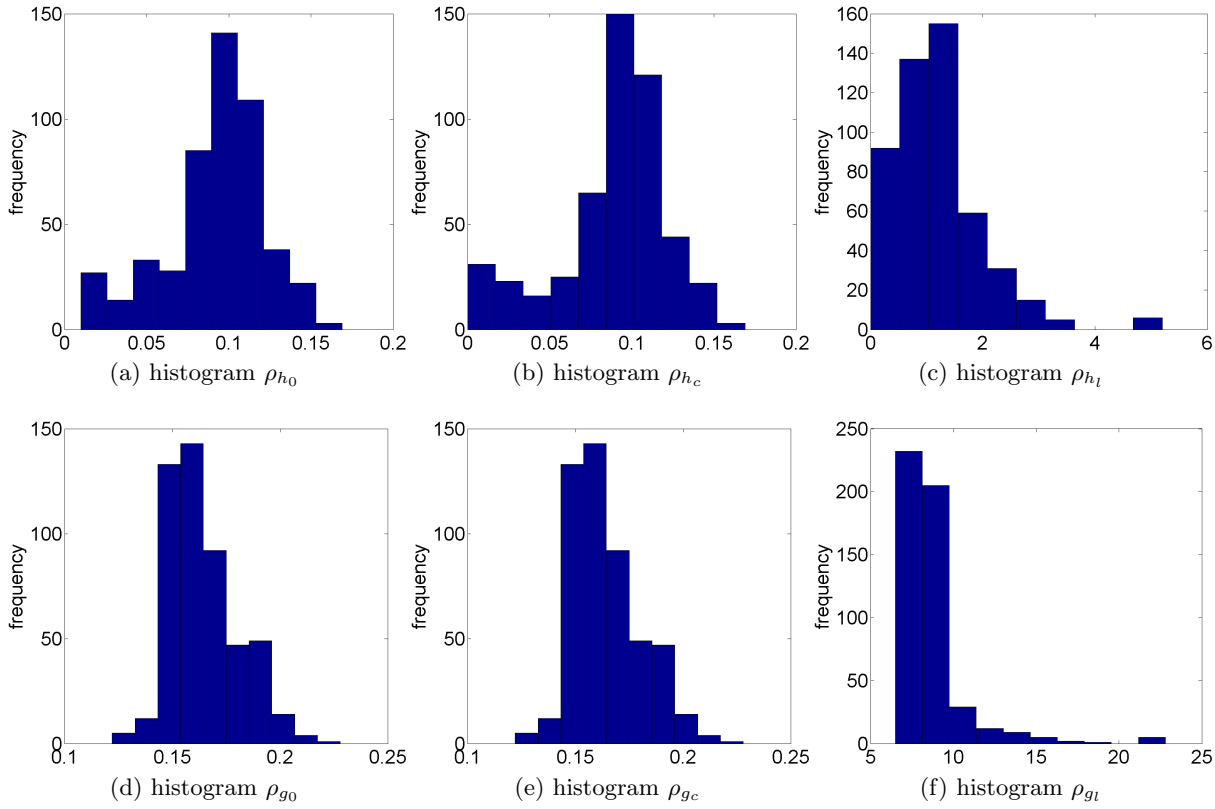


Figure 8.16: Five hundred bootstrap Replicates of $\hat{g}$ (left) and $\hat{h}$ (right) of US dollar fluctuation data.



(a) histogram ρ_{h_0} (b) histogram ρ_{h_c} (c) histogram ρ_{h_l}

(d) histogram ρ_{g_0} (e) histogram ρ_{g_c} (f) histogram ρ_{g_l}

Figure 8.17: Histograms of test statistics of US electricity price data.

$h = 0$ is therefore rejected. For testing $g = 0$, the test statistic is $\rho_{g_0} = 0.149$ with $\hat{se}_B = 0.016$ and a p value of 0. The null hypothesis $g = 0$ is therefore rejected.

- **Testing constant shape of trend and seasonality:** For testing $h = c$, the test statistic using bootstrap is $\rho_{h_c} = 0.0914$ with $\hat{se}_B = 0.0344$ and a p value of 3.97×10^{-3}. Therefore, we reject the null hypothesis: $h = c$. For testing $g = c$, the test statistic using bootstrap is $\rho_{g_c} = 0.149$ with $\hat{se}_B = 0.016$ and a p value of 0. Hence, we reject the null hypothesis: $g = c$.
- **Testing linear shape of trend and seasonality:** For the linearity of a trend, we obtain $\rho_{h_l} = 0.764$, $\hat{se}_B = 0.807$, and a p value of 0.171. Thus, the null hypothesis: h is a linear function can not be rejected. For the seasonality g being a linear function, we have $\rho_{g_l} = 7.634$, $\hat{se}_B = 2.141$, and a p value of 1.81×10^{-4}. The null hypothesis, g is a linear function, is rejected.

Table 8.2 summarizes the conclusions of the estimated trend and the seasonality under different hypothesis testings. Because the estimated $\hat{g}$ and $\hat{h}$ are both non-zero, their cross-sectional confidence bands with 95% confidence level are presented in Figure 8.18.

Table 8.6: Conclusions of bootstrap hypothesis testing of different shapes in UDS fluctuation data.

	Trend h	Seasonality g
H_0 : zero function	reject	reject
H_0 : constant function	reject	reject
H_0 : linear function	cannot reject	reject

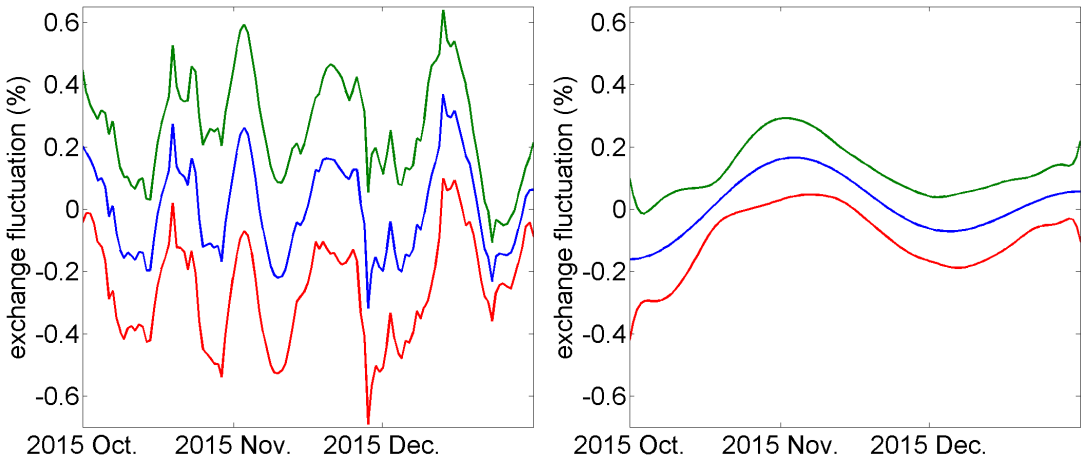


Figure 8.18: Cross-sectional confidence bands for estimates $\hat{g}$ (left) and $\hat{h}$ (right) in US dollar fluctuation.

CHAPTER 9

SIGNIFICANCE ANALYSIS OF TIME-COURSE GENE EXPRESSION DATA

9.1 Introduction

In the field of time-course genomewide studies, a key question is to detect a significant change between an experiment group and a control group under biological processes [BJGS12]. A microarray approach allows researchers to measure thousands of gene expression levels simultaneously, see [Dră03] for a review for this approach. Once the gene expression levels are obtained, statistical analysis for measuring the significance changes are applied [ST03].

Several considerations are important when e performing the significance analysis of the time-course gene expressions. Four of them are:

1. **Gene set analysis.** Although a considerable research has been done on gene-by-gene significance analysis ([SXL$^+$05, CNFT06]), it has significant limitations due to the lack of global interpretation and difficulty with compatibility between studies across phyla. On the other hand, gene set analysis divides the original genes (typically thousand or tens of thousands) into several subsets, studies changes in the gene expression for each subset, and mitigates these two limitations. More advantages of using gene set analysis can be found in [STM$^+$05], [ET07], and [Mac13]. We apply the MLE algorithm (Trend-L2) to the gene set analysis.
2. **Functional data analysis.** Statistical significance analysis on the time-course data treats each time point as contributing equally to the test statistics [CNFT06]. This may be an inappropriate assumption since it is very rare that gene experiments are measured with uniform timesteps. Recently, statisticians regarded these time-course data as longitudinal data by allowing uniform timesteps see [SXL$^+$05]. Figure 9.1 gives examples of longitudinal time-course gene data.

   Under the formulation of longitudinal data analysis, the number of time samples is assumed to be finite. An extension from finite to infinite time points with a smoothness assumption is called functional data analysis, first introduced in [RD91]. Applications of functional data analysis to the gene expression data were studied in [YMW05] and [HST15]. Examples of converting longitudinal data (Figure 9.1) into functional data are shown in Figure 9.2. In this chapter, we consider the functional data analysis approach.

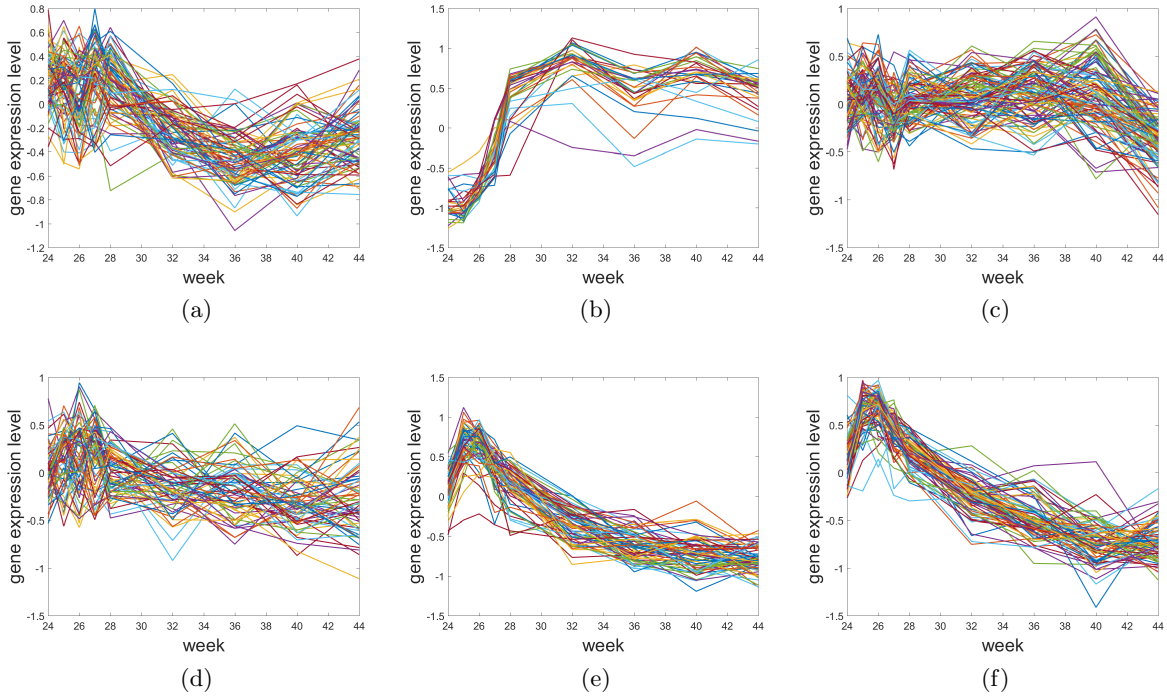


Figure 9.1: Examples of longitudinal time-course gene expression data.

3. **Time warpings.** Introducing warping functions into the statistical model improves classification and the clustering as discussed in [LM06] and [TM09]. Moreover, time warpings reflect the phenomenon that different genes have different responses to the biological processes. This dissertation is the first attempt to introduce time warping for significance analysis on time-course gene expression data

4. **Non-constant fixed effect of the gene expression.** In [SXL+05], [CNFT06] and [HST15], a constant fixed effect of the gene is assumed in the statistical model. We extend the constant effect idea to the non-constant case, i.e. a time dependent function.

9.2 Methodology

9.2.1 Application of Our MLE Algorithm

We use the trend and seasonality estimation model, from Section 2.2,

$$f_i(t) = (g, \gamma_i)(t) + h(t) + \epsilon_i(t),\ i = 1, ..., n \tag{9.1}$$

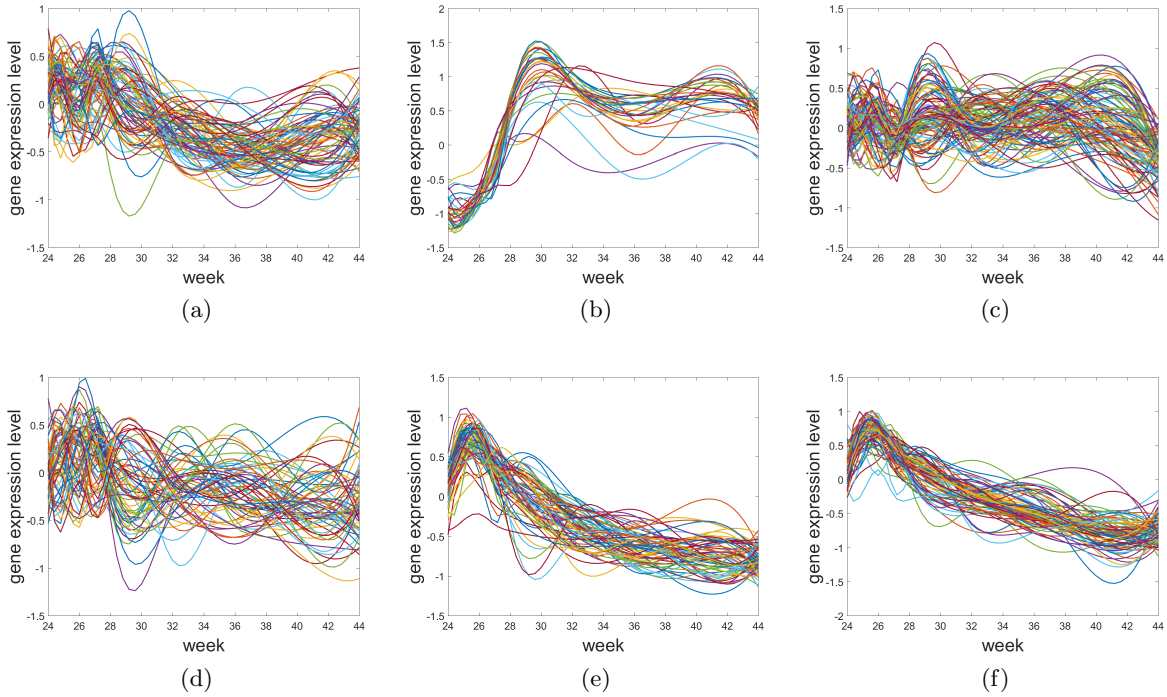


Figure 9.2: Examples of functional time-course gene expression data.

to analyze the significance of a gene set. The notation $f_i(t)$ denotes each gene's expression level at time t. The gene expression level is defined as the $\log_2$ of the fold change

$$\text{gene expression level} = \log_2(\text{fold change}) = \log_2\left(\frac{\text{experiment group}}{\text{control group}}\right).$$

Depending on the value and the sign of the gene expression, it has different biological interpretations:

$$\text{gene expression level} \begin{cases} \text{zero} & \Rightarrow \text{no changes, ex: Figure 9.1 (a),(c),(d)} \\ \text{positive} & \Rightarrow \text{upregulation, ex: Figure 9.1 (b)} \\ \text{negative} & \Rightarrow \text{downregulation, ex: Figure 9.1 (e),(f).} \end{cases}$$

In practice, biologists consider a gene expression to be **significant** if,

1. the average fold change across time is greater than 2.2 or less than $\frac{1}{2.2}$, i.e., the average gene expression level is greater than 1.137 or less than -1.137.
2. and a gene at a single time point has a fold change greater than 3 or less than $\frac{1}{3}$, i.e., a gene expression level is greater than 1.585 or less than -1.585.

Rather than using the heuristic approach, we develop a statistical model with bootstrap hypothesis testing to determine the significance of a gene.

In Equation 9.1, γ_i denotes a warping function that represents each gene having a different response to the biological process. The symbol g denotes the seasonality that is interpreted as the main gene expression level. The symbol h denotes the trend that is interpreted as the non-constant fixed effect for a gene set. To determine whether a gene set is significant, we use the hypothesis

$$\begin{cases} H_0 : \text{gene set is not significant} \\ H_1 : \text{gene set is significant} \end{cases} \iff \begin{cases} H_0 : g = 0 \text{ and } h = 0 \\ H_1 : g \neq 0 \text{ or } h \neq 0 \end{cases} \iff \begin{cases} H_0 : \rho_{g_0} = 0 \text{ and } \rho_{h_0} = 0 \\ H_1 : \rho_{g_0} \neq 0 \text{ or } \rho_{h_0} \neq 0 \end{cases}$$

The bootstrap analysis discussed in Chapter 4 is used to test the statistics ρ_{g_0} and ρ_{h_0}.

9.2.2 Method of Time-Course Gene Set Analysis (TcGSA)

We summarize the state-of-the-art method, time-course gene set analysis (TcGSA) [HST15]. Comparisons of TcGSA and our model are given later in Section 9.3. TcGSA uses a gene set approach, longitudinal data assumption, and a constant fix effect without time warpings. For each gene set, the statistical model of TcGSA is defined as

$$f_i(t) = h + g_i(t) + \varepsilon_i + \epsilon_i(t), i = 1, ...n \tag{9.2}$$

where

- time t is in the longitudinal sense that normally less than 9 points,
- h is a constant fixed effect for each gene set,
- g_i is the main structure of gene expression level, assumed to linear or cubic polynomial,
- ε_i is Gaussian noise with mean zero,
- $\epsilon_i(t)$ is also Gaussian noise with mean zero for each time t.

TcGSA uses restricted maximum likelihood to estimate $g_i(t)$. The hypothesis tests in their model is

$$\begin{cases} H_0 : \text{gene set is not significant} \\ H_1 : \text{gene set is significant} \end{cases} \iff \begin{cases} H_0 : g_i = 0 \text{ for all } i \\ H_1 : g_i \neq 0 \text{ at least one } i \end{cases}$$

Hejblum et al. suggests taking the test statistic as a likelihood ratio that follows a mixture of Chi-squares distribution, see [HST15] and [SL87]. Finally, Hejblum et al. selects a false discovery rate α and apply the Benjamini-Yekutieli procedure [BY01] to the multiple hypothesis test. If the null hypothesis is rejected the gene set is called significant.

9.3 Experimental Results

In this section we evaluate our approach to the gene significance analysis by comparing its statistical power to TcGSA and apply both methods to a real data set. In order to compare across different models, we analyze hypothesis testing results on the synthetic data. With similar significance level (FDR for multiple testings), the model with the greater statistical power is considered to be better, as suggested by [HST15]. We use time-course gene expression data post-ATI DALIA-1 that can be downloaded using the gene identifier GSE46734 or the link `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46734`.

9.3.1 Synthetic Data

Given the null hypothesis H_0 : gene set is not significant and its alternative H_1 : gene set is significant, the statistical power is defined as percentage of rejecting the null hypothesis when the alternative is true. The larger the statistical power, the better the model is considered.

We generate 100 significant gene sets and each gene set has 20 genes with 7 time points uniformly from 0 to 1. This time-course gene expression data will be re-sampled to 51 time points when fitting to our model. The gene expression level is generated by Equation 9.2, setting $h \sim \mathcal{N}(0.35, 0.087)$, $g = c_1 t + c_2 \exp(t^2) + c_3 \sin(\pi t^3)$ where $c_1 \sim \mathcal{N}(-0.2, -0.537)$, $c_2 \sim \mathcal{N}(0.3, 0.222)$, $c_3 \sim \mathcal{N}(-0.1, -0.0943)$, and $\varepsilon(t) \sim \mathcal{N}(0, 1)$ for each t. Because we simulate the gene data with 100 significant gene sets, the number of rejection H_0 becomes the statistical power. i.e., the number of significant gene sets from the conclusion of hypothesis testing becomes the statical number.

Synthetic gene data are shown in Figure 9.3 column 1, and their estimation results of three components are in Figure 9.3, columns 2 to 4. In each row (gene set), we see the main expression level $\hat{g}$ captures the shape of observed gene levels $\{f_i\}$ and the estimated effect $\hat{h}$ is not necessarily to a constant (see row 2, 4, 5, and 6 in Figure 9.3).

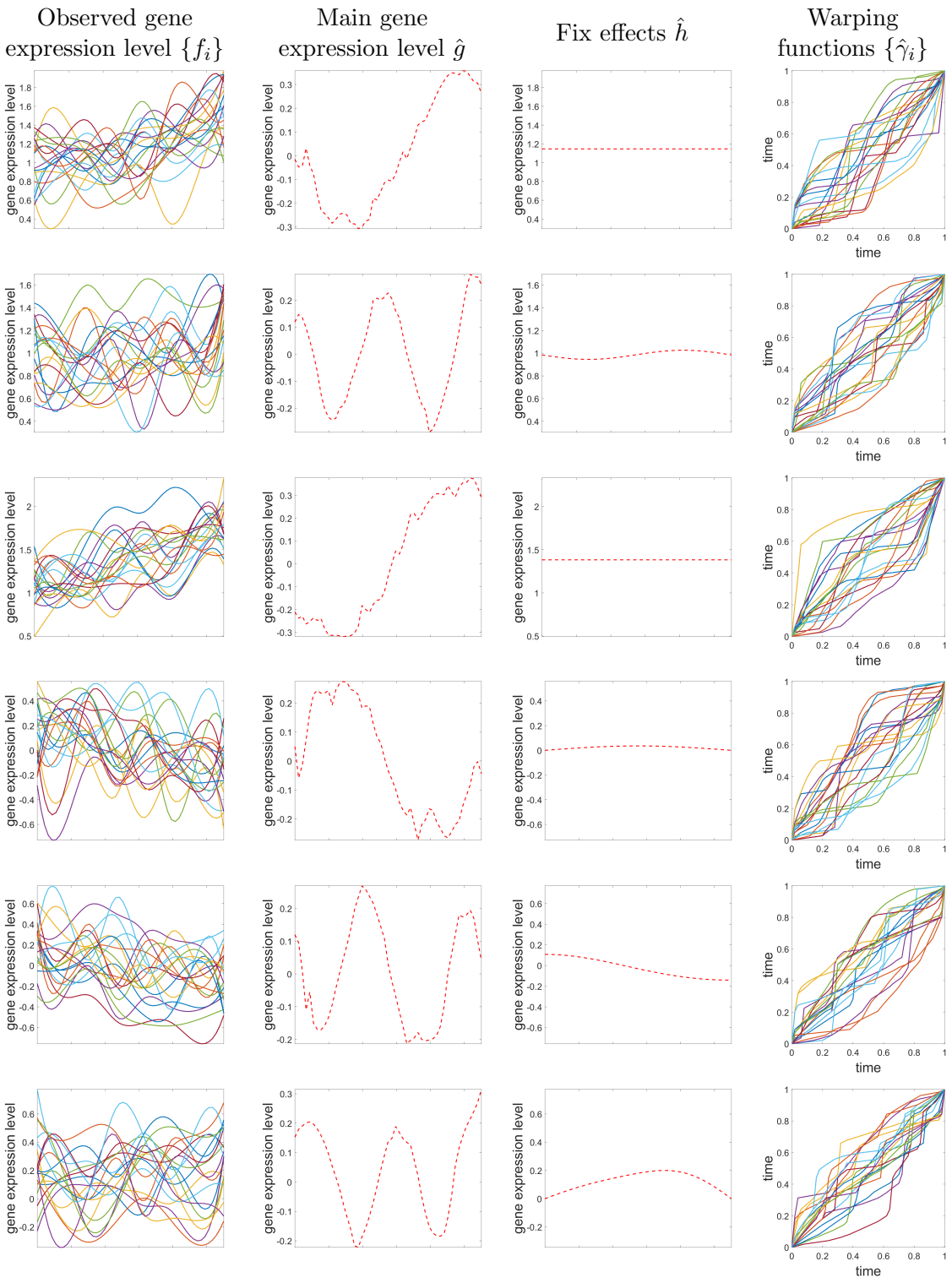


Figure 9.3: 6 out of 100 estimation results of three components in synthetic gene data.

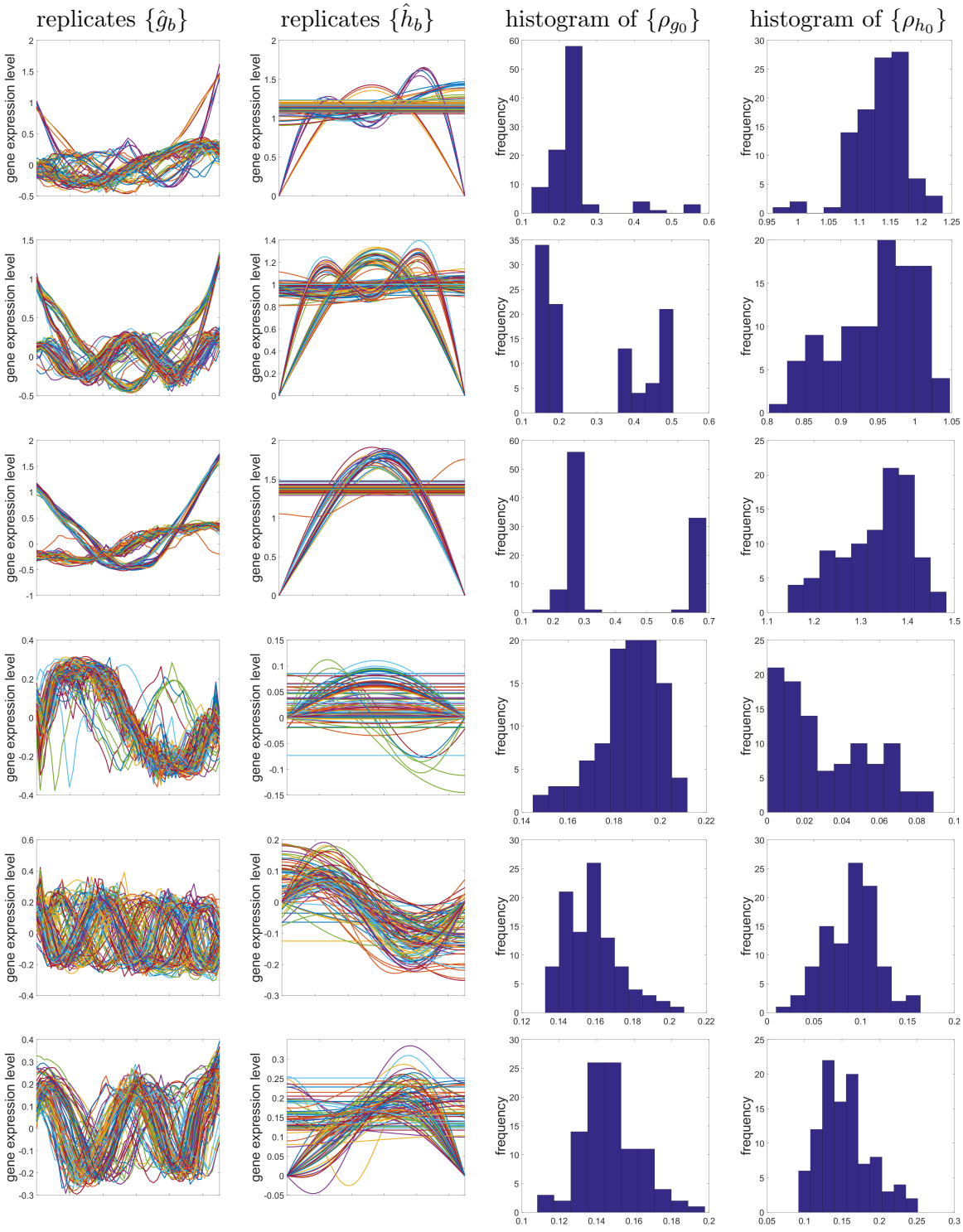


Figure 9.4: Bootstrap analysis to synthetic gene data: replicates (column 1 and 2) and histogram of test statistics (column 3 and 4). Only plot 6 out of 100 gene sets.

We set $B = 100$ for the bootstrap analysis. Bootstrap replicates $\{\hat{g}_b\}$ and $\{\hat{g}_b\}$ are shown in Figure 9.4, columns 1 and 2. Figure 9.4, columns 3 and 4, depicts the histogram of test statistics ρ_{g_0} and ρ_{h_0}. The decision to accept or reject the null hypothesis is based on the p values of these test statistics with a significance level $\alpha = 0.05$. We set the false discovery rate to be $\alpha = 0.05$ when applying TcGSA [HST15]. Table 9.1 lists the statistical power under different models when using synthetic gene data. We get a statistical power 81% in our MLE algorithm compared to 74% in the TcGSA model.

Table 9.1: Statistical power of different models in synthetic time-course gene data.

Method	Significant gene sets/Total gene sets	Statistical power
TcGSA linear [HST15]	73/100	73.00%
TcGSA cubic [HST15]	74/100	74.00%
MLE algorithm	81/100	81.00%

We continue the comparison by finding the Type I error rate for TcGSA and our MLE algorithm. The Type I error rate is the percentage of rejecting H_0 when H_0 is true. The smaller the Type I error rate, the better the method.

We generate 100 gene sets and each of them are purely noise with $\varepsilon_i \sim \mathcal{N}(0, 0.03)$ and $\epsilon_i(t) \sim \mathcal{N}(0, 0.05)$ for each t. In our setting, the number of claimed significant gene sets becomes type I error rate. We use the same bootstrap setting as the statistical power experiment. Table 9.2 summarizes the Type I error rate among different models. Although two models have similar Type I error rates, our MLE algorithm has a slightly smaller value. Since our MLE algorithm has a larger statistical power and a very slightly smaller Type I error rate than TcGSA, we conclude that our MLE algorithm has more statistical power.

Table 9.2: Type I error rate of different models in synthetic time-course gene data.

Method	Significant gene sets/Total gene sets	Type I error rate
TcGSA linear [HST15]	10/100	10.00%
TcGSA cubic [HST15]	12/100	12.00%
MLE algorithm	9/100	9.00%

9.3.2 Real Data

In the post-ATI DALIA-1 data set, there are 32,978 genes with their gene expression levels at 9 time points (week 24, 25, 26, 27, 28, 32, 36, 40, 44). We re-sampled the gene-expression data to 51 uniform time points using a cubic spline to give a functional data form. According to [CQS+08] and `http://www.biir.net/public_wikis/module_annotation/V2_Trial_8_Modules`, these genes are clustered into 260 gene sets. Figure 9.2, (a)-(f), are the first 6 gene sets of the data.

For each gene set, we apply the MLE algorithm (Algorithm 3) with subspace selection (Chapter 5). As an illustration, 6 out of 260 estimation results are plotted in Figure 9.5. We see all the seasonality estimates capture the main structure from the gene expression levels $\{f_i\}$ and both constant and non-constant estimates of fix effects are observed, see row 3 in Figure 9.5.

We analyze the bootstrap replicates ($B = 50$) of the main expression level $\hat{g}$ and the fixed effects $\hat{h}$ for each gene set, see Figure 9.6, column 1 and 2. Within each gene set, the variation of replicates $\{\hat{g}_b\}$ is smaller than the variation of replicates $\{\hat{h_b}\}$. The histogram of test statistics ρ_{g_0} and ρ_{h_0} are given in Figure 9.6, column 3 and 4. Given the empirical bootstrap distribution of the test statistics, we compute the p values of the null hypothesis: $g = 0$ and $h = 0$. For each gene set, we set the significance level $\alpha = 0.05$ and false discovery rate $\alpha = 0.05$ for TcGSA. The same significance analysis is repeated for all 260 gene sets. Table 9.3 summarizes the significance rate of the Post-ATI DALIA-1 data.

Table 9.3: Significance analysis of different models in Post-ATI DALIA-1 data. Use false discovery rate $\alpha = 0.05$ in the multiple hypothesis testing.

Method	Significant/Total	Significant gene set rate
TcGSA linear [HST15]	203/260	78.07%
TcGSA cubic [HST15]	216/260	83.07%
MLE algorithm	185/260	71.15 %

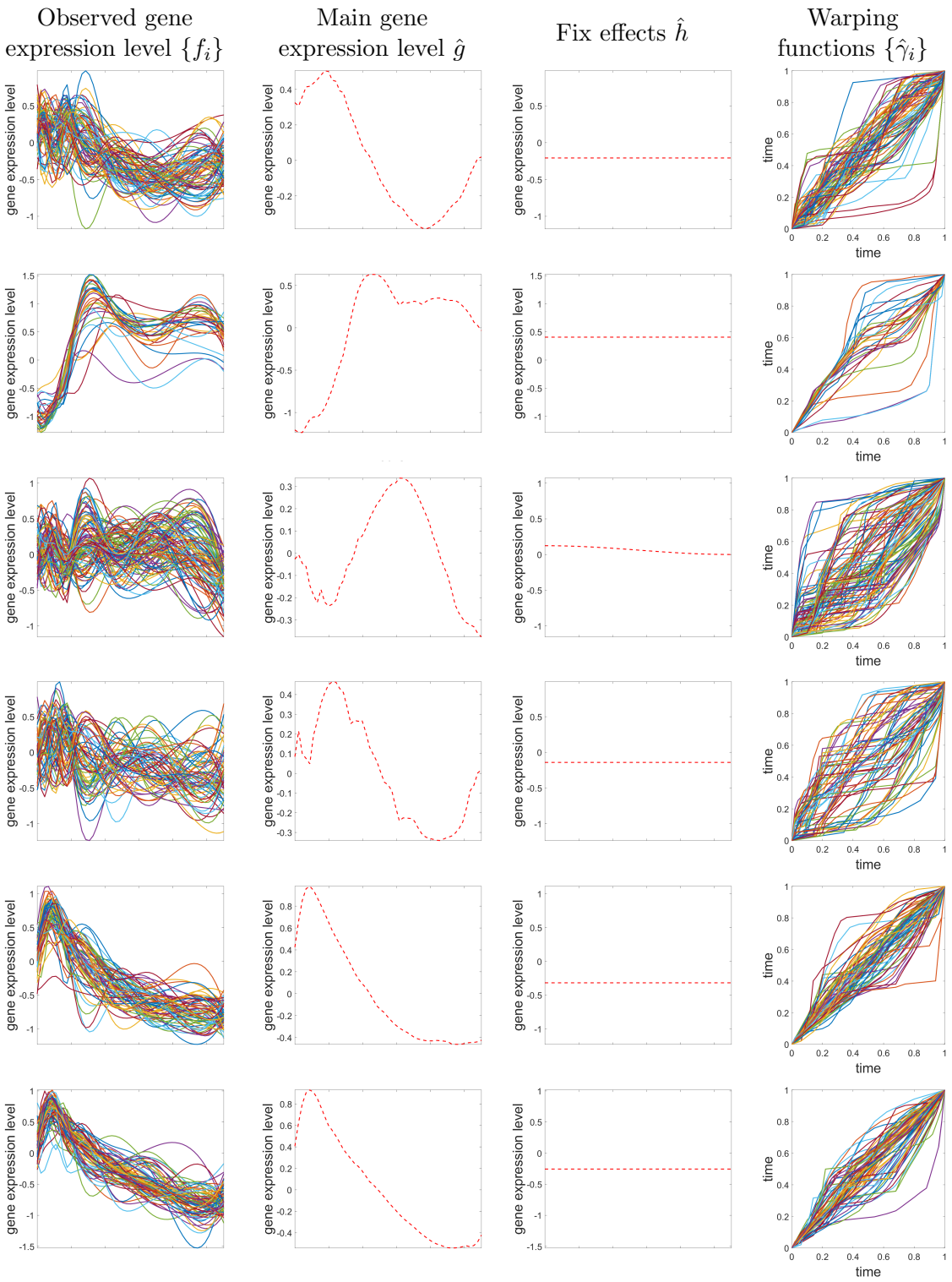


Figure 9.5: 6 out of 260 estimation results of three components in Post-ATI DALIA-1 data.

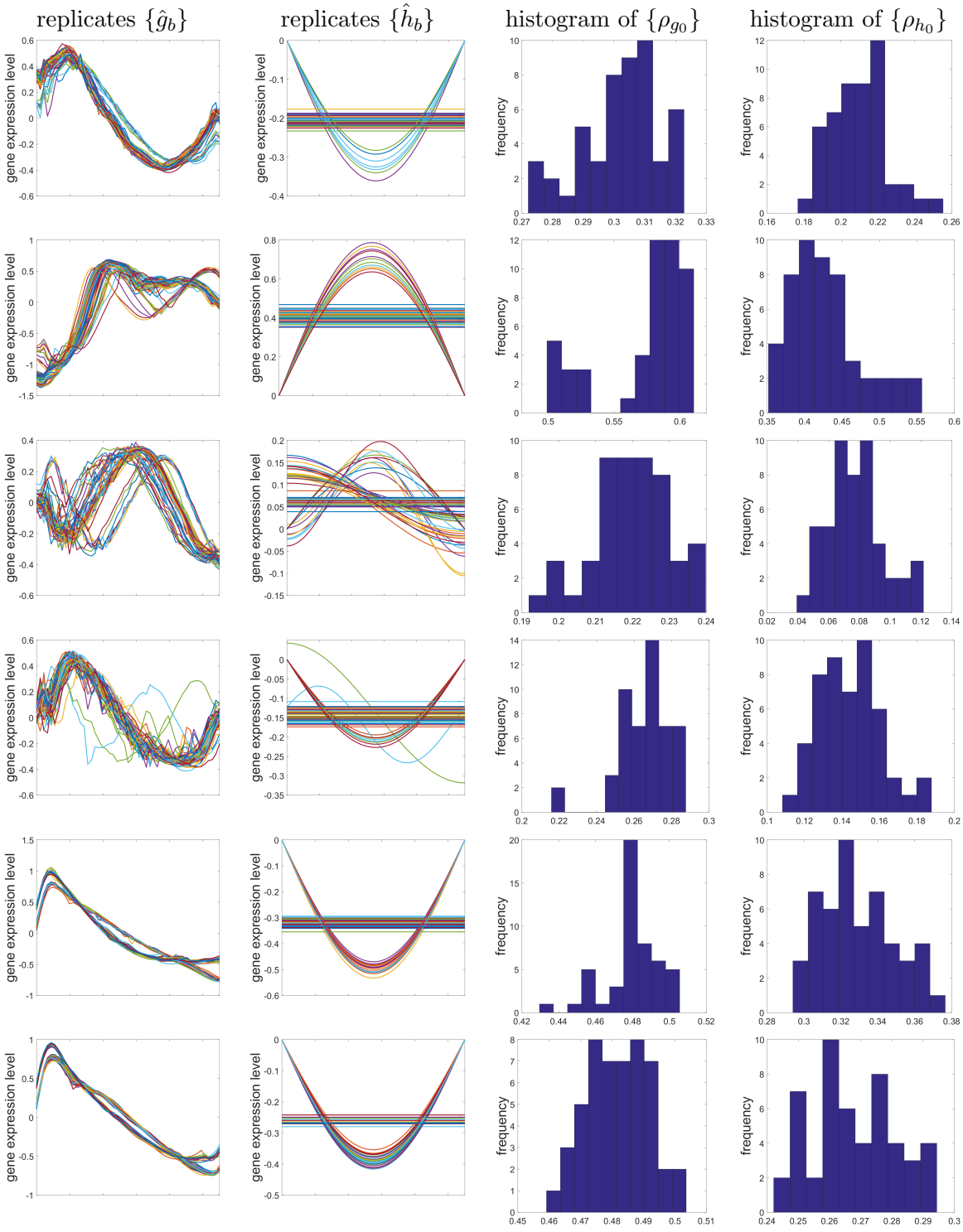


Figure 9.6: Bootstrap analysis to Post-ATI DALIA-1 data: replicates (column 1 and 2) and histogram of test statistics (column 3 and 4). Only plot 6 out of 260 gene sets.

From Table 9.3, although the significant gene set rates of TcGSA are higher than those from our proposed model, it is incorrect to conclude TcGSA is better than our MLE algorithm for two reasons:

1. The resulting significant gene set rate depends on the significance level and the false discovery rate α. The more stringent (smaller) α, the lower significant gene set rate. The choice of α is somewhat ad hoc and highly dependent on the experience of biologists [STM$^+$05].
2. The biological ground truth about the gene data is unknown. The purpose of significance analysis is to reduce the size of gene data from tens of thousands to thousands. This reduced gene data is further studied by designing sequence of experiments to understand their biological functions.

Rather than concluding the Post-ATI DALIA-1 data has significance gene set rate 78%, we would conclude 71% because our MLE algorithm is more accurate than TcGSA in the significance analysis, as shown in Section 9.3.1.

CHAPTER 10

CONCLUSIONS AND FUTURE WORK

10.1 Summary of Completed Work

We have developed a novel, model-based framework to solve the trend and variable phase seasonality estimation problem by estimating the trend h and seasonality g components from time series data in situations where the seasonal component exhibits random time warpings. The model subsumes those used by related approaches in the literature. We assume that the subspaces associated with these two components – trend and seasonality – are orthogonal, and the Karcher mean of warping functions is the identity, to ensure that the components are identifiable. Under these conditions we seek the MLE of h and g, using a coordinate descent algorithm that iteratively updates one component at a time, while fixing the others. We also use maximized likelihood to select an appropriate subspace $\mathcal{H}$ for the trend component. The estimated quantities – trend and seasonality – are tested, using bootstrap replication, for being null or having a specific simple shape, such as constant and linear.

Both synthetic data and real data have been used to demonstrate the effectiveness of this method. Using synthetic data, where the ground truth is known, we have demonstrated the robustness of our method in the presence of noise. We demonstrate this framework's ability to extract trend and seasonality using three real application datasets: the Berkeley growth data, U.S. electricity price data, and USD exchange fluctuation. Bootstrap hypothesis testing supported the presence of a trend in all of the datasets. For the USD exchange fluctuation data, a low frequency oscillation between revaluation and devaluation was extracted, along with a higher frequency seasonal oscillation. For the Berkeley growth data and electricity price data, we obtain estimates of the trends and seasonal effects that are expected from the nature of the applications. Finally, when studying the significance analysis of time-course gene expression data, our method is shown to have higher statistical power than the state-of-the-art TcGSA method.

10.2 Future Research

This dissertation motivates several directions for future work. These include:

1. Application
   (a) More real dataset experiments can be investigated. For instance, ten years monthly temperature data among several cities or ten years monthly revenue among several companies in the same industry are publicly available.
   (b) Alternate families of basis functions, e.g., Wavelets [D$^+$92], that can capture specific local characteristics of a function can be evaluated.
2. Algorithm
   (a) Real datasets may have additional properties that we have not exploited. For example, the growth velocity data has the property $f_i(t) \geq 0$. Therefore, it is reasonable to assume that both the seasonality (main structure) $g(t)$ and the underlying trend $h(t)$ are greater than or equal to zero because the growth rate should not be a negative number, i.e., we have the additional constraints of $g(t) \geq 0$ and $h(t) \geq 0$. Modifications to the algorithm or new algorithms that incorporate such additional constraints from the physical problem can be investigated.
   (b) Improvements to the algorithms can be investigated to increase robustness and efficiency. Modifications can be made to the current version of coordinate descent to avoid stagnation at a non-stationary point. Alternative optimization approaches can be considered to improve the convergence behavior and efficiency.
3. Bootstrap Analysis and its Inference
   (a) In the bootstrap analysis testing the presence of a trend or seasonality, we define the test statistic to be $\rho_{h_0} = \|\hat{h}\|$ which follows a normal distribution. However, there are other alternatives such as $\|\hat{h}^p\|$ where p is a positive integer. Additionally, the normal distribution assumption can be replaced by a percentile distribution or other distributions. To determine which is better, synthetic data experiments with statistical power comparison are necessary for supporting the bootstrap analysis.
   (b) Once the best test statistic is determined, the empirical size and the coverage rate of the bootstrap confidence band can be studied. Specifically, as the bootstrap size increases, we expect the bandwidth of the confidence band to reduce until an optimal bootstrap size is reached.
   (c) Explore the possibility of replacing the entire bootstrap inference process by a single statistic. In other words, once the estimated trend $\hat{h}$ and seasonality $\hat{g}$ are obtained, is it possible to have a conclusion about their presence rather than utilizing bootstrap analysis? If so, the inference process will speed up significantly.

4. Theory

   (a) As mentioned above, the modifications to the methods, e.g., changing coordinate descent to avoid remaining at a non-stationary point, can be considered. Further research can be made on the theoretical convergence behavior of the optimization problem generated from coordinate descent (or other methods) in a noise-free situation.

   (b) Theory about asymptotic normality of trend and seasonality estimators toward their approximated cumulative density function can be investigated, see [HK12] and [Bos12].

   (c) A fundamental open question is to prove a uniqueness theorem in a noise-free case assuming the existence of the three components g, h and $\{\gamma_i\}$. This may require nonlinear operator theory [Con99, BP67].

APPENDIX A

KARCHER MEAN OF WARPING FUNCTIONS

The Karcher mean was introduced by Karcher in 1977 [Kar77]. Let $x_1, ..., x_n$ be elements of a metric space M with metric d, then the Karcher mean x_{KM} of $x_1, ..., x_n$ is defined as

$$x_{KM} = \arg\min_{p \in M} \frac{1}{n} \sum_{i=1}^{n} w_i d^2(p, x_i)$$

where w_is are the weight for each $d^2(p, x_i)$. The Karcher mean is an intrinsic. The existence and uniqueness for some special cases of Karcher means were shown by Kendall [Ken90]. We adapt the Karcher mean to $\gamma_1, ..., \gamma_n$ on the manifold Γ. Hence,

$$\gamma_{KM} = \arg\min_{\gamma \in \Gamma} \frac{1}{n} \sum_{i=1}^{n} d^2(\gamma, \gamma_i) \tag{A.1}$$

where $d(\gamma, \gamma_i)$ denotes the geodesic distance from γ to γ_i.

Invariance of geodesic distance under warping is required. In other words,

$$d(\gamma_1, \gamma_2) = d(\gamma_1 \circ \gamma, \gamma_2 \circ \gamma).$$

However, the above equality does not hold for an arbitrary metric. In 1982, Cencov [Cen00] showed the Fisher-Rao Riemannian metric (FR-RM) is the only metric with this warping invariance. The geodesic distance under the Fisher-Rao Riemannian metric is denoted by $d_{FR}(\cdot, \cdot)$ and therefore the desired Karcher mean becomes

$$\gamma_{KM} = \arg\min_{\gamma \in \Gamma} \frac{1}{n} \sum_{i=1}^{n} d_{FR}^2(\gamma, \gamma_i). \tag{A.2}$$

To solve Equation A.2, the geodesic distance $d_{FR}(\gamma, \gamma_i)$ is defined by the infimum of length L of all the possible paths, α, that connect γ to γ_i.

$$d_{FR}(\gamma, \gamma_i) = \inf_{\{\alpha : [0,1] \to M | \alpha(0) = \gamma, \alpha(1) = \gamma_i\}} L[\alpha].$$

Let M be a Riemannian manifold. Let $\frac{d\alpha}{dt}$ be the tangent vector in the tangent space $T_{\alpha(t)}M$. The length of the path $\alpha : \mathbb{R} \to M$ is defined by the integral of square root of the metric Φ_α of $(\frac{d\alpha}{dt}, \frac{d\alpha}{dt})$,

that is

$$L[\alpha] = \int_0^1 \sqrt{\Phi_\alpha \left(\frac{d\alpha}{dt}, \frac{d\alpha}{dt} \right)} dt.$$

The choice of metric Φ is crucial. Here a non-parametric, extended Fisher-Rao Riemannian metric Φ_α is used because we want the warping invariance

$$\Phi_\alpha \left(v_1, v_2 \right) = \frac{1}{4} \int_0^1 \dot{v}_1(t) \dot{v}_2(t) \frac{1}{|\dot{\alpha}(t)|} dt$$

where v denotes the tangent vector of path α at time t. The geodesic distance d_{FR} using the Fisher-Rao Riemannian metric is

$$d_{FR}(\gamma, \gamma_i) = \inf_{\{\alpha:[0,1] \to M | \alpha(0)=\gamma, \alpha(1)=\gamma_i\}} \int_0^1 \sqrt{\Phi_\alpha \left(\frac{d\alpha}{dt}, \frac{d\alpha}{dt} \right)} dt. \tag{A.3}$$

The difficulty of finding such $d_{FR}(\gamma, \gamma_i)$ in Equation A.3 is that the metric Φ_α changes at every time point of path α due to computing new tangent vectors $\frac{d\alpha}{dt}$. The numerical implementation is very difficult. Fortunately, Bhattacharya [Bha43] provided a square-root representation that can greatly simplify the Fisher-Rao Riemannian metric calculation. Joshi et al. [JKSJ07] and Srivastava et al. [SWK+11] adapted the square-root representation and simplified the FR-RM into an $\mathbb{L}^2$-metric.

Define a new square-root transformation Q, such that $Q(\gamma(t)) = \sqrt{\dot{\gamma}(t)} = \psi(t)$. Notice that the following holds.

$$\|\psi\|_{\mathbb{L}^2}^2 = \left\| \sqrt{\dot{\gamma}} \right\|_{\mathbb{L}^2}^2 = \int_0^1 \left(\sqrt{\dot{\gamma}} \right)^2 dt = \int_0^1 \dot{\gamma} \, dt = \gamma(1) - \gamma(0) = 1 - 0 = 1.$$

This means all ψ are points on a Hilbert unit sphere $\mathbb{S}^\infty$. The geodesic distance between ψ_1 and ψ_2 under the $\mathbb{L}^2$-metric is the angle θ of the great circle connecting ψ_1 and ψ_2

$$d(\psi_1, \psi_2) = \theta = \cos^{-1} \left(\langle \psi_1, \psi_2 \rangle_{\mathbb{L}^2} \right) = \cos^{-1} \left(\int_0^1 \psi_1 \psi_2 dt \right)$$

where the property $\langle \psi_1, \psi_2 \rangle = \|\psi_1\| \, \|\psi_2\| \cos\theta = \cos\theta$ is used.

Because of the simpler geometry, it is better to work with the Karcher mean of $\psi_1, ..., \psi_n$ on the unit sphere $\mathbb{S}^\infty$, rather than Karcher mean of $\gamma_1, ..., \gamma_n$ in the space of Γ. Hence,

$$Q\left(\gamma_{KM}\right) = \psi_{KM}$$
$$\Rightarrow Q \left(\arg\min_{\gamma \in \Gamma} \frac{1}{n} \sum_{i=1}^n d_{FR}^2(\gamma, \gamma_i) \right) = \arg\min_{\psi \in \mathbb{S}^\infty} \frac{1}{n} \sum_{i=1}^n d^2(\psi, \psi_i)$$

Once ψ_{KM} is founded, it is converted to γ_{KM} using the transformation

$$\gamma_{KM}(t) = Q^{-1}(\psi_{KM}(t)) = \int_0^t \psi_{KM}^2(s)\, ds.$$

The Karcher mean of $\psi_1, ..., \psi_n$ requires solving the following manifold minimization [SWK$^+$11, SWK$^+$11, AMS08]

$$\psi_{KM} = \arg\min_{\psi \in \mathbb{S}^\infty} \frac{1}{n} \sum_{i=1}^{n} d^2(\psi, \psi_i).$$

Because of the unit sphere geometry, the Riemannian gradient $\mathbb{S}^\infty \to T(\mathbb{S}^\infty)$ of the cost functional $\min_{\psi \in \mathbb{S}^\infty} \frac{1}{n} \sum_{i=1}^{n} d^2(\psi, \psi_i)$ is given by [Kar77]

$$-\frac{2}{n} \sum_{i=1}^{n} \exp_\psi^{-1}(\psi_i)$$

where $\exp_\psi^{-1} \psi_i$ is the inverse exponential map $\mathbb{S}^\infty \to T(\mathbb{S}^\infty)$. This inverse exponential map is

$$v_i = \exp_\psi^{-1}(\psi_i) = \frac{\theta}{\sin\theta} \left(\psi_i - \cos(\theta)\psi\right)$$

where $\theta = \cos^{-1}\left(\langle \psi, \psi_i \rangle_{\mathbb{L}^2}\right)$. The retraction $R_\psi : T(\mathbb{S}^\infty) \to \mathbb{S}^\infty$, is taken to be the exponential map $\exp_\psi : T(\mathbb{S}^\infty) \to \mathbb{S}^\infty$ which is

$$R_\psi(v_i) = \exp_\psi(v_i) = \cos\left(\|v_i\|\right) \psi + \sin\left(\|v_i\|\right) \frac{v_i}{\|v_i\|}.$$

Taking the stepsize to be $\frac{\varepsilon}{2}$, one step of the Riemannian Steepest Descent for the Karcher mean of $\psi_1, ..., \psi_n$ is

$$\begin{aligned}
\psi^{(k+1)} &= \text{Retraction} \circ (\text{Stepsize} \times \text{Direction}) \\
&= \text{Retraction} \circ (\text{Stepsize} \times -\text{Riemannian Gradient}) \\
&= \text{Retraction} \circ \left(\frac{\varepsilon}{2}(-1) \left[-\frac{2}{n} \sum_{i=1}^{n} \exp_{\psi^{(k)}}^{-1}(\psi_i) \right] \right) \\
&= \exp_{\psi^{(k)}} \left(\frac{\varepsilon}{n} \sum_{i=1}^{n} \exp_{\psi^{(k)}}^{-1}(\psi_i) \right).
\end{aligned}$$

BIBLIOGRAPHY

[AB06] Charalambos D Aliprantis and Kim Border, *Infinite dimensional analysis: a hitchhiker's guide*, Springer Science & Business Media, 2006.

[ABD+12] Theodore Alexandrov, Silvia Bianconcini, Estela Bee Dagum, Peter Maass, and Tucker S McElroy, *A review of some modern approaches to the problem of trend extraction*, Econometric Reviews **31** (2012), no. 6, 593–624.

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeeton University Press, 2008.

[BCF97] Philippe C Besse, Hervé Cardot, and Frédéric Ferraty, *Simultaneous non-parametric regressions of unbalanced longitudinal data*, Computational Statistics & Data Analysis **24** (1997), no. 3, 255–270.

[BD06] Peter J Brockwell and Richard A Davis, *Introduction to time series and forecasting*, Springer Science & Business Media, 2006.

[Bha43] A. Bhattacharya, *On a measure of divergence between two statistical populations defined by their probability distributions.*, Bulliton Calcutta Mathemtatics Society **35** (1943), 99–109.

[BHT78] George EP Box, Steven Hilimer, and George C Tiao, *Analysis and modeling of seasonal time series*, Seasonal analysis of economic time series, NBER, 1978, pp. 309–344.

[BJGS12] Ziv Bar-Joseph, Anthony Gitter, and Itamar Simon, *Studying and modelling dynamic biological processes using time-series gene expression data*, Nature Reviews Genetics **13** (2012), no. 8, 552–564.

[Bos12] Denis Bosq, *Linear processes in function spaces: theory and applications*, vol. 149, Springer Science & Business Media, 2012.

[BP67] FE Browder and W Vi Petryshyn, *Construction of fixed points of nonlinear mappings in hilbert space*, Journal of Mathematical Analysis and Applications **20** (1967), no. 2, 197–228.

[BR98] Babette A Brumback and John A Rice, *Smoothing spline models for the analysis of nested and crossed samples of curves*, Journal of the American Statistical Association **93** (1998), no. 443, 961–976.

[BS13] Martin Philip Bendsoe and Ole Sigmund, *Topology optimization: theory, methods, and applications*, Springer Science & Business Media, 2013.

[BY01] Yoav Benjamini and Daniel Yekutieli, *The control of the false discovery rate in multiple testing under dependency*, Annals of statistics (2001), 1165–1188.

[CCMT90] Robert B Cleveland, William S Cleveland, Jean E McRae, and Irma Terpenning, *Stl: A seasonal-trend decomposition procedure based on loess*, Journal of Official Statistics **6** (1990), no. 1, 3–73.

[CDH+15] Wen Cheng, Ian L Dryden, Xianzheng Huang, et al., *Bayesian registration of functions and curves*, Bayesian Analysis (2015).

[Cen00] Nikolai Nikolaevich Cencov, *Statistical decision rules and optimal inference*, no. 53, American Mathematical Soc., 2000.

[CFF06] Antonio Cuevas, Manuel Febrero, and Ricardo Fraiman, *On the use of the bootstrap for estimating functions with functional data*, Computational statistics & data analysis **51** (2006), no. 2, 1063–1074.

[CH66] Richard Courant and David Hilbert, *Methods of mathematical physics*, vol. 1, CUP Archive, 1966.

[CiA02] Andrzej Cichocki and Shun ichi Amari, *Adaptive blind signal and image processing: Learning algorithms and applications*, John Wiley and Sons, LTD, 2002.

[CMP+06] Guillaume Charpiat, Pierre Maurel, Jean-Philippe Pons, Renaud Keriven, and Olivier Faugeras, *Generalized gradients: Priors on minimization flows*, Tech. report, Ecole Nationale des Ponts et Chaussees, 2006.

[CNFT06] Ana Conesa, María José Nueda, Alberto Ferrer, and Manuel Talón, *masigpro: a method to identify significantly differential expression profiles in time-course microarray experiments*, Bioinformatics **22** (2006), no. 9, 1096–1102.

[Con99] John B. Conway, *A course in operator theory*, American Mathematical Society, 1999.

[CQS+08] Damien Chaussabel, Charles Quinn, Jing Shen, Pinakeen Patel, Casey Glaser, Nicole Baldwin, Dorothee Stichweh, Derek Blankenship, Lei Li, Indira Munagala, et al., *A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus*, Immunity **29** (2008), no. 1, 150–164.

[CT76] William P Cleveland and George C Tiao, *Decomposition of seasonal time series: A model for the census x-11 program*, Journal of the American statistical Association **71** (1976), no. 355, 581–587.

[CVK03] Gerda Claeskens and Ingrid Van Keilegom, *Bootstrap confidence bands for regression curves and their derivatives*, Annals of Statistics (2003), 1852–1884.

[CWS16] Jason Cleveland, Wei Wu, and Anuj Srivastava, *Norm-preserving constraint in the Fisher–Rao registration and its application in signal estimation*, Journal of Nonparametric Statistics **28** (2016), no. 2, 338–359.

[D+92] Ingrid Daubechies et al., *Ten lectures on wavelets*, vol. 61, SIAM, 1992.

[Deg11] David A Degras, *Simultaneous confidence bands for nonparametric regression with functional data*, Statistica Sinica (2011), 1735–1765.

[Dră03] Sorin Drăghici, *Data analysis tools for dna microarrays*, CRC Press, 2003.

[E+79] B Efron et al., *Bootstrap methods: Another look at the jackknife*, The Annals of Statistics **7** (1979), no. 1, 1–26.

[Eck12] Andreas Eckner, *A note on trend and seasonality estimation for unevenly-spaced time series*, 2012.

[ET94] Bradley Efron and Robert J Tibshirani, *An introduction to the bootstrap*, CRC press, 1994.

[ET07] Bradley Efron and Robert Tibshirani, *On testing the significance of sets of genes*, The annals of applied statistics (2007), 107–129.

[FHT01] Jerome Friedman, Trevor Hastie, and Robert Tibshirani, *The elements of statistical learning*, vol. 1, Springer series in statistics Springer, Berlin, 2001.

[Fra02] J. B. Fraleigh, *A first course in abstract algebra*, Pearson, 2002.

[FZ00] Jianqing Fan and Wenyang Zhang, *Simultaneous confidence bands and hypothesis testing in varying-coefficient models*, Scandinavian Journal of Statistics **27** (2000), no. 4, 715–731.

[GG04] Daniel Gervini and Theo Gasser, *Self-modelling warping functions*, Journal of Royal Statistics Society **66 part 4** (2004), 959–971.

[GGC13] Jeff Goldsmith, Sonja Greven, and CIPRIAN Crainiceanu, *Corrected confidence bands for functional data using principal components*, Biometrics **69** (2013), no. 1, 41–51.

[Gho01] Sucharita Ghosh, *Nonparametric trend estimation in replicated time series*, Journal of statistical planning and inference **97** (2001), no. 2, 263–274.

[GK64] Michael D Godfrey and Herman F Karreman, *A spectrum analysis of seasonal adjustment*, Citeseer, 1964.

[GN70] David M Grether and Marc Nerlove, *Some properties of optimal seasonal adjustment*, Econometrica: Journal of the Econometric Society (1970), 682–703.

[HK12] Lajos Horváth and Piotr Kokoszka, *Inference for functional data with applications*, vol. 200, Springer Science & Business Media, 2012.

[HKO01] Aapo Hyvarinen, Juha Karhunen, and Erkki Oja, *Independent component analysis*, John Wiley and Sons, Inc, 2001.

[HST15] Boris P Hejblum, Jason Skinner, and Rodolphe Thiébaut, *Time-course gene set analysis for longitudinal gene expression data*, PLoS Comput Biol **11** (2015), no. 6, e1004310.

[HT82] Steven C Hillmer and George C Tiao, *An ARIMA-model-based approach to seasonal adjustment*, Journal of the American Statistical Association **77** (1982), no. 377, 63–70.

[HT83] Andrew C Harvey and PHJ Todd, *Forecasting economic time series with structural and Box-Jenkins models: A case study*, Journal of Business & Economic Statistics **1** (1983), no. 4, 299–307.

[Jam07] Gareth M. James, *Curve alignment by moments*, The Annals of Applied Statistics **1 No.2** (2007), 480–501.

[JCS+09] Miyoun Jung, Ginmo Choug, Ganesh Sundaramoorthi, Luminita A. Vese, and Alan L. Yuille, *Sovolev gradient and joint variational image segmentation, denoising and deblurring*, Tech. report, UCLA C.A.M., 2009.

[JKSJ07] Shantanu H Joshi, Eric Klassen, Anuj Srivastava, and Ian Jermyn, *A novel representation for riemannian analysis of elastic curves in rn*, 2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007, pp. 1–7.

[Kar77] Hermann Karcher, *Riemannian center of mass and mollifier smoothing*, Communications on pure and applied mathematics **30** (1977), no. 5, 509–541.

[Ken90] Wilfrid S Kendall, *Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence*, Proceedings of the London Mathematical Society **3** (1990), no. 2, 371–406.

[KG92] Alois Kneip and Theo Gasser, *Statistical tools to analyze data representing a sample of curves*, The Annals of Statistics **20 No. 3** (1992), 1266–1305.

[KLMR00] A. Kneip, X. Li, K. B. MacGibbon, and J. O. Ramsay, *Curve registration by local regression*, The Canadian Journal of Statistics **28 No. 1** (2000), 19–29.

[Kre89] Erwin Kreyszig, *Introductory functional analysis with applications*, vol. 1, wiley New York, 1989.

[KSW11] Sebastian A Kurtek, Anuj Srivastava, and Wei Wu, *Signal estimation under random time-warpings and nonlinear signal alignment*, Advances in Neural Information Processing Systems, 2011, pp. 675–683.

[LM04] Xueli Liu and Hans-Georg Müller, *Functional convex averaging and synchronization for time-warped random curves*, Journal of the American Statistical Association **99** (2004), no. 467, 687–699.

[LM06] Xiaoyan Leng and Hans-Georg Müller, *Classification using functional data analysis for temporal gene expression data*, Bioinformatics **22** (2006), no. 1, 68–76.

[LT92] Zhi-Quan Luo and Paul Tseng, *On the convergence of the coordinate descent method for convex differentiable minimization*, Journal of Optimization Theory and Applications **72** (1992), no. 1, 7–35.

[Lue69] David Luenberger, *Optimization by vector space methods*, John Wiley and Sons, Inc, 1969.

[Mac13] Henryk Maciejewski, *Gene set analysis methods: statistical models and methodological differences*, Briefings in bioinformatics (2013), bbt002.

[MRS+15] James Stephen Marron, James O Ramsay, Laura M Sangalli, Anuj Srivastava, et al., *Functional data analysis of amplitude and phase variation*, Statistical Science **30** (2015), no. 4, 468–484.

[MYC12] Shujie Ma, Lijian Yang, and Raymond J Carroll, *A simultaneous confidence band for sparse longitudinal regression*, Statistica Sinica **22** (2012), 95.

[Ner64] Marc Nerlove, *Spectral analysis of seasonal adjustment procedures*, Econometrica: Journal of the Econometric Society (1964), 241–286.

[Nes12] Yu Nesterov, *Efficiency of coordinate descent methods on huge-scale optimization problems*, SIAM Journal on Optimization **22** (2012), no. 2, 341–362.

[Ney37] Jerzy Neyman, *Outline of a theory of statistical estimation based on the classical theory of probability*, Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences **236** (1937), no. 767, 333–380.

[Que49] Maurice H Quenouille, *Approximate tests of correlation in time-series 3*, Mathematical Proceedings of the Cambridge Philosophical Society, vol. 45, Cambridge Univ Press, 1949, pp. 483–484.

[Ram06] James O Ramsay, *Functional data analysis*, Wiley Online Library, 2006.

[RD91] James O Ramsay and CJ Dalzell, *Some tools for functional data analysis*, Journal of the Royal Statistical Society. Series B (Methodological) (1991), 539–572.

[RL98] J. O. Ramsay and Xiaochun Li, *Curve registration*, Journal of Royal Statistics Society **60** (1998), 351–363.

[Ron01] Birgitte Ronn, *Nonparametric maximum likelihood estimation for shifted curves*, Journal of Royal Statistics Society **63, Part 2** (2001), 241–259.

[RS02] J. O. Ramsay and B.W. Silverman, *Applied functional data analysis: Methods and case studies*, Springer-Verlag New York, Inc., 2002.

[RSM14] Lars Lau Rakêt, Stefan Sommer, and Bo Markussen, *A nonlinear mixed-effects model for simultaneous smoothing and registration of functional data*, Pattern Recognition Letters **38** (2014), 1–7.

[SC78] H. Sakoe and S. Chiba, *Dynamic programming algorithm optimization for spoken word recognition*, IEEE Transactions on Acoustics, Speech, and Signal Processing **ASSP-26** (1978), no. 1, 43–49.

[SK16] A. Srivastava and E. Klassen, *Functional and shape data analysis*, Springer Series in Statistics, 2016.

[SKJJ11] Anuj Srivastava, Eric Klassen, Shantanu Joshi, and Ian Jermyn, *Shape analysis of elastic curves in Euclidean space*, IEEE Transactions on Pattern Analysis and Machine Intelligence **33** (2011), 1415–1428.

[SL87] Steven G Self and Kung-Yee Liang, *Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions*, Journal of the American Statistical Association **82** (1987), no. 398, 605–610.

[SSVV10a] Laura M. Sangalli, Piercesare Secchi, Simone Vantini, and Valeria Vitelli, *Functional clustering and alignment methods with applications*, Communications in Applied and Industrial Mathematics **1** (2010), 205–224.

[SSVV10b] ———, *K-mean alignment for curve clustering*, Computational Statistics and Data Analysis **54** (2010), 1219–1233.

[ST03] John D Storey and Robert Tibshirani, *Statistical significance for genomewide studies*, Proceedings of the National Academy of Sciences **100** (2003), no. 16, 9440–9445.

[STM+05] Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*, Proceedings of the National Academy of Sciences **102** (2005), no. 43, 15545–15550.

[SWK+11] Anuj Srivastava, Wei Wu, Sebastian Kurtek, Eric Klassen, and JS Marron, *Registration of functional data using Fisher-Rao metric*, arXiv:1103.3817 (2011).

[SXL+05] John D Storey, Wenzhong Xiao, Jeffrey T Leek, Ronald G Tompkins, and Ronald W Davis, *Significance analysis of time course microarray experiments*, Proceedings of the National Academy of Sciences of the United States of America **102** (2005), no. 36, 12837–12842.

[SZ92] Jan Sokolowski and Jean-Paul Zolesio, *Introduction to shape optimization*, Springer, 1992.

[TM01] Paul Tseng and O. L. Mangasarian, *Convergence of a block coordinate descent method for nondifferentiable minimization*, Journal of Optimization Theory and Applications **109, No. 3** (2001), 475–494.

[TM08] Rong Tang and Hans-Georg Muller, *Pairwise curve synchronization for functional data*, Biometrika **95:4** (2008), 875–889.

[TM09] Rong Tang and Hans-Georg Müller, *Time-synchronized clustering of gene expression trajectories*, Biostatistics **10** (2009), no. 1, 32–45.

[TS54] Read D Tuddenham and Margaret M Snyder, *Physical growth of california boys and girls from birth to eighteen years.*, Publications in child development. University of California, Berkeley **1** (1954), no. 2, 183.

[Tuk58] John W Tukey, *Bias and confidence in not-quite large samples*, Annals of Mathematical statistics, vol. 29, INST MATHEMATICAL STATISTICS IMS BUSINESS OFFICE-SUITE 7, 3401 INVESTMENT BLVD, HAYWARD, CA 94545, 1958, pp. 614–614.

[WG97] Kongming Wang and Theo Gasser, *Alignment of curves by dynamic time wapring*, The Annals of Statistics **25 No.3** (1997), 1251–1276.

[YMW05] Fang Yao, Hans-Georg Müller, and Jane-Ling Wang, *Functional data analysis for sparse longitudinal data*, Journal of the American Statistical Association **100** (2005), no. 470, 577–590.

[ZC+07] Jin-Ting Zhang, Jianwei Chen, et al., *Statistical inferences for functional data*, The Annals of Statistics **35** (2007), no. 3, 1052–1079.

BIOGRAPHICAL SKETCH

Liang-Hsuan Tai was born in Taitung, Taiwan (R.O.C). Taitung is a very small city with population less then 50,000 and a slow-paced life style. Liang-Hsuan enjoyed his childhood in this small town. Liang-Hsuan attended National Sun Yet-sen university in Kaohsiung and graduated with honors with a Bachelor Degree in Mathematics in 2010. Afterward, he spent one year mandatory military service as a navy lieutenant. In 2011, Liang-Hsuan was enrolled in the Applied and Computational Mathematics graduate program at Florida State University and defended his Ph.D. dissertation in April 11, 2017. Liang-Hsuan enjoyed six years of instructing seven college-level math courses and researching functional data analysis, optimization, and statistical modeling. For the following decades, Liang-Hsuan will keep expanding his knowledge in mathematics and statistics as he believes a life-long learning attitude.