2017

Modeling of biological data using longitudinal intraindividual means integrated with first and second power time-derivatives

Brandon Skylar Klinedinst
Iowa State University

Modeling of biological data using longitudinal intraindividual means integrated with first and second power time-derivatives

by

Brandon Skylar Klinedinst

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Interdisciplinary Graduate Studies

Program of Study Committee:
Auriel Willette, Major Professor
Elizabeth Stegemoller
Ranjan Maitra

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this thesis. The Graduate College will ensure this thesis is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

TABLE OF CONTENTS

ABSTRACT

This thesis will elucidate a longitudinal approach using intraindividual means, time-derivative values, and integration of these numbers to strengthen investigations that use linear models. This approach enhances the value of corrected responses and strengthens the quality of inference derived from independent variables upon dependent variables. Demonstration is given to several relevant examples in life sciences. There is potential that these efforts eventually contribute to biological study and the accessibility of empirical study itself.

CHAPTER 1. INTRODUCTION AND BACKGROUND

This paper presents results that improve on the value and utility of longitudinal linear models. I argue for the use of longitudinal data to produce and integrate intraindividual averages and their time derivative values, to better account for change in a given outcome measure over time. The problem addressed is that we want to improve upon both inference and validity in regression-based models and diagnostics (i.e. of diseases, states of mind, progression in development, etc). Additionally, one barrier is that similar methods, such as Jacobian determinants (Keller and Roberts, 2008) and growth curve models (Duncan, Duncan, and Strycker, 2013), involve convoluted estimation procedures which sacrifice simplicity for the claim that convoluted algorithms provide superior convergence to the true values. To address this point, a walkthrough is also provided on how to compute intraindividual and time-derivative values from a simple-to-use laymen's algorithm.

Biological systems and treatment of data from biological situations, characteristically and necessarily complicated, can be described by calculus and differential equations (often of more than one variable) in many scenarios. The orthodox reductionist approach focuses on pairing a biological outcome with a simple mathematical model based on isolated data. In the context of longitudinal variation in the outcome, this approach does not eliminate various interactive and iterative components that could otherwise reduce both noise and degrees of freedom, increasing complexity and error, as well as decreasing interpretability and validity of model inferences in curious and interesting ways. Regulation and oscillation of feedback control systems, homeostasis, one and two-compartment diffusion processes, metabolic turnover, enzyme reactions, blood flow measurement,

electrical activity, and drug metabolism represent some of the types of applied problems commonly subjected to computational analysis in the fields of biology, physiology, and medicine. Methods of curve fitting transformation techniques (e.g. Laplace transforms, partial fractional expansion, and the convolution interval), and systems of equations depict and particularize techniques where there is an emphasis to obtaining formal numerical results to infer about processes and make predictions.

Time-derivatives are values that can be computed using calculus on longitudinally observed data. Three computations are particularly noteworthy: position, velocity, and acceleration. The position describes where a subject is within the overall population distribution. Velocity describes how quickly a subject is changing from the position and in what direction (i.e. increasing or decreasing), and acceleration describes whether the subject is changing faster and faster or slower and slower over time. This manuscript advances demonstrations to elucidate how the calculus is used to assess time-derivative values and judge their predictive application in research by comparing their statistical properties, particularly including whether or not their linear models produce smaller error ranges for model estimates and greater R^2 for dependent variables.

There are many longitudinal techniques which have been developed and employ time-derivatives. Sometimes these methods pertain to how the time-derivative values are computed, other times they involve how the time-derivatives are implemented in models. For instance, structural equation growth curve models often evaluate each component (e.g. position, velocity, acceleration) individually in models. Other methods evaluate the components in a combined manner, such as multiplying the estimate of position by the sum of the time-derivatives, and entering

this in a linear model. Another issue is how to estimate the position. Many previous methods estimated the position using baseline observations.

This application estimates the position using an intraindividual average of all the time point observations under consideration. The demonstrations that follow suggest this is a superior estimate for the position because the procedure reduces noise and increases correlation strength, both of which address issues of validity. Additionally, the values of the position, velocity, and acceleration are integrated into a single component, giving them superior performance in statistical models because they elucidate more robust relationships between variables. This step of the procedure offered here has the added benefit of minimizing the total number of variables (vectors) that the research is working with during the modeling stages of data analysis.

The remainder of the introduction reviews background information on aging and Alzheimer's disease. The participants and their biological data used in these demonstrations, body mass index (BMI), radiolabeled fluorodeoxyglucose (FDG), and structural magnetic resonance imaging (sMRI), are described in Chapter 2. Chapter 3 is dedicated to the assessment of time-derivatives with vector operations using a numerical analysis technique known as the rectangle method (Weerakoon & Fernando, 2000). In chapter 4 some properties of time-derivative variables are explored, including changes in their means, variance, correlations, and discernibility between subpopulations. The goal of chapter 5 is to compare a MANCOVA (Multivariate Analysis of Variance and Covariance) model utilizing the latent time-derivative values to both a MANCOVA model utilizing cross-sectional, baseline values and a mixed effects model utilizing the same values from the latent time-

derivative model. A discussion of the results (chapter 6) and conclusion (chapter 7) carry through and bring the paper to a close by summarizing the findings, drawing appropriate inferences, and suggesting avenues of future research into the application of latent time-derivatives and the areas under them.

Background on Alzheimer's disease and metabolic etiology

Radiographic imaging is used to investigate Alzheimer's disease (AD), which is associated with abnormal changes in tissue volume that are distinct from normal aging (Desikan et al., 2009). Of the most common neurological disorders, AD is the most prevalent with ~45 million affected worldwide. AD comprises two-thirds of all dementia cases compared to Lewy Body Dementia (LBD) and Parkinson's disease (Vos et al., 2016). Characteristically associated with age > 65 years, AD can occur younger. AD presents with insidious memory loss as progressive dementia with expectation of pathological atrophy in various cerebral regions usually including medial temporal lobes, lateral and medial parietal lobes, and orbitofrontal cortex (Förstl & Kurz, 1999). Brain atrophy occurs with age even without AD, but at attenuated velocities of change.

AD is histopathologically understood as neuritic plaques containing neurofibrillary tangles of hyperphosphorylated *tau* filaments (TAU/pTAU) and accumulated amyloid in blood vessel walls and leptomeninges (Tiraboschi et al., 2004). Amyloid plaques alone are not diagnostic since they can be found in people with normal memory. One early landmark is falling at least 1.5 standard deviations below normal on a memory test predicting ~50% chance of progression to AD diagnosis. Although non-diagnostic in the absence of substantial memory loss, another protein linked to early AD that deposits in brain tissues additional to *tau,* is

beta-amyloid - both are assayable in cerebrospinal fluid by lumbar puncture (Seeley, 2012).

Comorbidities with cardiovascular disease and diabetes (Craft, 2007) are not uncommon, spurring research into their common underlying etiologies. Unfortunately, the Global Burden of Disease report estimated a worldwide 30% increase in the prevalence of diabetes in just the decade-span from 2005 to 2015. Diabetes predominates as a disease of insulin resistance (IR), whereas changes seem to occur in the affinity of agonists (specifically insulin) to the insulin receptor, followed later by decreased insulin production. Overall, systemic IR is understood as a multifactorial metabolic situation with below-normal ability of insulin to adequately perform physiologic functions.

Insulin receptors are of the tyrosine kinase receptor class and are activated by insulin, IGF1, and IGF2. Once active it phosphorylates protein substrates that are involved in either generating metabolic effects (such as glucose uptake) or the promotion of cellular growth (Ballotti et al., 1989). Blocked receptors prevents the insulin hormone from binding, thus blocking glucose from entering the cell and causing excess accumulation in blood. IR is a complex looped cascade of insulin dysfunction equivalent to glucose toxicity from dysregulation (Robertson et al., 2004). A targeted tissue for insulin-stimulated glucose disposal is skeletal muscle, an anatomic site of aberrant lipid accumulation in IR (Hulver, 2003). As lipolysis and free fatty acids (FFA) increase, lipotoxicity increases. As adverse adipokine secretions change, adiponectin decreases or leptin increases, and as glucose uptake decreases, glucose toxicity increases, and cells become starved or are forced to use more cumbersome and metabolically unfavorable sources of energy.

Provided adequate β-cell function remains intact in nascent stages, IR implies abundant insulin but a deficiency state metabolically. Over years as blood sugar levels increase, amylin proteins glycate, misfold, and promote atrophy of pancreatic β-cells as they are deposited (Höppener et al., 2000). This mechanism may explain transition from functional adequacy of endogenous insulin to artificial life-sustainment on exogenous insulin from recombinant DNA technology. Through similar means this could also account for the amyloidosis observed in AD (Ho et al., 2004).

Neurons only utilize glucose and ketones converted from fats as primary fuel sources (Bhattacharya and Datta, 1993), and half of whole body utilization of glucose in the fasting state occurs in the brain (Yki-Jarvinen, 2011). Moreover, individuals with AD have reduced metabolism of glucose in particular brain regions (Willette et al., 2015). However, beta-oxidation of fatty acids beyond the blood-brain barrier over longer periods presents unfavorable metabolic states characterized by increased risk for hypoxia due to greater oxygen consumption compared to glycolysis, greater pro-oxidant to antioxidant ratios because of superoxide production, and slower reaction times to delivering usable fuel to neurons (Schönfeld and Reiser, 2013).

CHAPTER 2. MATERIALS AND METHODS

Data used were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The cohort included 751 adults ages 55-90 who were diagnosed with varying degree of cognitive health ranging from normal, to the prodrome of Alzheimer's disease with mild cognitive impairment (MCI), to Alzheimer's disease (AD). A stratified mixed effects model and two stratified MANCOVA models were computed on data observed at five times in three years. One MANCOVA utilized latent time-derivative variables and the other used cross-sectional variables at baseline. Dependent variables were structural Magnetic Resonance Imaging (sMRI) volumetric data of brain grey matter and independent variables were Positron Emission Tomography (PET), Fluorodeoxyglucose data (FDG), representing metabolic rate in the brain, and measures of participant Body Mass Index (BMI). Models were further adjusted with participant age and sex.

sMRI data are intensity values representing amount of tissue in an area, thus making it useful for characterizing atrophy and growth in the nervous system. The FDG intensity values represent amount of metabolism in a region of brain tissue. PET and sMRI are techniques utilized in many fields, but in the present neuroscience demonstration this study pertained to the nervous system of aged adult humans. Altogether, the comparative distributions, means, variances, covariances (correlations), and model estimates are compared. Enumerations of intraindividual growth and atrophy were assessed using the calculus with vector computations in R to produce intraindividual averages, linear change, and quadratic change for each participant, described in the next chapter.

CHAPTER 3. ASSESSING TIME-DERIVATIVES WITH CALCULUS

The values assessed here are used in the demonstrations to follow. Below is an illustration of the vector operations involved in the rectangle method of estimating rates of change and the area under the growth/atrophy curve (**Figure 1**).

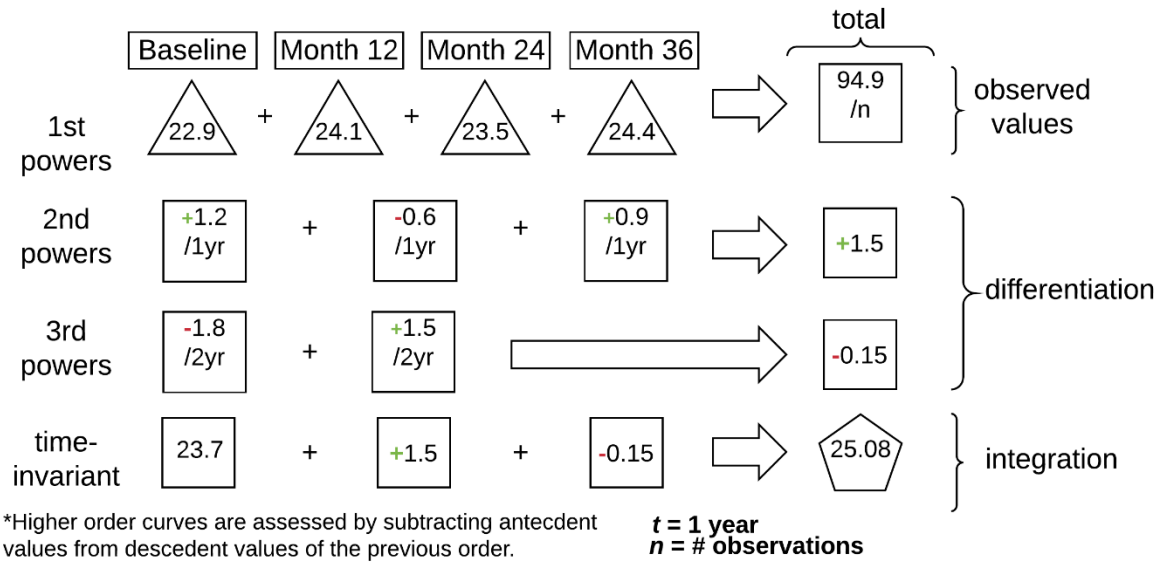

Figure 1. Demonstration of how time-derivative values are produced from temporal observations of a subject's BMI.

The final outcomes (the pentagon in **Figure 1**) are obtained using principles of vector calculus whereas a numerical analysis algorithm known as the rectangle method provides an estimation for two calculus fundamental theorems -- differentiation and integration. Differentiation, the method of difference ratios between quantities (BMI and Time in this instance), produces a series of new quantities that describe rates of change observed in trajectories. The quantities are also known as time-derivatives, because derivatives are calculated with respect to their denominator. Each new component contains information describing rates of

change observed in trajectories in vector-like terms of direction (accretion or depletion) and magnitude in a given reference frame.

The values observed for each timepoint are organized into distinct vectors, such that the row indices for each represents the same case. Completion of the process requires six steps: (1) Temporal means of the intraindividual values are calculated and stored in a separate vector. (2) New vectors of temporal difference values, called latent linear change variables in the time-derivative paradigm, are calculated for each immediate time point after and before the time point at hand, such that vectors with antecedent observations are subtracted from descendant vectors. (3) The resulting difference vectors must be each divided by the respective lapse of time. In this instance divide by 1 because 1 year has lapsed and *t* is defined as 1 year. If however *t* were defined as 6 months, then divide by 2 in this instance. (4) These first three procedures are repeated on the vectors of linear change to produce vectors of quadratic change, and such assessments could theoretically be continued infinitesimally until only one vector remains and therefore is not differentiable.

Integration, the method of summing over the derivatives within a range of the denominator (e.g. time), compacts the information from each rate of change. (5) With the rates of change vectors assessed, they are integrated into a single new value. (6) These values are the coefficients that compose the algebraic trajectory models (**Formula 1**). The solutions of these models are used in the coming demonstrations.

For each subject *i*

$$\mathbf{Y_i = \beta_0 + \beta_1 X_i + \beta_2 X^2_i}$$

Formula 1. Each β is a subject-specific position or rate of change component of each i^{th} trajectory.

The reference frame of the observer, the spatial and temporal conditions of each observation, have two properties that impact the properties of temporally-derived variables. First is the resolution of the reference frame, and the second is the range of view of the observer. Holding the observer's range of view constant, each additional temporal observation provides for the estimation of the next power of time, further increasing the observer's resolution of the trajectories and their rates of change (**Figure 2**).

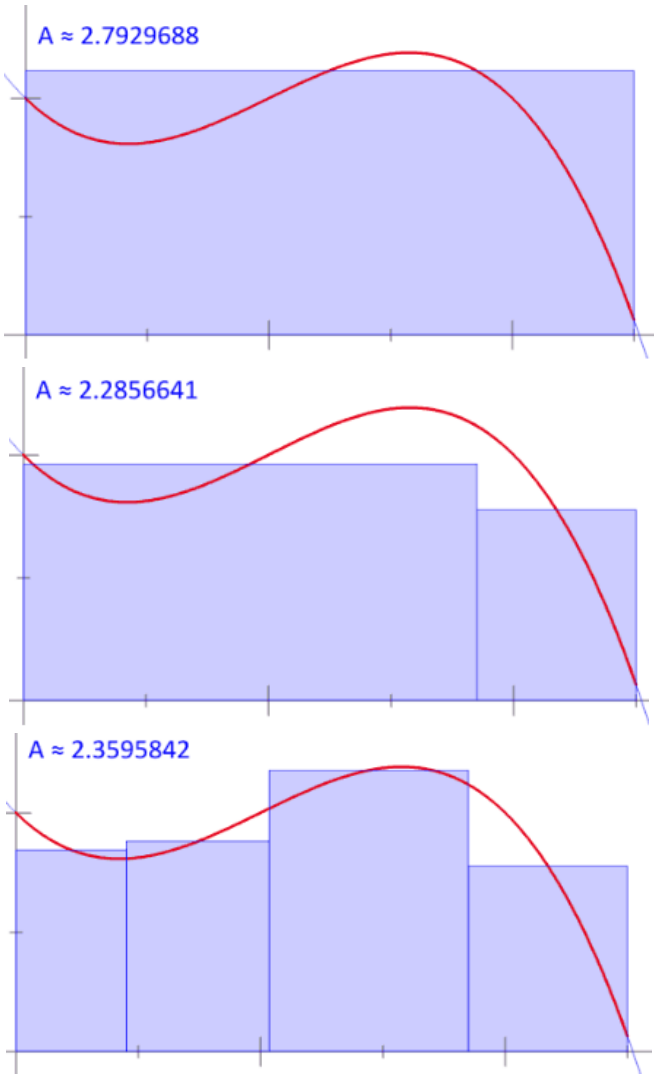

Figure 2. Algebraic curves and estimating area in the discrete case.

A trajectory's resolution is ultimately limited by the number of temporal observations. Increased resolution is useful in identifying rapid fluctuations that occur during, for example, critical periods of development by mapping local minima and maxima. More complex trajectories also require greater resolution for their accuracy. Higher resolution during pertinent critical periods can aid in advancing theory and understanding of phenomenon of natural aging, developmental, and infectious disease processes.

The second property is the observer's range of view, which is the portion of the window of time over which the trajectory is observed from its start to completion. For instance, was it observed from the trajectory's onset until the time of the end of the phenomenon, or was only a cross-sectional "snapshot" captured? Shorter ranges of view have less information to contribute.

CHAPTER 4. STATISTICAL PROPERTIES OF TIME-DERIVATIVES

The transformation of temporal observations into time-derivative variables impact their mean, variance, and covariance. Recall that the first stage in the methods is to compute the variable's intraindividual average. This produced variables with smaller sample variances, which may indicate decreases (or eliminations) in the components of variance that were infiltrated with random noise. Observed means also shift towards their 'true value' (**Figure 3**). Decreasing random noise translates into smaller likelihoods of observing spurious results and improved correction factors.

After integrating the rate of change vectors with the averaged density vector, the variance has subsequent modest increases, the magnitude of which is suspected to be dependent on the range of view of the observers, therefore supporting the notion that the time-derivative variables have "truer information" available for describing dependent variables. Because the components of variance, those which are neither measurement error nor random noise, represent relationships between variables, the net outcome is that the degree of correlation and covariation increase across the system (**Figure 4**). For instance, note the correlation between the entorhinal cortex and hippocampus increased from 0.63 for cross-section variables to 0.70 for these time-derivative variables.

Variable	Cross-section	Temporally-averaged	Time-derivatives
Body Mass Index	26.25	26.13	26.10
Frontal Lobe FDG	0.82	0.87	0.88
Temporal Lobe FDG	0.91	0.93	0.93
Frontal Pole Volume	639.27	627.21	624.39
Lateral Orbital Frontal Volume	6,209.06	6,101.29	6,071.91
Medial Orbital Frontal Volume	3,749.04	3,676.13	3,657.22
Hippocampal Volume	2,956.58	2,866.99	2,842.94
Entorhinal Cortex Volume	1,707.5	1,641.48	1,624.73
Parahippocampus	1,876.2	1,820.28	1,806.01

Figure 3. This table is comparing the means of **Cross-section**, **Temporally-averaged**, and **Time-derivative** variables. Note that the means' directions of change occurring between cross-section to temporally-averaged variables, and subsequently between temporally-averaged to mechanical variables always remain consistent.

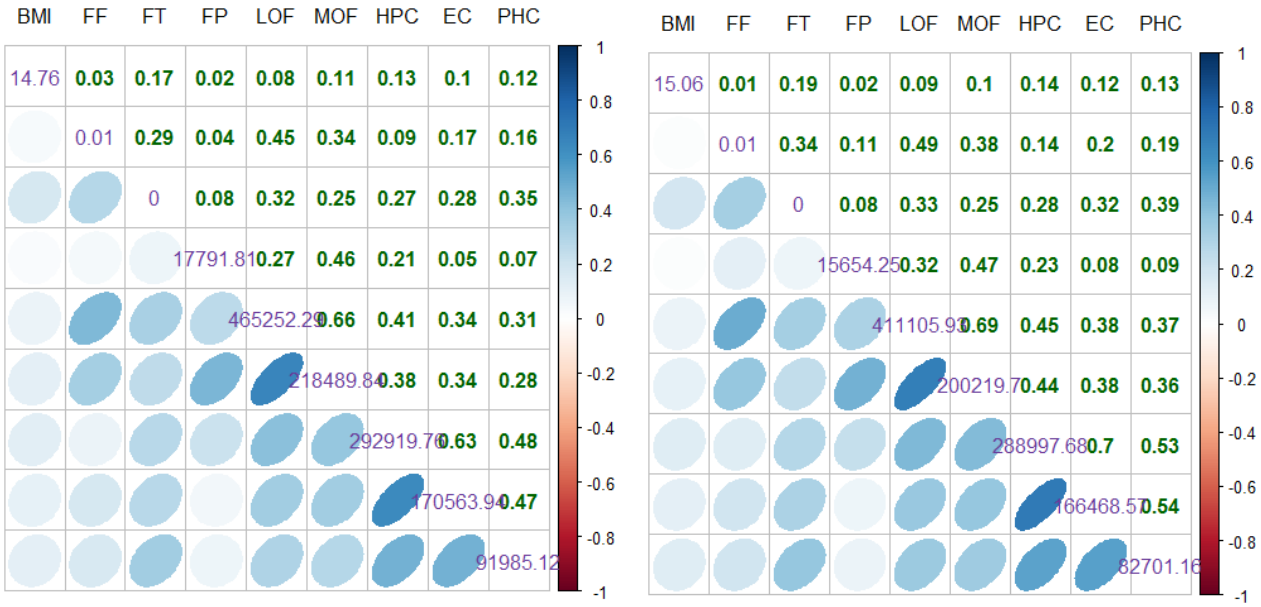

Figure 4. This figure was produced by a function made for R by Dr. Ranjan Maitra at Iowa State University. The higher off-diagonals show the correlations between variables, and the lower off-diagonals graphically illustrate the direction and degree of noise of the relationships. The diagonals represent observed uncertainties.
Legend: BMI (Body Mass Index), FF (FDG_Frontal), FT (FDG_Temporal), FP (FrontalPole), LOF (LateralOrbitalFrontal), MOF (MedialOrbitalFrontal), HPC (Hippocampus), EC (EntorhinalCortex), PHC (Parahippocampus).

Commonly, researchers want to distinguish between groups and subpopulations in their samples. This may commonly be in an experimental setting where conditions are tightly controlled, or perhaps in a clinical setting where the desire is to group and classify cases together. The density plots below shows the distribution of volume in the entorhinal cortices between cognitively normal, prodrome AD, and AD (**Figure 5**).

Comparing the figures on the left and right, the major change to note is that the difference value between the means remained stable for the prodrome AD and AD groups, but the difference between the means of the normal group from the other two groups increased by nearly 10%. This subtle effect will have pronounced results. As the differences in means between factor variables increases, be it a diagnostic label, education level, gender, or genetic variation index, the true effects will be easier to parse out.

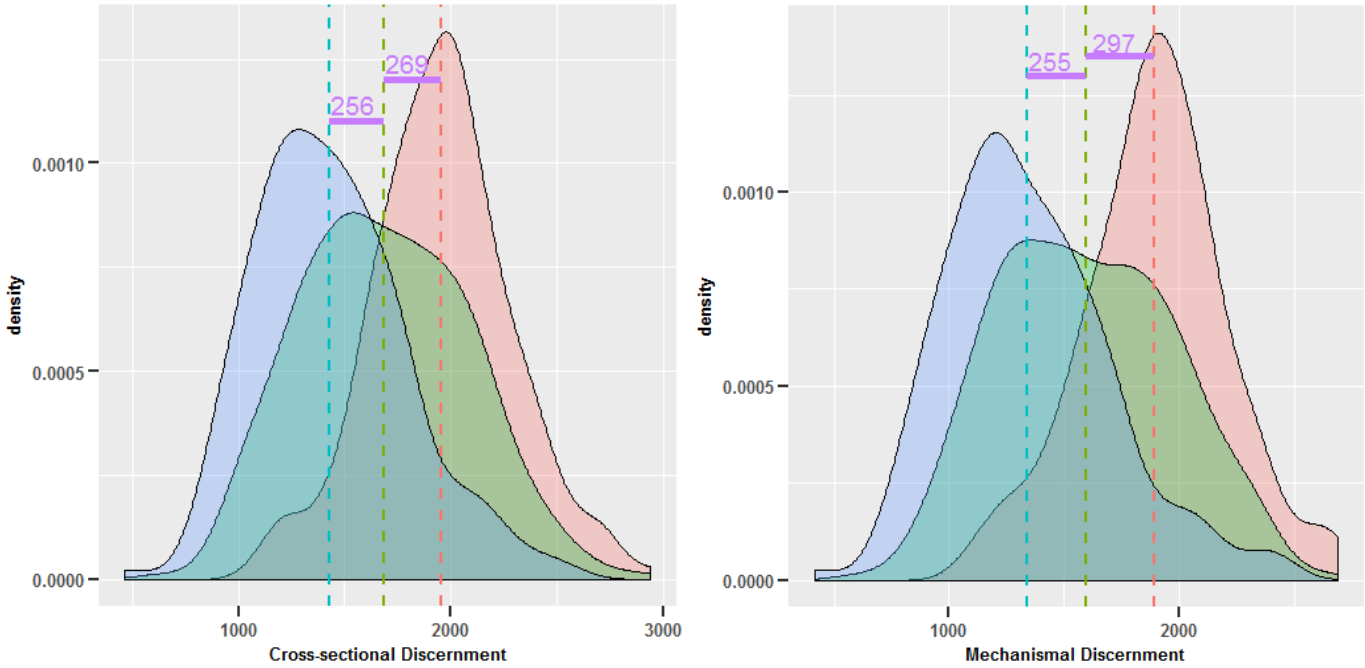

Figure 5. A side-by-side comparison of the density plots for the volume of the entorhinal cortices that are partitioned by diagnosis. **Legend:** Cognitively Normal, Mild Cognitive Impairment (i.e., prodrome AD), Alzheimer's disease

CHAPTER 5. RESULTS OF COMPARING MODEL PARAMETERS

The improvements in the means, variances, and correlations, discussed previously, translated into improvements in linear model parameters. This is of high value because linear models are one of the most widely used mathematical tools. The model estimates for the two MANCOVA models and their standard errors are organized into the tables below.

Interpretation of comparisons between the parameter estimates is less straight-forward than the prior demonstrations. One key finding here is that the directionality of some of the relationships reverse from cross-sectional to time-derivative variables. Out of a total of 54 parameters per model, 9 of them (16.6%) fail to have the same direction in both models. Of these 9 parameters, 7 of them pertain to relationships in the frontal lobe, with 4 of those in the frontal pole alone. The frontal pole is a small region notorious for high variances.

Comparing the ratios of each parameter mean to its standard error, the time-derivative model outperforms the cross-sectional model. Thirty (30) of the ratios are smaller (smaller standard errors, larger means) in the time-derivative models, compared to just 24.

The R^2 values of dependent variables are approximate measures of a model's correction and prediction ability, therefore ascribing them high value in the final judgement of a model's worth. For all three diagnostic groups the time-derivative variables had higher R^2 values for the lateral orbital frontal, entorhinal, and parahippocampal regions. From the 18 R^2 that were computed, 13 were greater in the time-derivative model, while only 5 were greater in the cross-section model. This is

almost a threefold difference, thus the time-derivative variables appear to have greater explanatory power in 72% of cases.

Cross-section	BMI	FF	FT	BMI	FF	FT	BMI	FF	FT
FP	**-1.8 (2.0)**	-17.7 (143)	**476 (230)**	**1.2 (1.9)**	174 (115)	319 (226)	**3.5 (2.9)**	**396 (143)**	**-454 (256)**
LOF	20.2 (8.7)	2616 (578)	**4596 (923)**	0.8 (8.1)	**2967 (457)**	**2701 (942)**	**-26.0 (15.3)**	3140 (676)	3060 (1211)
MOF	7.8 (6.4)	2066 (438)	**2175 (732)**	**8.5 (6.4)**	1646 (377)	**2442 (763)**	2.5 (9.4)	2286 (393)	831 (704)
HPC	10.5 (6.1)	**-922 (462)**	**2078 (760)**	12.3 (6.5)	**-167 (422)**	**1551 (766)**	-14.7 (10.9)	857 (617)	2197 (869)
EC	10.9 (5.0)	**590 (373)**	945 (615)	5.8 (5.4)	126 (342)	**1271 (620)**	-3.5 (8.0)	**262 (422)**	1330 (621)
PHC	8.8 (4.0)	631 (310)	**919 (516)**	**5.5 (4.0)**	-83 (245)	1606 (441)	**4.6 (7.4)**	-52 (383)	**1727 (543)**

Figure 6. This table is for comparing least squares parameter values and their (standard errors) for the cross-sectional model. The columns contain the IVs and the rows contain the DVs. Values with the more desirable estimate-to-standard error ratio are in **bold**. Values colored in red or blue represent a reversed relationship compared to the time-derivative model.

Legend: Cognitively Normal, Mild Cognitive Impairment (i.e., prodrome AD), Alzheimer's disease

In frontal ROIs, the time-derivative models outperformed the mixed effects models in 78% of cases. In the medial temporal ROIs the two methods tied, three explicit wins for each, and three explicit ties. time-derivative models always had greater predictive and corrective ability in participants with Alzheimer's Disease but only had greater ability in four of six regions in Mild Cognitive Impairment and four of six regions in Cognitively Normal participants.

Time-derivative	BMI	FF	FT	BMI	FF	FT	BMI	FF	FT
FP	-1.5 (2.0)	**76 (122)**	352 (245)	-0.3 (2.6)	**360 (117)**	**-455 (231)**	1.9 (1.8)	267 (100)	198 (202)
LOF	**24.6 (8.4)**	**3046 (478)**	3201 (966)	**-29.7 (14.8)**	3378 (615)	2173 (1284)	10.3 (7.6)	**3069 (397)**	**2579 (833)**
MOF	**13.5 (6.3)**	**2491 (362)**	947 (775)	-7.6 (9.1)	**2168 (364)**	425 (722)	**10.8 (6.1)**	**1598 (340)**	**2203 (705)**
HPC	**15.6 (6.7)**	-214 (490)	1891 (886)	**14.3 (6.7)**	74 (414)	1176 (762)	**-16.4 (10.3)**	**951 (540)**	**2236 (868)**
EC	**14.5 (5.1)**	364 (341)	**1073 (619)**	**9.3 (5.5)**	**221 (335)**	1159 (615)	**-8.5 (8.0)**	232 (398)	**1879 (643)**
PHC	**12.1 (3.9)**	**701 (283)**	806 (520)	4.5 (3.8)	**-298 (229)**	**1766 (414)**	0.1 (7.1)	**334 (343)**	1733 (555)

Figure 7. This table is for comparing least squares parameter values and their (standard errors) for the time-derivative model. The columns contain the IVs and the rows contain the DVs. Values with the more desirable estimate-to-standard error ratio are in **bold**. Values colored in red or blue represent a reversed relationship compared to the cross-sectional model.

Legend: Cognitively Normal, Mild Cognitive Impairment (i.e., prodrome AD), Alzheimer's disease

	Cross-sectional			Time-derivative			Mixed-effects		
	C N	MCI	AD	C N	MCI	AD	C N	MCI	AD
FP	0.051	0.047	**0.142**	0.042	**0.069**	0.130	**0.057**	0.019	0.109
LOF	0.471	0.428	0.408	**0.520**	**0.465**	**0.431**	**0.520**	0.423	0.328
MOF	0.376	0.314	**0.433**	**0.447**	**0.338**	0.416	0.435	0.280	0.251
HPC	0.257	0.244	0.320	0.221	0.211	**0.328**	**0.363**	**0.251**	0.217
EC	0.166	0.140	0.195	**0.183**	0.144	**0.220**	**0.187**	**0.187**	0.181
PHC	0.195	0.157	0.259	**0.266**	**0.193**	**0.279**	**0.270**	**0.193**	0.227

Figure 8. This table is for comparing the R^2 values between the cross-sectional, mixed effects, and time-derivative models. The model with the respectively greater R^2 is in **bold**; ties are defined as within a hundredth of a degree or closer.

CHAPTER 6. DISCUSSION

In biological science, hypotheses are used to assist in determining how a complex process should work. Hypotheses are distinct from demonstrations in that any number of hypotheses can be proposed as attempts to explain an undemonstrated, complex phenomenon, such as memory decline in Alzheimer's disease or cardio- and cerebrovascular degradation with age. Demonstrations extol hypotheses and extend the applied sciences. Here, I demonstrated that the use of time derivatives and intraindividual means in many biological systems as components in the analysis of data may help distinguish the more likely, successful, and falsifiable biological hypotheses from less likely ones.

Time-invariant variables can be defined as "the time-invariant changes in variable x over the course of *t*". These values are derived from intraindividually-averaged density values that are combined with their integral component rates of change. As demonstrated, variables whose variances contain components of their rates of change have superior performance across the tables. The variables have significant reductions in their variances that were likely random noise, however these can never be known for certain (Bell, 2001). Their means shift in modest amounts, presumably moving closer towards their truer values. With the help of density plots it was shown how this could help in diagnostics. These effects also suggest that use of these types of variables in models may lessen the need for post-hoc error correction.

Furthermore, the correlations between these variables is enhanced. Greater correlations mean reductions in the dispersions around their fit lines, and the movements in the means (respective to cross-section variables) also probably assists

with this (as they could be said to be corrected towards the fit line). These properties were shown to increase the R^2's of volumetric brain models, and improved standard error-to-mean ratios occurred. It's worth considering that the parameters which did not out-perform the cross-section model may in fact represent a reduction in spurious findings (e.g. a natural reduction in "type 1 & 2 errors"). This claim itself would seem spurious, if not for the substantial consistency of improvements in all the other properties that we examined.

Creating a distinction between cases that are increasing and cases that are decreasing is perhaps the most intriguing property of time-derivative variables of all. These components of variance are advantageous such that truer relationships between variables may be revealed. A great aspect of the linear model demonstrations was the finding that some relationships shifted directionality. One case in point are the three parameters for BMI in the frontal lobe: in a cross-section examination, one might falsely infer a positive relationship, but when time-derivative variables are added to the model, the interpretation changes altogether, such that one would infer a negative relationship between BMI and volume in regions of the frontal lobe.

CHAPTER 7. CONCLUSION

This thesis described how to assess and implement time-derivative variables, which were shown to enhance a researcher's power to discern relative to untransformed variables, whether or not the untransformed variables were used in a longitudinal mixed effects model or a cross-sectional model. Variables produced by the methods within had refined, high-resolution distribution spaces resulting from delineating temporal boundaries and differentiating cases with accreting or depleting values. time-derivative values enhance many properties of variables derived from temporal data, including efficacy, portion of resolved information-space, malleability in linear models, and magnitude of r-squared values.

At present early differential diagnosis and predicting AD conversion are not aided by neuroimaging studies (MRI and FDG) nor EEG. Images may look normal, not revealing disease-associated patterns. AD may take decades to detect (Seeley, 2012). A hope is that one day we can predict and identify persons very early on in their disease processes. The demonstrations offered here may help.

Time-derivative computations simplify ‘big data’. Biologists are mastering how the biological world can be reduced into its component parts as the life sciences are moving towards larger scale, shared, open initiatives in design, collection, and accessibility. Consequently, the ability to organize and model the resulting data is becoming extraordinarily complex. While the intricacies of large, temporal datasets may appear computationally immutable, time-derivative computations reduce entire datasets to cross-section-like variables. Consider the demonstrations expressed here, how quickly they made work of this phenomenon. Superior results from time-invariant models and the longitudinal nature of their computation suggest

observational designs may be advantageous over clunkier experimental designs that are laden with humanitarian, noncompliance, and bureaucratic problems.

It is emphasized that in any scenario it is desired to be able to derive expressions of the curves observed through time, as described in chapter 4 **(Formula 1**), because intuitively cause and effect information must be derived from temporal observation. The parameters in these expressions enumerate information about a variable that can only be gleaned from observing how it behaves over time.

The future aims remain to strengthen linkages of cause and effect, improve modeling, and demonstrate how to reproduce strong inferences with consistency. Future directions include testing how these values function for other types of biological variables, and whether observing processes further out in time (longer than 3 years) may continue to enhance their predictive power. There is now a need to focus on integrating and reconstructing the omics back into their whole, including methods on how to mathematically model entire biological systems.

REFERENCES

Ballotti, R., Le Marchand-Brustel, Y., Gammeltoft, S., & Van Obberghen, E. (1989). Insulin receptor: tyrosine kinase activity and insulin action. *Reproduction Nutrition Development*, *29*(6), 653-661.

Bell, S. (2001). *A beginner's guide to uncertainty of measurement*. Teddington, Middlesex: National Physical Laboratory.

Bhattacharya, S. B., & Datta, A. G. (1993). Is brain a gluconeogenic organ?. Molecular and cellular biochemistry, 125(1), 51-57.

Craft, S. (2007). Insulin resistance and Alzheimer's disease pathogenesis: potential mechanisms and implications for treatment. *Current Alzheimer Research*, *4*(2), 147-152.

Desikan, R. S., Cabral, H. J., Hess, C. P., Dillon, W. P., Glastonbury, C. M., Weiner, M. W., ... & Fischl, B. (2009). Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain*, *132*(8), 2048-2057.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2013). An introduction to latent variable growth curve modeling: Concepts, issues, and application.

Förstl, H., & Kurz, A. (1999). Clinical features of Alzheimer's disease. *European archives of psychiatry and clinical neuroscience*, *249*(6), 288-290.

Ho, L., Qin, W., Pompl, P. N., Xiang, Z., Wang, J., Zhao, Z., ... & Hof, P. R. (2004). Diet-induced insulin resistance promotes amyloidosis in a transgenic mouse model of Alzheimer's disease. *The FASEB journal*, *18*(7), 902-904.

Höppener, J. W., Ahrén, B., & Lips, C. J. (2000). Islet amyloid and type 2 diabetes mellitus. New England Journal of Medicine, 343(6), 411-419.

Hulver, M., Berggren, J., Cortright, R., et. al. Skeletal muscle in lipid metabolism with obesity. Am J. Physiology, Endocrinology and Metabolism. 2003: 284:E741-32.

Keller, S. S., & Roberts, N. (2008). Voxel-based morphometry of temporal lobe epilepsy: An introduction and review of the literature. *Epilepsia*, *49*(5), 741-757.

Robertson, R.P, et. al. Glucose toxicity in β-cells. In: LeRoith D., Taylor S.I., Olefsky J.M. (2004). Diabetes Mellitus, 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 129-137.

Seeley W.W. and Miller B.L. Dementia. In: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J eds. Harrison's Principles of Internal Medicine. 18th ed. New York, NY: McGraw-Hill; 2012: 3300-3316.

Schönfeld, P., & Reiser, G. (2013). Why does brain metabolism not favor burning of fatty acids to provide energy?-Reflections on disadvantages of the use of free fatty acids as fuel for brain. *Journal of Cerebral Blood Flow & Metabolism*, *33*(10), 1493-1499.

Tiraboschi, P., Hansen, L. A., Thal, L. J., & Corey-Bloom, J. (2004). The importance of neuritic plaques and tangles to the development and evolution of AD. *Neurology*, *62*(11), 1984-1989.

Vos, T., Allen, C., Arora, M., Barber, R. M., Bhutta, Z. A., Brown, A., ... & Coggeshall, M. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, *388*(10053), 1545.

Weerakoon, S., & Fernando, T. G. I. (2000). A variant of Newton's method with accelerated third-order convergence. *Applied Mathematics Letters*, *13*(8), 87-93.

Willette, A. A., Bendlin, B. B., Starks, E. J., Birdsill, A. C., Johnson, S. C., Christian, B. T., ... & Jonaitis, E. M. (2015). Association of insulin resistance with cerebral glucose uptake in late middle–aged adults at risk for Alzheimer disease. *JAMA neurology*, *72*(9), 1013-1020.

Yki-Jarvinen H. Pathophysiology of type 2 diabetes mellitus. In Wass JAH and Stewart PM eds. Oxford Textbook of Endocrinology and Diabetes. 2nd ed. New York, NY: Oxford University Press; 2011: 1740-1748.