The Saga of PLS

Gaston Sanchez

April 2015

Chapter 1

Introduction

A saga is a story of heroic events and achievements of a personage or family, typically written in medieval Icelandic or Old Norse in the form of a prose narrative. According to Wikipedia, the term saga originates from the Norse "saga" and refers to "what is said" or "story, tale, history." The closest term in English to the word saga is "saw" (as in *old saying*). In this regard, I'm using the word Saga to play with its meaning, and convey the ideas of history, story, and narrative of events associated to Partial Least Squares methods.

The main motivating trigger behind this book has been my long standing obsession to understand the historical development of Partial Least Squares methods in order to find the who's, why's, what's, when's, and how's. It is the result of an intermittent 10 year quest, tracking bits and pieces of information in order to assemble the story of such methods. Moreover, this text is my third iteration on the subject, following two of my previous works: chapter 2 *"Historical Review"* of my PhD thesis (Sanchez 2009), and the appendix *"A Historical Overview of PLS-PM"* from my book *PLS Path Modeling with R* (Sanchez 2013).

This is NOT a technical book. It doesn't cover theory, methodological aspects, nor technical details of how the various PLS methods work (no discussions about algebra, computational steps, interpretation issues, etc.). This is also not a book written with a particular reader in mind. Instead, I've written its content to organize the vast material I've collected so far, which includes not only information from papers, chapters, proceedings, and books, but also thoughts, memories, analysis, interpretations, and personal opinions. Having said that, I imagine that this book can be used as a **companion reading** for any course, workshop, or seminar about PLS methods, expecting to be enjoyed by anyone interested on this topic. Regardless of whether you're just taking your first steps on the PLS arena, or if you've already traveled a long way around PLS territory, I'm sure you'll find some value in the content of this work.

By writing this book, my aim is to shed some light about the teaching and understanding of the historical background surrounding Partial Least Squares methods. I'm convinced that this material will provide a fresh perspective and create awareness, debunk myths, and clarify misunderstandings. I'm not the one to judge whether the story of PLS has elements of an heroic tale. But I can tell you for sure that this story is formed with a long and complicated series of events, having scandinavian authors as protagonists. Hence the meaning of the title "The Saga of PLS."

The proof of the pudding

In his classic *"Soft Modeling: The Basic Design and Some Extensions,"* Herman (H. Wold 1982a), includes the following footnote on page 24:

> Most nonlinear iterative techniques of estimation are lacking an analytic proof of convergence. The proof of the pudding is in the eating.

Referring to the lack of a proof of convergence for his algorithm—when applied to the general case of multiple relations among multiple sets of variables—, Wold turned to the old saying of the **proof of the pudding** to invite readers to try for themselves his methodology. Michel Tenenhaus (pronounced *Tenen-os*), a French statistician and world renowned PLS expert, has also used the same famous saying when presenting Partial Least Squares methods to new audiences. For many years, Michel has included William Camden's (1623) version "All the proofs of a pudding are in the eating, not in the cooking," in the last slide of most of his presentations about PLS, inviting users to *taste* the method and experience for themselves whether they like it or not.

For almost three decades, the convergence of Herman's main algorithm was an unresolved issue. This was due in part to the various options in which the algorithm can be specified. Depending on the input settings, computations converge most of the times. But there are a few exceptions. Today, we have a much better understanding of the mechanics of the algorithm, and it is possible to prove the convergence under certain circumstances. Unlike Wold, we also have a better comprehension of how the steps operate, and we can even tweak the algorithm in a such a way that we are certain of what's going on inside of it. Thanks to these advancements, a couple of years ago, Tenenhaus slightly modified the pudding phrase to:

> All the proofs of a pudding are in the eating... but it will taste even better if you know the cooking.

He's right. Let me add to it that it doesn't hurt to know about the story and history behind the recipes, their chefs, their cooking styles, and their tasting preferences. If anything, it will make your tasting of the pudding irresistible.

Outline

This book is organized into three major sections:

1. Introductory Review (chapters 2 - 4)
2. Historical Narrative (chapters 5 - 9)
3. Assessment and Conclusions (chapters 10 - 11)

Part 1 covers a general introduction describing general aspects about Partial Least Squares methods. Part 2 presents a historical narrative focused on Herman Wold and the series of events that led him to develop the so-called *PLS Soft Modeling* framework. It also contains a narrative of the role played by Svante Wold and the development of PLS Regression. Instead of merely listing various events and polluting the text with dozens of citations, I've preferred to use a more fluid narrative style. Finally, the third part provides remarks and critical assessments of several points often ignored in the PLS related literature. My purpose is to give you a different perspective of Partial Least Squares that, hopefully, will help you gain insight into a better understanding and comprehension on the subject.

Acknowledgements

I've been able to gather invaluable unpublished information and details from my own involvement around the PLS community, as well as from personal communications with leading *PLSers* (via email, skype, and personal meetings).

My heartfelt thanks to Michel Tenenhaus, Wynne Chin, and Vincenzo Esposito-Vinzi. They all have being very gentle to answer to my emails full of questions and inquiries, providing responses full of details and milimetric precision.

Also many thanks to Christian Ringle, José Luis Roldán, and Aida Eslami for their emails with useful information and enriching comments. Likewise, I would like to thank the various friends, colleagues, and, for the lack of a better term, *followers* around the world that have been helping me proof-read the content, detecting bugs, and fixing typos. I take full responsibility for any prevailing errata in the text.

Last but not least, you wouldn't be reading this book if it wasn't for the patience and support of my loving wife Jessica. Not only she was willing once again to be my household editor, but she never complained my occupying of the dining table, and taking over the living room as my personal workspace. If you find any value from the content in this book, which I am sure you will, you owe her almost as much as you owe me.

Berkeley, California. April 2015.

Chapter 2

You've got to be kidding me

The first time I heard about PLS was in October 2004 when I first met professor Tomàs Aluja—my former PhD adviser. I had just recently moved to Barcelona to start my graduate studies in statistics, and I had been designated Tomàs's new mentee. At the end of our first meeting when I was about to leave his office, Tomàs asked me a seemingly intrascendental question: "Have you ever heard about PLS?"

"P. L. S.?"

I repeated those letters to myself without pronouncing any words. Fearing to give a bad impression to my brand new mentor, I quickly scanned the hard-disk of my brain in search for any hint. Unsuccessfully struggling to find any answer, I ended up with a perfect clueless expression in my face.

I had no idea what PLS was or what it meant.

The reason why Tomàs asked me about PLS was because he was in charge of organizing *PLS'05 - The 4th International Symposium on Partial Least Squares and Related Methods* for the following year (September, 2005). He wanted to know if I would be willing to help him with some of the organizational details for the symposium. Little I knew about the tremendous impact those three mysterious letters would have in the forthcoming years of my analytics career.

I became involved with PLS in a very unintended way. One of my first duties as part of the *PLS'05* staff, was being in charge of receiving the abstracts and paper submissions for the symposium, making sure they complied with the strict guidelines-for-authors. I spent my time sending emails and pestering authors kindly requesting them to fix the page layout, change the font size, reduce the number of pages, and things like that. One of the orders Tomàs gave me was to read all the papers. Grudgingly, I followed his command and sucked it all up. However, *reading* doesn't necessarily mean *understanding*. From all the submitted papers, I was able to understand three of them… at the most. And even then my understanding was very superficial. The common thing about those papers was their very applied and digestable examples that prevented me from getting lost with all the unintelligible details. By the way, this is NOT how anyone should ever try to learn about PLS methods.

I had a more formal and intensive introduction to the PLS framework with Tomàs's course "Advanced Methods for Multivariate Data Analysis." The outline of the course was based on the great book *La Régression PLS: Théorie et Pratique* by Michel (Tenenhaus 1998). The book is a marvelous piece, with Michel's unique eloquent writing style, explaining things in a very detailed way, using a clear notation, and full of examples and graphics. I guess everything would have been fine if it wasn't for the fact that I had a very limited knowledge of French. Which it only makes it harder when you're not only a PLS novice but also a French-language beginner. Luckily, *google translate* was already available at that time, providing a considerable help in deciphering some of the book's content.

Tomàs's class was a very hands-on course. Each student was responsible of studying a chapter from the book, programming the corresponding method in R, and giving a lecture to the rest of the class. I got to present chapter 8—PLS Regression for one response variable. Tomàs kept for himself the presentation of the PLS Path Modeling (PLS-PM) approach for the last lecture. That day he introduced the technique with an example using the famous Russett dataset, and then he proceeded to talk about how the method worked. When he got to the description of the PLS-PM algorithm, however, I was having a hard time understanding all the steps. The worst part was my fruitless attempt trying to connect the PLS regression algorithms with the PLS-PM algorithm. The question that was driving me crazy was: Where on earth were the PLS regressions in the PLS-PM algorithm? I raised my hand and asked Tomàs if he could please show the rest of the class where the PLS regressions were. To my surprise, his answer was even more disconcerting: "Regressions? What regressions? This has nothing to do with PLS Regression. This is PLS Path Modeling." I'm pretty sure Tomàs does not even remember these details. But they definitely had a profound impact on me.

All I remember thinking was something like: “You’ve got to be kidding me, this PLS stuff makes no sense whatsoever!” I was definitely biting more than I was able to chew.

I won’t lie to you. My introduction to the PLS world was not a smooth one. Besides my very limited knowledge of French, another difficulty was my equally limited knowledge of the techniques in Michel’s book like Canonical Correlation Analysis, Redundancy Analysis, and Inter-Battery Factor Analysis. One of the morals from this experience was that I needed to learn French; the other moral was that this was also not the best way to start learning about PLS from scratch.

I felt so confused and disoriented, but eventually I managed to my find way out through the PLS labyrinth over the following years. My biggest consolation, to my suprise, was when I heard Michel Tenenhaus story in front of the audience at one of the PLS Research Workshops in Paris, 2009. He told us about his PLS tale: having started studying PLS regression, he was later referred to PLS Path Modeling, only to find out that, initially, nothing made sense to him. I was so relieved in knowing that I was not the only one with the same big confussion. Even a PLS *jedi master* like Michel had faced the same puzzle I had struggled with.

I became obsessed with the story behind PLS since the very first moment I started to work on my PhD dissertation about PLS Path Modeling. As a grad student in the fall of 2005, I remember googling about “Partial Least Squares” and getting links back that talked about PLS Regression. When I tried to be more specific and searched for “Partial Least Squares Path Modeling,” there was nothing but again PLS Regression retrieved results only. Why was there almost nothing about PLS Path Modeling? There were virtually no resources about the original works by Herman Wold, or the so-called Structural Equation Models via PLS approach. Moreover, most of the available resources on the Web were basically chemometrics-related material. It is amazing to see how much things have changed in just a decade. But it hasn’t been easy. It has required a titanic effort coordinating the work of a large number of scholars. Gradually, the barriers have been pushed in order to revive a methodological framework that was heading toward the annals of statistical methods that sadly end up in the dead archives of university libraries.

One of the very first things I knew I wanted to do was to find out more about the history of PLS methods. The main resource I had in my hands was a tutorial on PLS Path Modeling, which was basically a draft for the famous article *"PLS Path Modeling"* by Tenenhaus et al (2005). I started tracking the references listed in the tutorial, and I was happy to find that the Economics and Math libraries of the University of Barcelona had copies of the two volumes *Systems under indirect observation: Causality, structure, prediction*, the intimidating treatise *Latent Variable Path Modeling with Partial Least Squares* (Lohmöller 1989), the *Encyclopedia of Statistical Sciences* with the entry of PLS (Wold 1985), and *The Making of Statisticians* containing the autobiographical essay of Herman (H. Wold 1982b). Little by little I started to dig out and uncover fragments of a fascinating story.

All I wanted was to be able to answer the *who*, *when*, *how*, and *why*. Who created PLS? When was it developed? How was the development process? And why? How does something such as PLS is developed? What were the circumstances that gave birth to that framework? In this sense, the quest became a quest not only about PLS itself but also about its creator, about his mind, his ideas, and surrounding context. Among all my inquires, one question has remained for years in the back burner of my head: Why was PLS Path Modeling not that well known? Think about it. By 2005, the first works about PLS had already been present for almost 30 years! Yet, having access to those references was—and still is—ridiculously difficult. I think I have now most of the elements to answer the questions that have been lingering since that fall of 2005.

Chapter 3

A little bit about Partial Least Squares

One of the ever present tasks of almost every data analysis endeavor is the study of relationships between variables. In its simplest form, the analyst wants to know whether two values are related with one another. For instance, I might be interested in studying if there is a relation between the number of calories in the coffees my friends and I drink, and the price we pay for them. The question about the relation between calories and price can be further broken down in other inquiries: What is the nature of the potential relationship? What is its magnitude? Or how important is it? I could also expand the analysis and ask about a possible dependency: Can I predict the prices of coffees based on the number of calories they contain? Or viceversa: Knowing the price of a coffee, what can I say about its number of calories? To answer these questions we can turn to the rich set of analytical tools that statistical methods give us by using a number of graphical displays and charts, measures of simultaneous variation such as covariance, coefficients of association such as correlation, and modeling options such as regression analysis.

Often, we don't have just two variables but many more. If we were hired to do a deeper and broader analysis on data from coffee shops, we surely would have more available information like the area size of the establishments, or records about frequent customers. Hence we could ask about the relationship between the number of square meters and the cost of the coffees. Furthermore, we can think about the differet aspects that go into the preparation of a cup of coffee—the beans, water, milk, sugar, etc—and how they may be associated with the financial performance of the businesses (earnings, losses, costs, etc). These questions and many others arise naturally when we do data analysis, and we aim to tackle them at various levels of statistical complexity.

Depending on the circumstances, we may face situations in which the analyzed data are divided—or can be divided—into sets of variables. With the hypothetical coffee shops data, one set of variables may be comprised by quantity of ingredients (beans, water, milk, sugar), a second set may be formed by utilities consumption (gas, electricity, water, internet, etc), and a third set may be composed of sales by size of container (sold units by size: small, medium, large). Sometimes we just care about describing and summarizing the relationships. This is typically done in the initial phase of the analysis when we are in exploratory mode. Sometimes we may be interested in relationships of prediction and explanatory nature. Based on the analysts's preconceived ideas or hypotheses about the data, more advanced and sophisticated tools can de devised. In all these cases we can rely on a wide range of multivariate statistical methods that allow us to generalize the analysis of two variables into two or more sets of variables. Among such methods, Partial Least Squares techniques stand out as one remarkable toolbox that, simply put, provides a versatile data modeling platform for analyzing relations among one or more sets of variables.

PLS

Rooted in a couple of data model-building and computational ideas developed by Herman Wold in the 1960s, **Partial Least Squares** methods have traveled a long journey until reaching the high praise and recognition they receive today. Experiencing gradual changes, and bifurcating into several branches, PLS methods have gone through a fascinating—although not always smooth—evolutionary process. Nevertheless, they all have proven to be tremendously valuable on both theoretical and practical aspects.

At its heart, Partial Least Squares provide a versatile platform for analyzing multiple relationships among one or more sets of variables—measured on some objects. Among their attractiveness we can highlight several features:

- their deceptively simple iterative mechanisms
- their ease of programming implementation
- their estimation ideas anchored on the principle of least squares
- their marked geometrical flavor of projection-based methods
- their inherent dimension reduction nature
- their rich possibilities for data visualization
- their strong inductive spirit
- their fit within statistical learning approaches

Above everything else, the most captivating trait—in my opinion—is that Partial Least Squares offers us a wide analytical platform that covers a large number of multivariate data analysis techniques. Plus, it gives us the ability to connect a number of seemingly unrelated methods. Such a platform, encompassing both its computational side, and its flexible analytical essence, is of a great richness and applicability very few times encountered in other data analysis approaches.

A broad definition of PLS

If what you are looking for is an official definition of PLS I'm afraid you won't find such thing. Instead, we could give different meanings to the PLS term. From a narrow point of view, we could talk about PLS as an algorithmic template for computing several multivariate models. From a more applied angle we could also talk about PLS as a family of regression-type data analysis methods. From a richer and more comprehensive standpoint, we can regard PLS as an umbrella term for dealing with multirrelational systems of one or more sets of variables following the principle of partial least squares. It is the latter description that I like to consider for defining PLS as a general framework. Here's my definition. Broadly speaking:

> Partial Least Squares is a versatile multivariate data modeling framework for analyzing multiple relationships among one or more sets of variables measured on some objects.

One thing for sure is that there's a large number of methods labeled with the PLS acronym. Such an amazing variety of methods and approaches, so diverse and plural, makes them seem practically impossible to view under a united scope. In my opinion, many headaches would be avoided if there was a unique reason for a method to be labeled with the PLS acronym. For better or worse, there is not just one but several sources that make a method be sheltered under the PLS brand. Methods carrying PLS as prefix or suffix are so varied and so imbued in different fields and areas of application that it seems hard to picture them all under the same lens.

The literature about PLS methods is incredibly rich. If you take a look at the dozens of references about Partial Least Squares, you will be shocked with a mind blowing puzzle. The major benefit of such abundance of material is the richness of concepts, principles, applications, extensions, and new proposals. The ugly side is that there's no universal view, no unified notation, no unique terminology, and even worse, slightly different connotations of the term PLS. Which only makes it easy to get lost along the way. After all, this reflects the evolution of PLS methods, how they have spread out, how they have permeated into different subfields, and how they have come to occupy multiple places within the collective consciousness of Multivariate Data Analysis. Everywhere a PLS method has taken up residence, it dresses in different clothes, does its hair differently, and speaks a slighty different dialect.

In summary, as soon as you take a first step into PLS territory, you can immediately get disoriented in the middle of the trees without being able to appreciate the forest in all of its magnitude. No wonder why many practitioners are confused about exactly what PLS is, what it does, how it works, and how they can benefit from it.

Some words of caution

I would be a fool to tell you that there is one single universal perception of Partial Least Squares. Depending on who you talk to, you can get different meanings for the term PLS. If you ask your favorite search engine about "PLS," you might get answers like "please," or "political science," which is obviously not what we are talking about. If you search for the term "partial least squares," I bet you will find most results related to what is known as *PLS Regression*—the most popular version of PLS methods. Not surprisingly, most people think of PLS as just another technique for solving regression-like problems. This is not a mere coincidence but the result of the intricated history behind PLS, splattered with random twists of fate. Now, while it is true that the regression version may be the most common presentation people know, it is by no means the only one. For other people, PLS is an approach for estimating what are known as Structural Equation Models with latent variables.

Path Modelism and Regressionism

Today we can distinguish two main branches of Partial Least Squares approaches: 1) the **Path Modeling** branch, and 2) the **Regression** branch. The former introduced by Herman Wold, and the latter headed by Svante Wold. The recognition of these two large categories has to do with the way in which they have subsequently unfolded. By taking different directions, the branches have produced two major *movements* that for the most part, have grown apart, remained disconnected, and even unaware of one another in some areas of application.

PLS: Noun or Adjective?

The study of Partial Least Squares (PLS) methods begins with potential confusion—the term “PLS” has become broader and looser over the years. One problematic area has to do with the way authors use the term PLS. Sometimes PLS is used as a noun, while others is used as an adjective. As a noun, “PLS” is not one method but a set of methods. Talking about Partial Least Squares as a single methodology is like talking about Athletism as a single discipline, or like talking about the Himalayas as a single mountain. Hence, we should always let users/readers know that PLS implies a big family of methods.

As an adjective, “PLS” describes the *flavor* of a given method. Generally, authors have baptized PLS algorithms after the methodology they sprout from. In this sense, the acronym PLS is used as a label to indicate the estimation approach of a given technique. For instance, PLS as used in “PLS Regression” works as an adjective, describing the estimation procedure for a regression model.

Algorithmic approaches

Another problem is when we hear or read about "the PLS algorithm." The truth is that PLS methods are a bit of an anomaly in that they are both **methodologies** (with specifications, assumptions, and motivational reasons) and **algorithms** (with computational and operational aspects). This means that all PLS methods have their associated algorithm. Often, many different PLS methods are presented by their authors in such a way that the unaware reader is mislead to believe that a particular method is *the* PLS method. This is probably one of the main sources of confusion around the PLS literature. There is not such thing as *the PLS algorithm*, instead there are the PLS algorithms (in plural). For example, there is the PLS regression algorithm, the PLS path modeling algorithm, the PLS algorithm for Principal Components Analysis (a.k.a. NIPALS-PCA), or the PLS algorithm for Canonical Correlation Analysis (a.k.a. NIPALS-CCA), to mention but a few.

If there is anything in common among Partial Least Squares methods is that they all have an associated algorithm with a fairly uniform format. PLS methods proceed in a less intuitive way compared to classical statistical procedures where an optimization criteria is solved algebraically. PLS approaches are not formulated in terms of a global criterion to be optmized. That is, PLS procedures are stated without any maximization or minimization criterion. Just the data decompositions and the system of regression equations are declared. Usually, we express a model (in PLS mindset) in such a way that we identify the components, and the equations between components. Instead of looking to derive an analytical solution, we walk through the solution via a series of repetitive sequential steps until reaching a good stable approximation. In other words, Partial Least Squares are pure algorithmic approaches that consists of a series of steps approximating a stable solution. In many cases, the PLS algorithms coincide with standard algebraic solutions—usually involving an eigendecomposition.

Fragmentation

Another problematic issue is that PLS methods are usually presented in a fragmented way. Most book chapters, papers, courses, workshops, and even conferences deal with only a subset of the dozens of methods, rarely covering a panoramic view. The main reasons for this compartmentalization has to do with historical issues, and the way PLS methods have evolved. Furthermore, accessability to the original references is largely difficult to most practitioners and users.

The initial versions of PLS methods are documented in the 1960s works of Herman Wold, and there is no controversy about the PLS origins. Throughout the entire body of references and publications over several decades about PLS methods in general, all authors acknowledge the roots of the so-called NIPALS procedures Wold (1966a, 1966b) as well as (Wold and Lyttkens 1969). However, the place PLS methods occupy in contemporary multivariate statitics is a bit more complicated. PLS's rising global popularity has led to disagreements about what constitutes an authentic PLS algorithm. The same name is used to refer to different things, while at the same time different terms are used for naming the same concept. Even within PLS community, ideas about what PLS is varies and has caused debate and tension every once in a while.

The two Wolds

Another source of confusion is the fact that two of the main leading authors behind PLS methods share the same last name "Wold". The unaware reader may think that Wold is one single author without realizing that it can be either Herman (the father) or Svante (the son). In fact, the way and style in which the two Wolds presented their works have left a profound footprint in their ulterior developments. The framework of Herman, emerged from econometric's systems of equations with latent variables, seems to have insurmountable disparities with Svante's Regression Models stemmed from chemometrics. Both branches, with its various subdivisions, show contrasting differences at various levels, notably at the area of application, but also at a philosophical, ideological, conceptual, language, technical, spread and diffusion levels. Albeit their common mathematical and operational elements, and even their shared genes, the contrasting physical appearance between both frameworks puts an illusory divide between them that can easily mislead all inexperienced users, and even some well versed PLS *connoiseurs*.

Pedagogy of PLS

Another hurdle for studying PLS methods has to do with the fact that they have evolved on top of existing techniques, which produces some pedagogical side-effects. Why? Because if you want to learn about a specific PLS technique, usually you must go through a double learning process. Consider for instance Principal Components Analysis (PCA), and its corresponding PLS version NIPALS-PCA. If you want to study NIPALS-PCA, you would begin studying PCA, and then switch to the NIPALS-PCA algorithm. In general, you must first learn about the general problem—and its standard solution—, be it regression, discrimination, principal components, or canonical correlation, just to mention some of them. And then you have to learn how to look at a given technique from the PLS angle. Consequently, this adds an extra layer of concepts, terms, and jargon that you have to deal with: one from the standard approach, and the other one from the PLS approach. No wonder why PLS methods, despite their practical usefulness and methodological attractiveness, have no reserved seats in undergraduate syllabus, and have limited attention in postgraduate programs.

Current Status

Today Partial Least Squares methods are well recognized in pretty much all fields of knowledge where they have been applied. There's a tradition of biannual symposiums exclusively dedicated to discuss advancements and state-of-the-art around PLS and related methods. We also have special publications such as the *Handbook of Partial Least Squares* (Esposito et al, 2010), *New Perspectives in Partial Least Squares and Related Methods* (Abdi et al, 2013). On the computational side a wide range of software programs are available both commercially and free. The largest group of tools is the dozen of R packages freely available in CRAN; there are also XLSTAT plug-ins (by Addinsoft) for MS Excel; SmartPLS (Ringle et al, 2005), ADANCO, SIMCA (by Umetrics), procedures for SAS, and libraries for Matlab, and Python. Also important is the amazing amount of online learning resources with tutorials, slides, videos, seminars, webinars, workshops, and specialized courses. All this ecosystem reflects the success of PLS methods in industry, academia, research, and governmental spheres.

But it hasn't always been like this.

Chapter 4

Commented Timeline

Before presenting the historical narrative in the following chapters, I would like to provide a brief timeline featuring events that have guided the path traveled by PLS methods until our current days.

Mid-1960s: NIPALS Procedures

The origins of all PLS methods can be traced back to the mid 1960s where the precursors of modern-day PLS tools were developed by Herman Wold and his research group at the Uppsala University, Sweden. From his work on econometric models of simultaneous equations, Herman Wold modified the algorithm of his Fix-Point method in order to solve a series of ad-hoc data analysis problems. Historically, the first type of PLS algorithm is a power method (H. Wold 1966a) for computing Principal Components. This procedure was almost immediately extended to a series of ad-hoc procedures among which there was a version to calculate Canonical Correlations. Under the name **NILES**, short for "Nonlinear Iterative Least Squares," Herman (H. Wold 1966b) presented a collage of examples solved by means of iterative procedures based on steps of least squares regressions.

Interestingly, these initial works contained the fundamental mathematical elements of all subsequent PLS methods: computation of *data components* as weighted sums of variables, operationally obtained through steps of least squares regressions. Not long after their presentation, (Wold and Lyttkens 1969) replaced the term "NILES" by "NIPALS" (Nonlinear Itearive PArtial

Least Squares), consequently shifting from *NILES procedures* to *NIPALS procedures*. Because these first publications around NILES, subsequently NIPALS, emphasized the computation of Principal Components Analysis, today most authors refer to NIPALS as the PLS algorithm for PCA. To avoid confusion, I prefer to use the term NIPALS-PCA for this special case.

It is worth noting that the first NIPALS procedures **were never a single methodology nor a single approach**. They were rather a collection of more or less disconnected and different algorithms for solving a diversity of methods such as PCA, CCA, regressions, and systems of econometric equations. The emphasis of these procedures was put on the idea of linearizing problems that were inherently non-linear (or bilinear) in their parameters and unknowns. In addition, there was no use of path diagrams.

Early 1970s: NIPALS Modeling

In the early 1970s, the so-called NIPALS procedures experienced a wave of significant modifications. The most impacting factor was the inclusion of (at the time) recent simple path models with latent variables (i.e. models associating two blocks of variables). Such models were the result of the breakthrough achieved by Karl Jöreskog, one of Wold's former PhD students. One of Jöreskog's major accomplishments was the synthesis from merging econometric simultaneous equations models, psychometric latent variable models, sociology causal analysis, and biometric path analysis, in an estimable way via computer algorithms and Maximum-Likelihood approach. Attracted by the new model-building opportunities offered by the combo of path models and latent variables, Wold realized that some of the NIPALS procedures he had already proposed could be adapted for this new type of models. In 1973 Wold changed again the label of the methods: from *NIPALS procedures* to *NIPALS modeling*, clearly reflecting a more mature—but still incomplete—modeling framework (Wold 1973). In retrospect, this is the period of time where we can truly talk about the origins of the PLS Path Modeling branch.

Keep in mind that the term NIPALS, as used by Herman Wold, was a very broad label. It was so broad that renowned American mathematician Joseph Kruskal once asked Herman (Wold 1973) "whether an explicit definition can be given for the class of nonlinear models that constitute the scope of NIPALS modeling." Wold answered:

> NIPALS modelling is highly flexible, allowing the combined used of several devices, including parameter grouping and relaxation; auxiliary transformation of the model; and modeling the predictors in terms of indirectly observed manifest variables and other hypothetical constructs. Hence I see NIPALS as an open ended array of models with unlimited complexity in the combined use of several devices.

Mid 1970s: NIPALS Soft Modeling

The mid-1970s is a time of extensive work and refinement. Seeing and seizing the opportunity to develop his own modeling approach, Wold's team, this time at the University of Gothenburg, refined and polished different versions of a general methodology for estimating path models with observed and unobserved variables. Throughout the 1970s, Herman Wold led the development of an alternative methodology to estimate path models by applying iterative algorithms of least squares regressions. In first place there is the extension of the algorithms from handling two blocks (2 latent variables) to handling three blocks (3 latent variables); secondly there is the extension of handling one between-block relation to more than one between-block relation (Wold1974, Wold1975a, Wold1975b).

It is also during this second half of the 1970s when Herman Wold introduces the fanciful notion of "Soft Modeling," which he will adopt later as the insignia to wrap his model-building framework via the PLS approach. To the best of my knowledge, the first time that Wold publishes a paper mentioning the term "soft modeling" is in *"Soft Modelling by Latent Variables: The Non-Linear Iterative Partial Least Squares (NIPALS) Approach"* (H. Wold 1975a).

> The NIPALS approach is applied to the 'soft' type of model that has come to the fore in sociology and other social sciences in the last five or ten years, namely path models that involve latent variables which serve as proxies for blocks of indirectly observed variables.

The "soft modeling" notion is more of ideological nature and will also evolve in the following years. Its underlying meaning implies the idea of modeling in "complex situations where data and prior information are relatively scarce and without requiring assumptions about the stochastic-distributional properties of variables and residuals" (H. Wold 1975a).

In *"Path Models with Latent Variables: The NIPALS Approach"* (H. Wold 1975b), Herman writes on page 352:

> It sometimes happens that the model builder has little or no more prior information at disposal for the model construction than its intended operative use. The NIPALS models are designed with particular view to applications in such low-information situations.

Late 1970s: PLS Soft Modeling

Toward the ending of the 1970s, after a long simmering and cooking, Wold and his team arrive to a more defined framework. Finally, the acronym "NIPALS" is shortened to "PLS", and the more or less uncoordinated types of models are reduced and emphasized to handle path models with latent variables indirectly observed. In other words, the other types of models, of a more econometric flavor (GEID, REID, etc) that previously overlapped NIPALS modeling become secondary topics within the PLS framework.

1977-78 see the appearance of a number of articles:

- *"On the Transition from Pattern Cognition to Model Building"* (H. Wold 1977a)
- *"Open Path Models with Latent Variables"* (H. Wold 1977b)
- *"Ways and Means of Multidisciplinary Studies"* (Wold 1978)

Early 1980s: The Basic Design and Extensions

The end of the 1970s decade sees the official presentation of the so-called **Basic Design** for PLS path modeling. This is the what can be considered to be the *stable* version, or paraphrasing in marketing terms, Wold's *minimal viable product.* The first publication with all the elements of the basic design is *"Causal-Predictive Analysis of Problems with High Complexity and Low Information: Recent Developments of Soft Modeling"* (Wold 1979). This is the first time where Wold uses the definitive terminology that will prevail in his subsequent manuscripts. More specifically, it is the first time that the terms "Mode A" and "Mode B"—for the two main types to estimate the so-called weight relations—appear on print. The so-called Basic Design, is the basic method for PLS Path Analysis with Latent Variables, and it is the one on top of which all extensions and modifications are based on. Further discussions are provided in (H. Wold 1980a), (H. Wold 1982b), and (Wold 1985).

Also in 1979 the LISREL-PLS meeting jointly organized by Karl Jöreskog and Herman Wold takes place in Cartigny, Switzerland. The proceedings of this meeting will be published in 1982 in the form of the classic two-volume book: *Systems under indirect observation: Causality, structure, prediction.*

Under the term *PLS Soft Modeling*, Herman Wold proposed a generic modeling framework, wrapped around his unique philosophical perspective, with the goal of solving path models with latent variables. Among the vast array of references, the classic works are those of Wold (H. Wold 1980a), (H. Wold 1980b), (H. Wold 1982a), (Wold 1985).

From (Wold 1985):

> Soft Modeling is the name for the methodology for PLS estimation of path models with latent variables indirectly observed by multiple indicators.

The methodology involving the basic design is fully described in (H. Wold 1982b). Geometric interpretations are provided by Fred Bookstein (Bookstein 1980), (Bookstein1982a, Bookstein1982b). What is perhaps the first pseudo-code description of the basic algorithm is provided by Jan-Bernd Lohmöller (1989, p. 29).

Late 1980s

In 1989, the book *Latent Variable Path Modeling with Partial Least Squares* by Jan-Bernd Lohmöller is published. From the list of researchers and contributors of PLS Path Modeling, Lohmöller is the figure that compiles various extensions of the basic design. Although his writing style is not friendly at all, his contributions touch several directions: multiblock data, multiway data. More importantly, he also develops the software program LVPLS with its first versions dating from 1982 till the well-recognized LVPLS ver-1.6 from 1987.

In the last of his works, the book *Theoretical Empiricism: A Rationale for Scientific Model-Building*, edited by Herman (H. Wold 1989b), perfeclty reflects and summarizes his standpoint about model-building and Soft Modeling via PLS approaches. The introductory chapter "Introduction to the Second Generation of Multivariate Analysis," (H. Wold 1989a) is a compendium that shows both sides of Wold's interests: path models with manifest and/or latent variables; an also theoretical empiricism—his ideological posture about phylosophy of science.

1990s

In the 1990s, there's basically no theoretical progress. The most interesting work is on the computational side with the development of PLS-Graph by Wynne Chin.

2000s

The beginning of the XXI century saw a renaissance of interest in PLS-PM. Major contributions were made by French researchers. Among the primary landmarks we have the paper *PLS Path Modeling* by Tenenhaus et al (2005), in the journal *Computational Statistics & Data Analysis* (CSDA). At the time of this writing, this paper accumulates 1921—the most cited paper of CSDA!.

On the software side, 2005 saw the launching and presentation of the software "SmartPLS 2" by Ringle, Wende, and Will (2005). Their work has been an on-going process with a series of versions (the current one being SmartPLS 3). Currently, SmartPLS has a bit more than 2000 citations, and growing.

On the theoretical side, the works of Hanafi (2007) and Tenenhaus (2010), are fundamental. Their work has brought a better understanding of the algorithm. They've continued the multiblock extensions initiated by Lohmöller, and Hanafi has even resolved some of the issues around the convergence of the PLS-PM algorithm.

More recently, Tenenhaus has proposed new modifications that allows us to tweak the PLS-PM algorithm in such a way that we guarantee its convergence, be sure of how the algorithm works, and includes PLS-Regression as one of its special cases. Extended on the same direction, Arthur and Michel Tenenhaus (2011) have proposed their *Regularized Generalized Canonical Correlation Analysis* (RGCCA).

In summary

Herman Wold led his team, constantly carving and painstakingly polishing his procedures over a long period of time; taking shape and maturing, until he got a ripe version dubbed "basic design", wrapped around the notion of "soft modeling." Wold envisioned a modeling approach for analyzing systems of linear relations with observed and unobserved variables. Given the fact that PLS Path Modeling was inspired by Karl Jöreskog's approach, it is undeniable that both methodologies shared many things in common. Notably, the concept of latent variables, the systems of linear relations, and the graphical representation of the models. However, Wold conceived PLS with significant and unique differences to LISREL. There is not only a difference in estimation procedures, but there is also a major ideological difference, heavily influenced by Wold's posture regarding econometric simultaneous equation models. The overall "PLS Soft Modeling" idea represents Herman's culminating work in all of its conceptual, intellectual, philosophical, scandinavian glory.

PLS Regression Origins

The beginnings of the PLS Regression branch date back to the early 1980s when some of the PLS principles were modified and applied to tackle regression problems of chemical data.

Under his father's suggestions, Svante Wold attempted to apply the PLS Path Modeling methodology for regression analysis with chemical data. After some modifications by Svante Wold and Harald Martens, a new algorithm is born. This is the beginning of the PLS Regression framework. The seminal work *"The multivariate calibration method in chemistry by the PLS method"*, corresponds to Svante Wold, Harald Martens, and Herman Wold (Wold, Martens, and Wold 1983).

Like PLS Path Modeling, the roots of PLS Regression have the signature of Herman Wold with the use of iterative least squares algorithms. Unlike PLS-PM, it was not Herman who took charge of the bifurcated ideas but his son, Svante Wold, who almost immediately became the flag bearer behind the PLS Regression sprout. Taking advantage of the core computational ideas of his father's methodology, and stripping them away from the more philosophical-epistemological elements, Svante and his colleagues subsequently launched a series of algorithms with an emphasis in multivariate regression problems, decisively more pragmatic, and without the econometric-psychometric speech of Herman's original methods.

Chapter 5

The Econometrician

The father of Partial Least Squares was the great Swedish statistician and econometrician Herman Wold. In fact, his full name was Herman Ole Andreas Wold, and he was not born in Sweden but in the neighbor country of Norway. He was born on December 25th, 1908, as the sixth child of Edvard and Betsy Wold, in the small town of Skien, the administrative center of Telemark county, 133 kilometers (82 miles) south of Oslo. He spent there his first three years of life until 1912 when his parents decided to move to Sweden. Because of harsh economic times in Norway, the Wold family, except the two oldest children, relocated to Lidköping, a small town southwestern Sweden, near Stockholm. Here Edvard Wold—a skilled furrier—made a living designing and making leather coats lined with fur, an essential piece of apparel for the cold Scandinavian weather.

Herman Wold grew up and went to elementary school in Lidköping. However, since there was no high school in town at that time, he had to attend high school in Skara, one of the oldest cities in Sweden, with a long educational and ecclesiastical history—40 kilometers (25 miles) away from Lidköping. At young age, Herman showed a good talent for mathematics, and after high school he enrolled at the University of Stockholm in 1927 where he studied physics, mathematics, and economics. There he met Harald Cramér, the renowned Swedish professor of Mathematics and Statistics who made Herman change plans. "I was greatly impressed by him", Herman said, "and interested in his work." Since Wold was very interested just in statistics he decided to stay under the tutelage of Cramér learning about elements of probability, statistics, and risk theory. In 1930 Herman finished his degree and he took his first job in the insurance industry where he started to do actuarial work.

Herman's interest in statistics was greater and deeper than his intention to stay in the actuarial field, and so he decided to go back to academia and get a PhD degree. Once again under the mentoring of Harald Cramér, Herman took courses on stochastic processes and time series. Moreover, he soon got caught up in the excitement surrounding the emergence of probability theory recently introduced by the famous Russian mathematician Andrey Kolmogorov.

Time Series Studies (1932-1938)

In 1938 Wold received his doctorate with the thesis *A Study in the Analysis of Stationary Time Series*. For his dissertation, Wold did his research on stationary stochastic processes, studying the one-step prediction of a time series, and proposing his *Decomposition Theorem* —one of the most famous results by Wold, and one of the essential elements in the foundation of time-series analysis and forecasting. From the historical point of view, it is worth mentioning the use that Wold made of the least squares principle for his doctoral work. Herman studied the one-step prediction of a time series using the principle of least squares. Basically, Wold proved that any stationary time series can be partitioned into a deterministic component precisely predictable from its past, plus a random component which can be modeled as a weighted sum of "innovations." In simple terms, given a series of values at different times

$$x_t$$

,

$$x$$

,

$$x_{t-1}$$

,

$$x_{t-2}$$

,

$$\dots$$

,

$$x_{t-r}$$

, the decomposition theorem is used to express

$$x_t$$

in terms of the preceding values as a weighted sum:

$$x_t = a_1 x_{t-1} + \cdots + a_r x_{t-r}$$

in which the coefficients

$$a_1, \ldots, a_r$$

are obtained by least squares regression of

$$x_t$$

onto

$$x_{t-1}, \ldots, x_{t-r}$$

. Accordingly, Wold's decomposition showed that the three classic time-series models—the model of hidden periodicities, the moving-average model, and the autoregressive model—could be seen as different cases of the same general model. "The role of least squares," Herman wrote, "was important too." To the naked eye this might not seem to be very relevant but the truth is that Wold had already started to embrace least squares as one of his favorite analytical tools. From that moment on, the principle of Least Squares would occupy a central place in Wold's mind and, without exaggeration, even a sacred place in his heart. Whatever model-building endeavors that he would later faced, he would always try to find a way for using Least Squares to estimate the parameters of a model, no matter how simple or complex the models and equations were.

Consumer Demand Analysis

After his doctoral studies, Herman remained in the University of Stockholm as a lecturer on actuarial mathematics and mathematical statistics. He married Anna-Lisa Arrhenius in 1940, and they had three children: Svante, Maria and Agnes. He was very proud to be the son in law of Svante Arrhenius—the famed Swedish scientist founder of electrochemistry, and Nobel Prize for Chemistry in 1903. In 1942 Wold accepted the Chair of Statistics at the prestigious University of Uppsala, the oldest university in Sweden, founded in 1477. At the time of Wold's arrival to Uppsala, however, the Statistics Institute was a small one, formed by one professor, one half-time assistant, and one half-time secretary. This would change in 1945 right at the end of World War II when the government decided to invest and expand Sweden's Universities.

With a fresh tenure position, Wold started his own research on Demand Analysis and econometrics modeling. As a matter of fact, he had already started to work on demand analysis the summer before defending his dissertation, appointed by the Swedish government to perform such studies on the national economy. More specifically, Wold carried out the study of consumer demand analysis from 1938 to 1940. The main line of approach was to combine the analysis of family budget data and market statistics (time-series data) so as to obtain a unified picture of the demand structure in Sweden between 1920 and 1938. The airs of war were starting to fill Europe's atmosphere and it was clear that if a conflict broke out, government rationing policies for food and goods would need to be implemented. As Europe entered the War period, Wold's work intensified, measuring price and income elasticities of demand—how sensitive the consumer demand was to changes in prices and income.

Although the government commission only lasted two years, Wold spent about 14 years doing research and time series analysis of the collected data, as well as publishing several articles between 1938 and 1947. One of such publications appeared in 1940 in the dual form of a research report; written by Herman Wold and Lars Juréen that material would later be used for a specialized textbook on econometrics *Demand Analysis: A study in econometrics* (published in 1952)—which became a classic in the field. "The monograph is written," Wold said, "in the dual form of a research report and a specialized textbook on econometrics." Like in his doctoral research, Herman Wold made extensive use of Least Squares for estimating the parameters in his models and making accurate forecasts.

The Least Squares Affair

Directly related with his work on Time Series and Demand Analysis, Wold got involved in a peculiar confrontation that happened within econometrics during the 1940s: Ordinary Least Squares against Maximum Likelihood. This period is crucial for the development of PLS because this is the time when Wold, somewhat stubbornly, embraced the Least Squares principle against most other methods, especially against the overuse of Maximum Likelihood. To understand why, we need to talk a bit more about econometrics and demand analysis.

During the first half of the twentieth century, one of the greatest challenges in econometrics was the estimation of demand analysis equations, and simultaneous equation systems. Before the 1940s, the main analytical tool used for this task was ordinary least squares (OLS). Although it was not the perfect tool, OLS was able to get the job done in most occasions. However, as models started to become more sophisticated, there were more and more cases where OLS simply didn't seem to work. Consequently, the method of least-squares started to being criticized and some econometricians began to stress out that the device was not foolproof if applied uncritically.

Partly due to lack of understanding, partly due to conceptual confussions, econometric theorists were burdened with a myriad of problems that took many years to be solved or to be aware of. We're talking about things like *identification*, *measurement errors*, *multicollinearity*, *model choice*, and so forth, that are known to affect not only OLS but many other estimation methods when applied to economic time-series and simultaneous equations. These issues are now taught in most introductory and intermediate econometrics courses, but back in Wold's days all these problems were the cause of so many headaches for econometricians trying to figure out how the economic systems work.

Around the early 1940s, the challenges posed by econometric models attracted many mathematicians, statisticians, and economists. This "gold rush in macroeconomic model building," as Wold used to call it, captivated the minds of very talented economists. Among the most *avant-garde* groups was the Oslo school of thought, whose most distinguished protagonist was the Norwegian Trygve Magnus Haavelmo (Economics Nobel in 1989). Haavelmo's most proclaimed accomplishment was the *Probability Revolution in Econometrics* initiated in 1943 with the premise of adoptating a probability approach in Econometrics. Among other things, he was also the one who highlighted the problems of Least Squares when applied for estimating simultaneous equation systems.

Trygve wrote two influential articles: the *The statistical implications of a system of simultaneous equations* in 1943, and *The Probability Approach to Econometrics* published in 1944 as a supplement in the journal *Econometrica*. Both works set a before-and-after in Econometrics. On the one hand, Haavelmo introduced modern statistical inference based on probability models to economics. Although he was not the first one to introduce elements of probability theory into economics models, he was the first one to import the statistical inference approach for testing hypotheses. On the other hand, Haavelmo proposed the idea of using several structural equations *simultaneously* to construct econometric models following a probability approach.

Broadly speaking, one of the main questions among economic researchers had to do with: How to mathematically model the economy's behavior? For instance, how to model a system of equations for demand and supply? In a very simple model the *Demand* of a good, say coffee, depended on the price. In turn, the *Supply* of coffee depended on how much demand was for coffee, as well as the cost of production. While there was no doubt about the general theory of demand-and-supply, there was a heated debate on the mathematical and statistical form such theory should took.

The Oslo school advocated for a system in which demand and supply affected each other simultaneously—at the same time:

$$D_t = a_1 S_t + \epsilon_t \quad \text{demand} S_t = b_1 D_t + b_2 P_t + \delta_t \quad \text{supply}$$

The Demand equation reflects the behavior of the *consumers* of coffee: they respond to the price of coffee—reflected in

$$a_1 S_t$$

. The Supply equation reflects the behavior of the *producers* of coffee: they set the price

$$P_t$$

depending on how much demand

$$b_1 D_t$$

there is for coffee, as well as the cost of production. In other words, the demand and supply for coffee are simultaneously determined. Simultaneous equation systems can be much more complex and sophisticated, but one of the basic ideas, as its name indicates, is that of *simultaneity*. More important, Wold disagreed with this model of demand and supply in which both relationships were determined within the same time period. He found it difficult to believe that the economic system was determined simultaneously. For him, causal forces only worked in one direction at any one time and these forces should be reflected in the models.

The Stockholm school, strongly based on the works of Dutch economist Jan Tinbergen—first Nobel Prize in Economics, 1969—thought of an economic system not in terms of simultaneity, but in terms of time-lagged periods. Tinbergen called these types of model *causal chain systems* and he used to illustrate them with *arrow schemes*, very similar to the path diagrams that would later be employed within structural equation models as wel as within path models.

Herman Wold had been formed under the Stockholm school tradition, in which the econometric models of causal chains included an *ex ante–ex post* component. Basically, these models did not include the idea of simultaneity, but rather a time-lag component in which one thing had to occur first, and then another thing would followed. Taking the demand-and-supply example of Coffee, the associated equations would be expressed as:

$$D_t = a_1 S_{t-1} + \epsilon_t \quad \text{demand} S_t = b_1 D_t + b_2 P_t + \delta_t \quad \text{supply}$$

Wold put emphasis on knowing the interrelations of all the variables in terms of the time lapse between causes and effects. In this way, one could reduce the system of dynamic causal relations to one final form equation. The condition to achieve such reduced form involved no simultaneously determined equations like Haavelmo suggested.

Before Haavelmo's contributions, the errors

$$\epsilon_t$$

and

$$\delta_t$$

in the models were thought to be due to measurement discrepancies. What Trygve proposed instead was to treat the errors in the statistical sense of random noise, and assign them a certain distribution. In this way, the Maximum Likelihood device had the door open, allowing researcher to test and discard hypothesis. His advice was to avoid using Ordinary Least Squares (OLS), since this method provided inconsistent results when applied to simultaneous equation models. Instead, he favored and promoted the use of Maximum Likelihood.

Wold's effort to rescue OLS

Haavelmo's advice against OLS was a shocking statement for Wold. "I felt so disturbed," Herman wrote, "by Haavelmo's wholesale dismissal of it." Considering that all of Wold's work had been based on OLS, and that he had achieved very good results with it, he was very surprised by the bad press least squares received. Herman was anything but "spurred" by Haavelmo's "rejection of OLS regression." Could it be possible that Haavelmo was right and that OLS were to be banned? If that were the case, all results previously obtained by Herman were useless, something that didn't seem to match at all with his analysis and practical evidence.

In the middle of the 1940s, together with Ragnar Bentzel, Herman set out on a task to see whether Trygve Haavelmo was right or not. They intensely studied the simultaneous systems of equations and, to Wold's relieve, they found hope for Least Squares, publishing the proofs and conclusions in the 1946 article *"On statistical demand analysis from the viewpoint of simultaneous equations"*. Simultaneous equation models could be divided in two broad categories: *recursive* simultaneous equations and *non-recursive* simultaneous equations. Or as Wold preferred to call them: **causal chain** systems and **interdependent** systems, respectively. When a system could be expressed in recursive form, Wold and Bentzel showed that the method of Ordinary Least Squares could be perfectly used to estimate such models and give equivalent results to those under Maximum Likelihood. The challenge, however, was with the non-recursive models where OLS did not provide adequate results.

This issue would obsess Wold for the next decades, trying to find a way that would let him estimate non-recursive systems with Least Squares.

By the end of the 1940s, Herman Wold had already been working for over 17 or so years (1932 - 1949) on a number of topics including time series, insurance statistics, analysis of consumer demand, and econometric systems of equations. Moreover, he was basically the only Scandinavian econometrician outside of Oslo with international reputation. Niels Kaergard gives a revealing description of how Wold was perceived among the econometrics community (Kaergard 2012):

> "Where the Oslo school was a central part of the internatinal mainstream econometric tradition in the 1930s, 1940s, and 1950s, Wold was seen as a man with a rather special point of view."

In summary, the main question of debate became whether the structure of economic models was simultaneous or recursive? This was reflected in the theme *"Toward a verdict on macroeconomic simultaneous equations"*, title of a publication edited by Wold. The main stream in econometrics for estimating simultaneous equation systems was the one based on the method of Maximum Likelihood. Wold was not part of this group. On the contrary, Herman was a prominent figure in favor of recursivity and use of least squares, a stance that he would maintain for the rest of his life. However, it was Haavelmo's point of view that became the dominating one. The Maximum Likelihood convenience and elegance offered the mechanism of hypothesis testing, too tempting and irresistible to let it go by most researchers.

Wold always tried in one way or another to find a solution by least squares. Making distinction between simultaneous equations of recursive and non-recursive type, Bentzel and Wold showed that Haavelmo's wholesale dismissal of OLS estimation of the structural equations was applicable to non-recursive systems only, but not for recursive systems. For approximately the following 15 years, Wold went on a tortuouse mission to find an OLS solution for non-recursive systems.

Chapter 6

The Philosopher of Science

During the 1950s and 1960s Wold's research interests broadened from econometrics to other non-experimental analysis in general, and to the area of philosophy of science. Throughout the 1950s he dedicated a considerable amount of time to discuss notions of *model building*, and in particular, a lot of his work spinned around the prickly notion of *causality*. These topics (causality and model-building) were among Herman Wold's favorite subjects of study. Everything that had to do with theorizing about causation and models deserved his attention, especially as viewed from the more abstract and intellectual perspective of Philosophy of Science. In particular, Wold formed part of the main trend among econometricians that were keenly interested on making econometrics as scientific as possible. To achieve this, it was believed that researchers needed to be able to apply a scientific method of proposing hypotheses and test them.

In Wold's eyes, based on his extensive hands-on experience, research in economics as well as in other social sciences had "two heavy handicaps." One of the handicaps had to do with working "without the guidance and support of controlled experiment, the supreme tool of natural sciences," Wold believed. This was something not only attributable to economics but it was common to other social sciences. The second major handicap had to do with the type of data "notoriously unreliable" for the major part, and sometimes even "scarce or completely lacking." Suffering from those drawbacks "it is no wonder," Wold concluded, "that quantitative economic research displays little of the rigour and precision attained in many natural sciences."

If there was something that characterized the type of data Herman was working on was the fact of being *observational data* or nonexperimental data. His preferred tool of analysis was without a doubt the method of Least Squares, which he believed was better suited to deal with nonexperimental settings. At the time, the method of Maximum Likelihood (ML) was becoming extremely popular in most scientific fields. Introduced by Ronald Aymler Fisher in the early 1920s, ML provided an elegant statistical mechanism for theory testing. Emerged from agricultural and biological applications where experiments were the rule rather than the exception, ML proved to be highly valuable. Imported by Trygve Haavelmo to the Econometrics arena, ML had also been welcomed and given a primordial place of preference over other estimation methods such as Least Squares.

Maximum Likelihood, as attractive and convenient as it can be, requires the researcher to know the distribution of the analyzed data, and observations to be independent—assumptions that in practice may be impossible to have. For Herman Wold, the pay to price for using Maximum Likelihood on nonexperimental socio-economic data was unacceptable. It was not that he rejected the method of ML per se, but the assumptions on which ML was required to be applied made Wold had trouble accepting it. He argued that socioeconomic observational data didn't meet the ML assumptions. In economic analysis, it was not possible to make experiments and manipulate variables as in a laboratory or biological related setting.

Thinking about Causality

If it wasn't enough with the handicaps of lacking experiments and quality of data, a "more serious" issue, Wold observed, was the "research attitude" adapted in economics. "In economics," Wold wrote, "these checks on the scientific conscience are weaker." Consequently, "the average quality of the research is lower." All this situation was "paradoxical" to Herman eyes. Given all the "big arsenal of efficient methods that modern statistics has placed at the research work's disposal," how was it possible to have so many results of poor quality? His main conclusion was that the statistical methods—such as Maximum-Likelihood—had "been devised primarily for the treatmeant of experimental data," not for observational data as all of economics data was. Wold pointed out to analysis of variance and the maximum likelihood method, saying that they were “handicapped when applied to economic statistics or other non-experimental data.” The solution, instead, was in using more traditional methods “which have sometimes been declared obsolete,” Wold wrote, “like the least squares regression.” In Wold’s opinion, “these methods are essentially sound.”

Herman Wold was deeply concerned on identifying causality relations which could be used for predictions and forecasts. He didn’t like the idea of finding “laws” which he thought it was hard to do in economics given the fact that one could not do experiments. But he believed in the possibility to discern and determine which variables were the cause for a certain target or response variable. The problem was that, at the time, nobody was willing to talk about relations in causal terms, something that Wold found deplorable. For example, there was—according to Wold—“a deliberate disregard of causal interpretation,” in the way simultaneous equations were being used. Moreover, the terminology used in simultaneous equation models “avoids the concept of causality.” Such an evasion was pretty much inconceivable to Wold’s position, so much that it was for him “a break with scientific tradition by and large.” “The concept of causality”, Wold wrote, “is indispensable and fundamental to all sciences.” Terms that Wold considered “with a causal content” were influence, dependence, effect, stimulus-response, active substance. Terms like *functional relation* or a *predictability* hidden a causal meaning in Wold’s opinion.

> “if scientific analysis were stripped of all terms with a causal content, nothing would remain but description and formalism.”

“There was the question of causal interpretation of OLS regressions”, Wold recalled, although “causal arguments were still practically taboo in the socioneconomic and behavioral sciences.” At that time, the notion of causality was a hot topic in social sciences where causal analysis and causal models had attracted a lot of interest. Simply put, Herman Wold formed part of those thinkers that considered causality in the sense of discerning among a set of variables, which ones should be considered to be causal. To Wold, knowing which variable causes which within a system of equations was of paramount importance.

However, reading Wold’s papers about causality and model building, it is not clear what was the exact definition that he gave to the notion of *causality*. A good example about this issue is Wold's article *"Causality and Econometrics"*, appeared in *Econometrica* in 1954. This is one of the most enygmatic pieces of work, with a dense writing style, rhetorical questions, murky structure, and heavy philosophical content. Doubtless, the most philosophical of all his publications. Despite the title, the article is not much about econometrics but about philosophy of science. In it, Wold attempts to show his stance around the debate and controversy about causality in econometrics. The structure of the paper can be divided in three acts.

Act I is about the philosophical discussion of causality. Above all, it is a response to Bertrand Russell's seminal paper: *On the Notion of Cause*. It is also the opportunity for Wold to reveal his philosophical preference: "Logial Empiricism." For Wold, it is necessary (almost mandatory) to use the term "causality" related to terms in scientific models. For him, just talking about "functional relation" or "predictability" is not enough.

Even though Wold acknowledges that "the first requisite for a fruitful discussion about causality is an adequate definition of the concept," he does not provide such definition. "What then is to be understood by causality?" Herman asks. But he's not very precise in his answer. Instead, he talks about "causal relationship," viewed from both perspectives: controlled experiments—as in natural sciences—, and nonexperimental cases as in social science, and econometrics in specific.

In Act II Wold presents his famous and recurrent *duality* which he illustrates in the form of a contingency diagram. On one side there is the description and explanation. On the other side there is the non-experimental and the exeprimental aspects. Finally, Act III touches on causal concepts in econometric models. More specifically, he delves into his recursive models estimated by OLS, ending in causal chains. One of the few occasions he mentions the word "structural" appears in page 176. The structural term as used in structural models in the econometrics literature, is not the same that Wold uses. For him, "the causal interpretation of the relationships is entirely different."

"It is fundamental," Wold said, to separate "empirical observations" (or facts) from "speculative thoughts" (or theory). He recognized that the "weakest point in the philosophical discussion" was an adequate definition of the causality concept. For Wold most definitions were either "not defined at all" or simply "too narrow." "What then is to be understood by causality?" Wold asked. The answer he provided was that of "a causal relation in the case of a controlled experiment.” In the case of nonexperimental observations—as in economics—a relationship

$$y = f(x_1, \ldots, x_p)$$

is causal “if it is theoretically permissible to regard the variables as involved in a fictive controlled experiment with

$$x_1, \ldots, x_p$$

for cause variables and

$$y$$

for effect variable.”In the experimental case the typical situation is one in which the causal relation enters as a hypothesis to be tested or demonstrated." Wold stressed that “the test of the causal hypothesis is more direct and indisputable in the experimental case.”

Perhaps one of the best explanations about Wold’s stance, can be found in American sociologist Hubert Blalock’s 1991 paper *“Are There Really Any Constructive Alternatives to Causal Modeling?”* (pp 327). Blalock writes:

> “The handful of us who introduced causal modeling, path analysis, and structural equation modeling into the sociological literature during the 1960s encountered a problem similar to that faced by Wold (personal communication) a decade or so earlier. At the time, regression analysis was being used atheoretically and *causation* was a dirty word.”

In the same paper:

> “our attention at the time was devoted to developing the rationale for employing statistical techniques *along with supplementary assumptions* as an aid for assessing the fit between empirical data and predictions made from causal *models* about covariances and temporal sequences.”

As Blalock admits “Much of the flavor of that earlier literature has now been lost from view.”

An Intellectual Turmoil

While it is true that Wold was interested in various topics, most of his personal research was focused on the problem of simultaenous equations. The lack of a definite solution based on Ordinary Least Squares was accumulating a ticking bomb in his mind. As the decade of the 1950s approached its end, Herman Wold was having more and more difficulties handling all the prickly issues associated with causation and model-building.

A few months before turning fifty years old, Herman Wold packed his suitcases and flew to the United States, taking some time off from his Chair position of the Statistics Institute at Uppsala University. He had gladly accepted the invitation to spend some months between 1958 and 1959 as a Visiting Professor at Columbia University where he was expecting to make some progress around the econometric systems of multirelational equations. This topic was in reality the outer layer of a deeper problem, something much less mathematical and much more philosophical: the causal interpretation of statistical models. After a decade of vigorous attemps to make a point among the rest of the econometrics community, the harder he pushed on the causality issues, the more isolated he found himself. Trapped in the intellectual maze of his philosophical thoughts and his statistical modeling approaches, Herman hoped switching the Scandinavian airs for the estimulating New York atmosphere would help him find a solution.

Somewhere during his visiting period at Columbia University, Wold found himself lost in the middle of his philosophical arguments and his econometric multirelational systems. "An intellectual turmoil," Wold wrote, "built up within me." Unable to include a satisfactory causal specification in his models, he finally "gave up."

How does all this relate to PLS? Well, the period of time at the end of the 1950s is crucial for the posterior development of PLS. Wold had already proved that Least Squares could be applied for recursive models. His big obstacle, however, was the estimation of non-recursive systems, or what Wold used to called: interdependent models. This was one of the biggest challenges he had to deal with in his research career, at least during the 1950s. In his quest to find an OLS solution for non-recursive systems, but also in his journey across the causal labyrinth, all the roads he explored were dead-end roads. One attempt after another, Wold was not able to solve the non-recurvie problem and he finally ended up in a blind alley.

A New Start

At age 50 and after a decade of work Herman had still not found the dream solution for the interdependent systems. He was proud of his salvage of OLS applied to causal chains; that is something that kept him safe, allowing him to keep floating, and preventing his sinking. But exhausted after many years of wandering with no clear direction, he returned to the shore. Pondering between his philosophical causal notions and his statistical reasoning, Herman "decided to push the causal arguments on causal chains and interdependent systems aside, and make a new start."

If something was clear to Wold among all of his intellectual turmoil, it was his blind faith on Least Squares, which he was not willing to abandon. Wold agreed to push the causal arguments on causal chains and interdependent systems aside, making a radical decision: give up all what he had done and start all over again from a new conceptual and philosophical perspective. Wold made a new start on the basis of what he denominated *predictor specification*. This new approach in turn led very soon to the Fix-Point method and eventually to Partial Least Squares.

For his new start, Herman found the solution in a couple of convenient assumptions: the conditional expection in regression analysis for OLS. This offered him a way to tame the causality whirlwind and encapsulate it in a statistical concept that would let him speak in terms of prediction while, most important, kept using OLS for estimation purposes.

Roughly speaking, if we assume a functional relation between a response variable

$$y$$

and a set of predictors

$$x_1, x_2, \dots, x_p$$

:

$$y = f(x_1, x_2, \dots, x_p) + error$$

the conditional expectation:

$$E(y|x) = f(x_1, x_2, \dots, x_p)$$

implies that the the

$$error$$

term is uncorrelated with the

$$x$$

variables. If we take a step further an assume a linear function:

$$f(x_1, \dots, x_p) = b_1 x_p + \cdots + b_p x_p$$

then:

$$E(y|x) = b_1 x_p + \cdots + b_p x_p$$

which is the well known regression expression. Under mild assumptions, the regression coefficients

$$b_1, \dots, b_p$$

can then be estimated by OLS.

Today we don't even stop to think about its philosophical causal implications. However, to someone as deeply concerned with causation and model-building as Wold, this was not something of minor importance. He first referred to that relation as *unbiased predictor* (Wold 1963), then he changed it to *eo ipso predictor*, and finally he ended up calling it *predictor specification*. To Wold, this conditional expectation provided the "general rationale for LS specification and LS estimation." Predictor Specification, or *Presp* as he used to called it, "marks the comeback of least squares"—Wold declared.

What's behind this so-called *predictor specification*? Basically, Wold decided to leave causality aside and asked himself what could be achieved with a predictive relation. The joint multivariate distribution assumptions typically used in Maximum Likelihood were also to be left aside. The only thing that Wold wanted to take into account was the systematic part of a predictive relation expressed as the corresponding conditional expectation. For most of us this might seem trivial but for Wold it was the theoretical mechanism that allowed him to seize Least Squares and had a convincing epistemologycal argument that justified the use of OLS. In this way, by requiring just mild supplementary assumptions, Wold found a scape route to use OLS regression for estimating coefficients that are consistent in the large-sample sense. Moreover, he extended the notion of predictor specification to models with multiple relations of blocks of variables, which in turn would lead to PLS. # The Soft Modeler

Herman's agenda fell beyond his interests on econometric analysis. Above all, he was passionate about model-building and philosophy of science. He was also very much interested in more general modeling applications within socioeconomic and behavioral disciplines. After meeting the famous American psychometrician Louis Leon Thurstone and his wife Thelma in the early 1950s, Herman Wold was so impressed by the multivariate models of psychometrics that he decided to organize and host a Sympoisum of Psychometrics at Uppsala in 1953.

One of the things Wold found of enormous theoretical value was the use of *latent variables* or *factors*—as they were commonly referred—in psychometrics. The theories and multivariate tools around Factor Analysis had a longer tradition than econometrics. While simultaneous equations had been proposed in the early 1940s, the Factor Analysis methods could be traced back to at least 1904 with the works of English psychologist Charles Spearman and his theory about the *General Factor of Intelligence*. Likewise, the term Econometrics was coined in 1926 by Frish?, while Psychometrics had been in used since 1886, when James McKeen Cattell wrote his PhD thesis *Psychometric Investigations*. Both disciplines had in common the foundation of societies (Psychometric Society in 1936, and Econometric Society in 1934?).

Psychometrics and Econometrics both deal with phenomena and data of social nature. They both are based on theories that make use of abstract or latent concepts. The most famous concept in Psychology is that of *Intelligence*. Economics has its own latent concepts like *Utility*. The difference was in the way such theoretical concepts were handled in the mathematical models. Psychometrics literature had a longer tradition of including theoretical variables in their models playing an instrumental role. Instead, the theoretical variables in economics didn't have that status of *latency*. The forces of demand and supply were assumed to be like the force of gravity or any other physical force: invisible yes, but present. The force of demand was measured with the demanded quantities, and the force of supply was measured with the supplied quantities; there was not need to include variables of latent nature in Economics. The most latent concept perhaps was the "invisible hand of the market" in the metaphor by Adam Smith; but his concept was not directly modeled or incorpored in any model, it was the *equilibrium* or intersection between curves of demand and supply that were taken into account instead.

Expanding Interests

Herman Wold was soon captivated by the possibilities of the analytical tools and theories provided by the Psychometricians. And he was not the only one. It was a matter of time for a full crosspolination of ideas to take place within quantitative methods in Social Sciences including psychology, economics, sociology, and education, to mention just a few. The introduction of causal path models to Sociology and its expansion was a watershed. The excitement in the quantitative analysis among social sciences disciplines was enormous after the the seminal work of Duncan (1966), as classes of problems suddenly seemed to open to new approaches with the conceptual tools it provided.

A crucial fact that would have tremendous consequences for the years to come, was the proposal of Herman Wold to Karl Jöreskog, one of his students, to write a doctoral dissertation on Factor Analysis.

Iterative Least Squares

In the early 1960s, Wold was elected to the prestigious Swedish Academy of Sciences. His increasing reputation and recognition as a leading figure in Scandinavian Econometrics was unquestionable. It was also in this decade that his trajectory in econometrics reached the summit, and in 1966 he became President of the Econometric Society. Still, he continued to focus his energies on tackling the problem for estimating non-recursive—or interdependent—systems of equations.

After giving up the causal arguments for the estimation of structural equations, Herman Wold started to work on a new methodology that he dubbed the **Fix-Point** (FP) method. Using Least Squares and his theoretical argument of predictor specification, the FP method was developed as his new proposal to estimate interdependent (ID) systems via an iterative procedure of least squares regressions.

At the end of 1964, having a first stable version of the FP method, he went on a roadshow to the USA to promote his new analytical tool. One of the stops was the State University of North Carolina. There, during one of his seminars, G. S. Tolley, a professor in the audience, asked Wold whether the FP method could be applied to compute principal components on a dataset he had collected. The previous year, Tolley had published a study on farmer skills. After the seminar Wold discussed in more detail with Tolley and R. A. Porter the possibility to adapt Fix Point for computing principal components. This discussion "gave me the clue for computing principal components by an iterative procedure," Wold wrote. This unintentional discovery must had reinforced Wold's enthusiasm about his new method. Besides, it had all the elements that Herman liked: based on least squares, aimed at practical application, and intended for prediction uses. Such a case of "serendipity," Wold observed, "led me to an iterative procedure for the computation of Hotelling's canonical correlations."

Almost immediately a plan began to take shape in Wold's mind to further investigate the properties and scope of his recently discovered iterative least squares procedures. He suspected that much more could be gotten out of its new method. As soon as Wold returned to Uppsala, he set to work on the "by-product" of his Fix-Point method to compute Principal Components and Canonical Correlations. He recalled the work of his former PhD student Peter Whittle on factor analysis and principal components, and he decided to use Whittle's data set for experimenting purposes. Since the method involved iterations, Wold knew that a hand calculator was not going to be enough, so he asked his son Svante to write the computer programs for him. Likewise, his colleague Ejnar Lyttkens got his hands on to further study the iterative nature of the algorithm, proving the convergence of this new class of procedures. An opportunity to introduce his findings showed up in the summer of 1965 with the International Symposium on Multivariate Analysis at Dayton, Ohio. There, Wold presented an application of his **NILES procedure** on how to compute PCA using an iterative algorithm alternating simple least squares regressions. The following year the proceedings of the symposium were published in *Multivariate Analysis* (Krishnaiah, 1966) containing the paper by Wold *Estimation of Principal Components and Related Methods by Iterative Least Squares*. At the same time, he took the opportunity to write a more detailed and extensive paper when he was invited to participate in the Festchrif for Jerzy Neyman, also appeared in 1966: *Nonlinear Estimation by Iterative Least Square Procedures*. Historically, these publications are the roots of what years later would officially become the Partial Least Squares framework.

The work around NILES was definitely a *hat-trick* for Wold's team. In a one-year period they had accomplished fruitful results with all the elements that Herman liked: least squares, practical application, prediction uses… and it even provided the opportunity to treat data with "partial information" or missing values. One could use the iterative least squares procedures for Principal Components, for Canonical Correaltions, and for a handful of other applications. At last something good was emerging out of his new start. Confident that he was on the right track, he pushed his research on his FP method for tackling the problem of interdependent systems. Yet, Herman was cautious enough not to prematurily celebrate or spreading his findings. "All of the fresh procedures," Herman said, "are in an early stage of development." Besides, his primary focus and motivation was the interdependent systems. His NILES procedures were after all a "by-product," interesting and useful, yes, but not the main line of research for the rest of the second half of 1960s.

Psychometrics Influence

Meanwhile, Karl Jöreskog, one of Wold's former PhD students, had been opening paths and achieving breaktroughs in factor analysis. In 195X he achieved the computation of Maximum Likelihood estimation first proposed by Lawley in 1941. At the middle of the 1960s he went to the USA with the Educational Testing Service (ETS)—the famous educational testing and assessment organization in charge of developing various standardized tests primarily in the United States. With eocnometrician Goldberger, Karl started to work on finding a way to merge Psychometric models with econometric systems of equations. The so-much awaited synthesis would be achieved in 1969. This feat marked a borefe-and-after moment in multivariate model-building. Something which Herman Wold couldn't have ignored.

In the summer of 1957, when Karl was getting ready to start his professional career as a high school teacher of mathematics and physics, an unforeseen opportunity came to his life. He was visiting a friend one afternoon when the telephone rang. Professor Herman Wold was calling the friend offering him a research assistant position at the Department of Statistics. Since Karl's friend had already taken a job, he told Herman that he was not available but that he knew of a friend that was a good candidate and that he was right there, passing the phone to Karl. After a brief chat between Herman and Karl, Wold offered the young aspiring teacher the research assistant possition in the Statistics Institute at Uppsala University.

A year later, Herman Wold told Karl that his research position required him to study more statistics. So Karl "took all the undergraduate courses in statistics and then continued for a master's degree, which was completed in 1961". Showing a great talent and dispossition, Karl went on to do the PhD and finished it in December 1963. For his doctoral research, Wold had suggested Jöreskog to do a dissertation on Factor Analysis, continuing the work on the topic previously done by Peter Whittle in 1952.

In February 1964 Fred Lord invited Jöreskog to do research on the renowned *Educational Testting Service* (ETS) in Princeton. After one year, Karl returned to Uppsala "for half a year and taught as a lecturer in statistics". However Fred Lord made Karl an offer as Senior Research Statistician with tenure that he was not able to reject. Karl spent the next five years in Princeton until the summer of 1971 when he went back to Uppsala convinced and pushed by Wold, who even quit his Chair and took an early retirement so that Karl could have the Professor possition formerly occupied by Herman.

During the 1960s Karl stood out as one of the greatest figures and minds in Psychometrics, both at the methodological and computational levels. He accomplished various landmarks, and produced the necessary software that made possible the application and analysis of the new methods for practitioners. Among his most amazing achievements was the merging of the simultaneous equations in Econometrics with the Cofirmatory Factor Analysis of Psychometrics. With his approach to estimate *Linear Structural Relationships* models and his software **LISREL**, Karl would revolutionaze the way people think about models, latent variables, and testing of theories in social sciences. Following the tradition of Fisher, and Haavelmo, he performed all his work under the Maximum Likelihood principle.

Although formed under Wold's mentorship, Karl Jöreskog didn't embrace Wold's deep love for Least Squares. Actually, Karl didn't followed the econometric path of Wold's main research group. Seeing the modeling potential of psychometric models, the late promising advances in the field, and the talented skills for mathematics and statistics shown by Karl, Wold must have seen something special in his young research assistant to entrust him a research topic in which Herman was not precisely an expert. Wold was not mistaken. Karl stood out in Psychometrics and became full involved in the field during his ETS period.

Inspired by his former student's achievements and his LISREL models, Wold would had a revelation in the early 1970s. "In 1971 when I saw Jöreskog's LISREL algorithm for ML estimation of path models with latent variables, it struck me that the same models could be estimated by suitable elaboration of my two algorithms for principal components and canonical correlations." Wold would probably have pursued the ML approach of his pupil had he not being a Least Squares fervent advocat. Unlike the rest of the world that blindly followed the ML-approach for linear relation models, Wold set out on his own path.

The graphical representation of the models was practically the same in the *arrow schemes* used by Wold. He was well aware about what was hapenning not just in Econoemtrics but also in Psychometrics, Sociology, Education, and Causal Modeling. Not so long, Wold wondered whether it was possible to apply his iterative least squares procedures for the type of models proposed by Jöreskog. It was a very tempting yet natural question for Wold. The latent variables could be treated like principal components or canonical variates, and the structural relations could be posed in terms of regression equations, which in turn could be used for prediction purposes. A whole new unexplored route to a promising land was opening in front of Herman Wold's eyes. Freed from the administrative responsabilities of his previous chairman possition, and having an enviable research experience, he set out to see how far he could get with his new ideas, now as a professor at the Univeristy of Gothenburg, the third-oldest university in Sweden.

It didn't take Wold long before he had some preliminary results using his NILES procedures for path models with latent variables. Starting from two blocks of variables and two components, he jumped to models with three blocks and three structural relations. Like in 1965, the opportunity to show his findings came with the third International Symposium on Multivariate Analysis, again held in Dayton Ohio. Unconvinced of the name **NILES**, he changed the name to **NIPALS**, short for Nonlinear Iterative Partial Least

Squares. Also the term *procedures* was dropped and replaced by *modeling*. **NIPALS modeling** slowly reflected a more mature ideological framework: *"Nonlinear Iterative Partial Least Squares (NIPALS) Modelling: Some Current Developments"* (Wold 1973). "I see NIPALS modeling," Wold wrote, "as an open ended array of models with unlimited complexity in the combined use of several devices."

The middle of 1970s saw the evolution of NIPALS modeling from the first tentatives of expanding to three blocks of variables, to a more robust framework capable of handling any number of blocks and more complex structural relations. In 1977 Wold declared the ripenes of his PLS basic design method. Two years later, in 1979, Wold and Jöreskog organized a joint meeting at Cartigny, near Geneva, Switzerland, presenting both of their approaches (LISREL and PLSPM). The proceedings of that meeting would appear three years later in the form of a two-volume publication *Systems under indirect observation: Causality, structure, prediction*.

Chapter 7

The Chemometrician

Meanwhile in the late 1960s and early 1970s, another type of *metrics* discipline was experiencing revolutionanzing changes in Scandinavian countries and Northern Europe. In this case, the field was not part of the socio-economic sciences scene but part of the life and natural sciences, concretely Chemistry.

With the arrival of computers during the 1960s, and the increasing appearance of new electronic measurement devices, the field of Chemistry was irremediably touched by these technological changes. Consequently, a transition from "wet" chemistry to a more instrument based "electronic" chemistry gradually took place. One of the effects of having new electronic devices was the production of more data than what was commonly available before. So with the new instruments also came the need for analytical tools to crunch the raising tide of emerging data.

Traditional mathematical and statistical methods soon proved to be not enough for the new types of data. One of the main characteristics was the presence of many more variables than observations. Among the leading figures recognizing this change, and taking advantage of it, was Herman's son Svante Wold. Named after his grandfather Svante Arrhenius, the famous Swedish chemist—founder of electrochemistry, and Nobel Prize for Chemistry in 1903—Svante Wold also decided to follow Arrhenius's steps by studying Chemistry. He was also interested in computers, and he was a well versed programmer, sometimes helping his father implement various of his least squares-based algorithms.

The Born of Chemometrics

In 1971, Svante Wold, as a young professor at Umea University, Sweden, invented the word *chemometrics* for a grant application. "Chemometrics, the art of extracting chemically relevant information from data produced in chemical experiments." Svante, however, was not the only one concerned about extracting relevant information from chemical data. In the early 1970s, a couple of publications about pattern recognition of chemical data by Bruce Kowalski and Bender appeared in the *Journal of the American Chemistal Society*. "I read and reread and reread the two articles on pattern recognition", Svante recalled. To him, those publications were more than simple papers, "they were a revelation." Svante quickly realized something equally important: "suddenly, I was not alone in my feelings about the state of chemistry." The same ideas of analyzing multidimensional chemical data, through the application of statistics and applied math, via computers, was shared by other chemists in the west coast of the USA. Svante "was thrilled" by all this situation.

Invited by the renowned American statistician George Box and Bill Hunter, Svante spent the academic year of 1973-1974 at the Statistics Department of the University of Wisconsin in Madison, Wisconsin. In mid-October 1973 Box asked Svante to go to the University of Arizona, in Tucson, for an Office of Naval Research symposium on chemistry and computers. Coincidently, in that meeting, the so much awaited moment by Svante became true. He finally had the opportunity to met the young chemist professor Bruce Kowalski, whose papers had deeply impacted him so much.

At the end of the symposium, Svante had the chance to talk to Bruce about their common interests. One of Bruce's initial questions to Wold was: "What are you doing?," to which Svante answered "chemometrics." Surprised by Wold's answer, Kowalski continued: "What is that?" Svante patiently explained the meaning of the term, asserting that it was basically the same thing Bruce was doing, but "much less advanced." The word "chemometrics" was precisely the term Kowalski was looking for. Positively impressed by the young Swedish chemometrician, Bruce invited Wold the following year (1974) to Seattle for a month to discuss in more detail what they were doing.

Between May and June 1974, Svante went to visit Bruce in Seattle. There, Wold had the opportunity to show his–at the time new—pattern recognition methodology “soft independent modeling of class analogy”, most commonly known as SIMCA. Given the good performance of the method, it was decided to include it in ARTHUR, the package that Bruce and his research group were developing during that period.

Few disciplines can claim the exact day they were born. And Chemometrics is one of them. Just before Svante’s returning to Umea, Svante Wold and the group of Kowalski gathered on the evening of June 10, 1974, in a small Tex-Mex restaurant in Seattle. “After a number of Tequila shots”, Wold wrote, “we decided to form *The Chemometrics Society*” (which soon became “The International Chemometrics Society”).

The Born of PLS Regression

The early interest of chemometrics were mainly about pattern recognition and classification. Borrowing techniques from statistics (like discriminant analysis), and other computational approaches for pattern recognition, the new chemometricians soon discovered that such approaches were limited for the type they had to analyze. As Svante recognizes, “these approaches were still handicapped by the dogma that data had to have substantially more observations than variables.” This forced analysts to devise their own dimension reduction approaches and new methods of pattern recognition. Then, more predictive problems were added to the table.

By the late 1970s, as the PLS framework of Herman became more mature, Svante began to show some interest in the idea of latent variables, and the opportunities they seemed to open for the analysis of high dimensional chemical data. “The concept of latent variable,” Svante said, “appealed to me as it is very similar to the ‘effects’ we have in organic chemistry.” It didn’t take long for him to be “greatly attracted by the PLS philosophy” of his father. The problem, however, was on the implementation side which Svante recognized “impressed me less.” Svante’s concern had to do with an over-fitting tendency of his father’s method, mainly due to the large number of treated variables. Initially, he did not take PLS “very seriously.” Thanks to his father’s “enthusiasm and patient explanations of how latent variables stabilized the situation,” the reluctance shown by Svante slowly went down until he finally became convinced that the PLS methodology “was a well working approach with great potential.”

Having grasped the basics of Herman's PLS framework, Svante "started to work with the simplest PLS model (two-blocks) in the beginning of the 1980s. To his surprise, he realized that the model"could be looked on as a regression model in latent variables with a simple geometrical interpretation." This discovery made Svante "enormously happy," and he immediately decided to carried out a series of simulations to see "what happened with many variables." The results confirmed that many descriptor (explanatory) variables could be handled with no or minimal over-fit (provided cross-validation was used to check the number of components in the models). This confirmation showed him "that Herman was correct."

About 1979 or 1980, Svante met Norwegian chemist Harald Martens in Oslo. At the time, Harald was working with predictive models, using Principal Components Regression (PCR)—which not always provided good results. After some exchange of ideas, Svante "managed to convince Martens that PLS would be much better for his problems." Together, they "started to apply Herman's 2-block PLS," Wold wrote. But they didn't achieve much. Once again, the first results were not what they expected. "We had a big scientific crisis," Svante declared. Something was not working, but what? Svante and Harald shared their procedures, and they worked for days to find out the cause of the problem. "Harald and I talked on the phone for hours everyday between our programming." The main issue was related to the performance of the extracted PLS components: their models did not work beyond the first component. Cleverly, Harald realized that the problem had to do with the way the latent variables were obtained. Step by step they figured things out and "discovered the fantastic properties of the 2-block PLS approach." They had to make a couple of adjustments to Herman's algorithm for the method to work, but it proved very successful and promising. "We had a wonderful time," Wold remembered. "Those years between 1981 and 1984,", he said, "were among the happiest—scientifically—in our professional lives."

Svante presented a first version of the new PLS regression method at a conference on Data Analysis in Food Research arranged outside Oslo in 1982 by Harald and his boss Russwurm. Formalizing the procudere, Svante, Harald, and Herman proposed the new technique for regression analysis based on Partial Least Squares in 1983. They published the first papers on PLS and multivariate calibration in 1983, and the analytical chemistry community was informed soon after. Improved and modified during the middle 1980s, slightly different versions of the PLS regression were developed by Svante, Harald and other colleagues. The PLS regression framework started to revealed as very practical, useful and valuable. In an industry where there was a need for analytical tools capable of dealing with multicollinearity, missing values, and large data sets, PLS Regression fulfilled all those needs.

Like Herman Wold in econometrics, his son Svante Wold became also a pioneer and leading figure in his field of expertise: chemometrics. Unlike his father, Svante's trajectory would take a different path. Svante Wold and colleagues adapted and reshaped Herman's ideas. Stripping away the more theoretical issues, they focused more on the practical issues, and computational aspects. Their motivation was more pragmatic and driven by the type of data challenges they were dealing with. While Herman Wold's developments were always framed within an econometrics model-building tradition, and looked at through the philosophy science glass, the behavior of Svante and his collaborators was more industry influenced. There was not much room for philosophical considerations and causality considerations.

The Eclipse of Herman's PLS

At the time of Herman's retirement—which were also his last years—the body of knowledge about Partial Least Squares had grown considerably. By 1989, three years before his passing away, Herman had produced around 20 publications, as well as a couple of dozens reports and working papers about his PLS methods. Also in 1989 the book *Latent Variable Path Modeling with Partial Least Squares* by Jan-Bernd Lohmöller was published, and the software LVPLS was available. PLS, in its Path Modeling version, was already ripen, and it seemed to be a matter of time before taking off.

PLS did take off, but not the father's version, but the son's regression framework. At age 83, Herman passed away on February 16th, 1992 in Stockholm. He had left a tremendous legacy, and he had already passed the torch to various of his students and collaborators. Unfortunately, the most promising figure, Lohmöller, also died in the early 1990s. Orphan and abondoned to its own luck the so-called "PLS soft modeling on path models with latent variables" started a slow but constant decline. There was nothing that stopped PLS Regression from eclipsing his older sibling PLS Path Modeling.

By the early 1990s, PLS Regression was on the summit. Svante Wold was promoting it within Chemical industries all over the world. Not only that, Tormod Naes and Harold Martens were also among the prolific authors, proposing adaptations and expanding the method to new directions. Software was not an issue, there was SIMCA-P (now Umetrics) and Unscrambler. SAS would later add its PLS-R procedure. Consequently, it soon attracted the attention of other experts in the field, expanding to oil industry, food science, cosmetics, and other related subfields of Chemistry.

In 1992, French practitioners and researchers started to have their first contact with the PLS regression methods. Among the consulting activities of Svante, he was difussing his methods and software around the European Chemical Industry. In France, he consulted for L'Oreal introducing his tools and software. Before long other companies were catching up with this method. One of those companies was Rhone-Poulenc. At that time, the head of the data analysis research department, Jean-Pierre Gauchy, asked statistician and professor Michel Tenenhaus to explain the PLS methods and train the company's researchers. Michel, a french professor in business school and also a consultant for industry, accepted the request of Jean-Pierre and started to study the regression via PLS. Nothing would prepare Michel for the coming revolution of the following years.

Chapter 8

The Renaissance of Herman's PLS

The PLS Regression of scandinavian chemetricians was the perfect match to the French multivariate analysis school *Analyse des Données*: both favoring the same core concepts of projections, inductive style, interpretable tools, and graphical displays. With a long tradition of Multidmensional Analysis methods, French data analysts have had their preferences for tools with a heavy use of geometry, algebra, and projection methods. Rooted in the seminal work of Jean-Paul Benzécri during the mid-1960s, data analysis *alla Française* privileged data analysis with a decisive inclination to exploratory and descriptive approaches—similar to the spirit of John Tuckey's *Exploratory Data Analysis* (EDA). One limitation of the French vision, however, was the lack of more explanatory and causal modeling techniques that matched the geometric spirit without privileging inferential procedures based on unrealistic assumptions about the data (like the "soft modeling" ideals of Herman Wold).

PLS Regression was a revelation to Michel Tenenhaus, and he was immediately caught up studying and researching the method. Soon after, he started to prepare a book on that topic. "During the writing of this book," Michel remembers, "I had many interrogations on many points." Invited by Svante Wold, Michel spent some time in Sweden to discuss various issues. Michel asked a number of questions about PLS which Svante answered and clarified. But not all them, "some formulas in SIMCA-P are still secret!" Michel admits. Among the copious bibliographic sources Svante gave Michel, there was a green hardcover book that Svante's father, jointly with Karl Jöreskog, edited in 1982: *Systems Under Indirect Observation*, now considered one of the classical references about the PLS framework of Herman Wold.

Being a professor in the prestigious Business School HEC, Michel asked Svante about PLS applications in management. But Svante didn't have much material around management or business applications, so he referred Michel to Swedish professor and consultant Clases Fornell, Business and Management professor at University of Michigan, and also an expert in Customer Satisfaction Measurement.

Customer Satisfaction

Claes Fornell had been introduced to PLS around the late 1970s by American Professor and morphometrician Fred Bookstein, while both were teaching at the University of Michigan. In turn, Fred Bookstein was an expert in morphometrics who had had the opportunity to learn about Herman Wold's methods and even collaborate with him. In fact, a couple of geometrical interpretations of PLS-PM were provided by Bookstein (1980, 1982). Among one of the most cited PLS papers on marketing applications is the 1982 article "Exit Voice", written by both Fornell and Bookstein in the Journal of Marketing Research. Claes found PLS very interesting and convenient, and he even edited a two volume book published under the title *A Second Generation of Multivariate Analysis* (Fornell 1982). And he also taught those methods to students in the Business PhD Program at Michigan University.

Claes Fornell used the PLS Path Modeling approach for Customer Satisfaction, and he, like Svante, used it not only for academic research but also for consulting purposes. In this case, the method was employed by the firm *Claes Fornell Group* (CFG). The main application of PLS-PM by Claes Fornell was on Customer Satisfaction. Among his entrepeneurial research projects he developed the *Swedish Customer Satisfaction Barometer* in 1989. And then he started the project for developing the *American Customer Satisfaction Index* (ACSI) in 1994.

Michel Tenenhaus traveled to Michigan to meet with Claes Fornell who talked to him and presented him PLS applications in management. Claes showed Michel applications in Customer Satisfaction and talked about the

PLS algorithm. But something didn't click with Tenenhaus. Although Fornell was talking about Partial Least Squares, his version was not exactly the one used by the chemometricians. There was obviously something that Tenenhaus didn't quite get. After all the research Michel had done on PLS regression, he was totally perplexed by the method Fornell was describing to him. According to Michel this was "a very unpleasant situation". Had Svante forgotten to mention something else? Was Michel missing something important?

There was clearly something wrong. Under the same term "Partial Least Squares," the version of the chemometricians was not at all the same version Claes Fornell had shown Michel. Strangely though, since both methods had the PLS label and *Wold* was the last name of the main author for both versions. Where did Tenenhaus's confusion come from?

American Influence

Jointly with Claes Fornell, there was another leading expert of PLS-PM who was part of the very reduced group of American researchers making intensive use of the PLS approach: Wynne Chin, professor of Decision and Information Science.

In the early 1990s, Chin found himself in the middle of a crossroad. On its way to becoming a leading figure in the research field of *Management Information Systems* (MIS), Wynne Chin had also been developing the software program *PLS-Graph*, named after the little known—at the time—data modeling approach of Partial Least Squares Path Modeling.

Basically, research on the field of *Management Information Systems* is concerned with the problems related around management and usage of information technology and information systems: the use of Information Technology systems, how they are developed, used, and applied in organizations and industries. Among the quantitative tools for studying these problems, MIS researchers use a set of data modeling methods better known under the umbrella term of *Structural Equation Models*. Historically, this type of models have been borrowed, used and adopted from other disciplines, mainly from social and behavorial sciences rooted in psychometrics and causal modeling literature.

In his early 30s, Chin was in that singular spot that makes someone being a skilled master of some unique tradition, and at the same time being one of the few possessors of such knowledge. It looked like if Wynne was in a privileged power play position... except for the fact that almost no one else seemed to care much about that obscure PLS methodology. Trained in biophysics and chemical engineering, Wynne Chin enrolled to the MBA program at the University of Michigan in 1980, and then he continued with a PhD in Computers and Information Systems. He got his first contact with PLS-PM from the seminars taught by Claes Fornell. But more important, Chin intermittently worked over a period of ten years developing a series of computer programs that eventually would led to the final version of his *PLS-Graph* software. In the early 1980s he helped with the programming for a version on the then still in use main frame computers. Shortly after the introduction of the first IBM PCs, he also started to develop a version for MS-DOS.

Wynne Chin considerably helped spread PLS-PM among MIS researchers, and he even released for free his PLS-Graph, under the condition that it would only be used for academica purposes, and that it wouldn't be extended or developed wihtout his approval—something that didn't happened. Although the development of PLS-Graph has considerably slowed down, at the time of this writing, it is still available previous contact with Wynne Chin.

Before Chin's software was available, the other complete computer program for PLS-PM was Lohmöller's *LVPLS*. Distributed at nominal charge, LVPLS was not a commercial software but an academic one for research-purposes. This short supply of PLS software was a substantial difference compared to the software available for covariance-based Structural Equation Models. By that time the most popular software was *LISREL*, developed by Karl Jöreskog and Dag Sörbom. At the end the 1980s *LISREL* was already in its version VI with its *microcomputer* PC-based program. Version 7 was launched in 1989 integrated within the larger software *SPSS*, and it already had a graphic interface intended to make it ease for users.

Since none of the PLS versions had a graphical interface, Wynne Chin started to work on a software consistent with Lohmöller's PC version around 1987. Although it was working on Windows 2.0 back then, the reliable version required Windows 3.0—mainly for Wynne Chin's own use. According to Wynne, the first official presentation of his software was in 1988. However, the so-called PLS approach remained largely unknown eclipsed by the more commercially successful *LISREL*, widely available and converted in the *gold-standard* for SEM at that time. The PLS method was so unique that it suffered from a lack of interest in the best of the cases, and from multiple misundertandings in most occasions.

The first half of the 1990s decade also coincided with the passing away of Jan-Bernd Lohmöller and Herman Wold. “At that point, it was clear no one else was supporting the use of PLS,” Chin recalled. Watching how the method was fading away at an alarming rate, Chin made what he believed was the most appropriate decision at that time. As a last resort to keep alive the PLS approach among his colleagues, he decided to share *PLS-Graph* within the MIS community where he kept it “officially free for academic research” purposes.

The Path Modeling Renaissance

After the initial confusion Michel had with PLS Path Modeling, he soon realized about the attractive modeling possibilities the method provided. Greatly interested in this method, he began a long trip that would eventually rescued PLS-PM from its ostracism. In 1999 Michel Tenenhaus and Alain Morineau organized an international symposium in Paris bringing together users of both methodologies: PLS Regression as well as PLS Path Modeling. Given the produced enthusiasm and interest of the participants, it was decided to organize a PLS Symposium every two years dedicated to all PLS methods.

Likewise, the Customer Satisfaction movement of the late 1990s inspired European members to proposed a similar alternative to the American Customer Satisfaction Index, but this time adapted to the European geography and economy. A consortium of universities and researchers was formed and the ESIS project was born with the goal, among other purposes, to continue the development of PLS-PM software. This also sparked a fire of interest for the methodology, bringing a revival and mass difusion.

Chapter 9

Critical Review

In the previous chapters I've tried to provide a comprehensive account of the many events that have shaped the evolution of PLS methods. However, they are not enough to have a complete understanding of the status PLS methods reached at the end of last century. The lack of initial interest and attention that Herman's PLS Path Modeling framework suffered, was not an issue for Svante's PLS Regression proposals. No one can deny the overwhelming popularity that PLS Regression-related techniques have over PLS Path Modeling approaches. Likewise, no one argues that LISREL and covariance-based SEM approaches are also better known than SEM-PLS.

For better or worse, Herman's PLS framework has occupied an awkward place in the multivariate data analysis spectrum. His legacy is astounding and remarkable. But it was packaged in his very particular idiosyncratic style. However, there have been very few attempts to clarify this situation in the related literature. Consequently, I think it's important to provide a critical review that allows us to scatter the fog that has impeded a clearer panaroma. In this chapter, I want to provide my personal assessment of some of the crucial factors that determined the destiny of PLS methods in general, and PLS-PM in particular. I will focus on the following points:

- Software availability
- Presentation and notation formats
- Communication and channels of distribution
- Herman Wold's personality

Software Availability

A crucial factor to understand the itinerary followed by PLS methods has to do with the computational tools. One fundamental difference between PLS Path Modeling and PLS Regression had to do with the initial availability of software programs, evidenced by the lack of programming skills of Herman versus the highly talented coding abilities of Svante. This is the perhaps the most important aspect that negatively affected the spread of Herman's PLS during the 1980s and 1990s. Although Herman Wold did use computers for calculations and estimations, he never wrote a single line of code.

In the *"The ET Interview: Professor H.O.A. Wold (1908-1992)"*, David Hendry and Mary Morgan presented the transcribed interviews they had with Herman Wold. When they asked him about his involvement in computing, Herman answered:

> I never did any computer programming, but I did use computers a great deal and did a substantial amount of computational work. I never relegated this purely to research assistants, and at one time I was one of the main computors. In the thirties and forties we mainly used calculating machines; later it was computers based on programs written by others.

Herman was dependent on the help of his students and colleagues, and even on the help of his son Svante to program his methods. While it is true that there was always some program to perform PLS analyses, they were mostly academic versions written by members of Wold's team such as Areskog, Baldwin Hui, H. Apel, and S. Wold. Those programs, mentioned in footnotes of several of Wold's papers, were available in exchange of a fee (nominal value). However, nothing indicates that there was ever an attempt to develop a commercial version like those available for LISREL. It wasn't until the mid-1980s that a full version became available with Jan-Bernd Lohmöller's remarkable LVPLS software. However, its development was truncated because of Lohmöller's premature dead.

Jöreskog's LISREL, in contrast, almost from the beginning, had accompanying software. Also PLS Regression. Above all, this reflects the generational gap between Herman, who never became involved in programming, and Karl

Jöreskog and Svante Wold who were skilled programmers, and had the spirit to develop commercial software. Not having soon a dedicated software greatly affected the spread of Herman's PLS. Had users have access to PLS-PM software, it wouldn't have taken so long for practitioners to see the potential and advantages. The proof of this is given with the formidable PLS-Graph program developed by Wynne Chin in the early 1990s. It's not an exaggeration to say that thanks to PLS-Graph the PLS-PM usage was kept alive, prevented from extinction. This required Chin to make a desperate move by freely releasing his software for academic purposes. Looked in retrospect, his actions had a tremendous positive impact, blessing the Information Systems as well as the Marketing Research communities.

Presentation and Notation Format

A less obvious factor but perhaps one with an equal negative contribution has to do with the format and language used by Herman to present his methodology. The way Herman Wold presented PLS Path Modeling requires much more effort to be fully understood, compared to the less demanding requirements to grasp the descriptions offered by Svante Wold and colleagues.

The Path Modeling work of Herman is described with a long list of theoretical model specifications that, when looked at for the first time, enormously difficult its understanding. In constrast, the Regression manuscripts are characterized with a simpler, cleaner, and straightforward presentation. Few statistical methodological works have engraved in them so much of an author's idiosyncrasy as PLS Path Modeling with Herman Wold's ideological stance. This is also evidenced in the philosophical canopy that covers the most rich mathematical, algebraic, and algorithmic elements.

Notably, there is no pesudo-code description of the PLS-PM algorithm in any of Herman's papers. Although this may be a bit shocking, it is not of a big surprise given the fact that Herman never did any computer programming. In contrast, almost all manuscripts by Svante Wold and other chemometricians, include a description of the PLS regression algorithm using pseudo-code notation, even sometimes providing the instructions in some programming language code (e.g. FORTRAN or matlab). Clearly this has a major impact on users who want to write their own scripts. For a user with basic programming skills, implementing PLS regression from the provided pseudo-code is straightforward. A completely different story is when someone tries to implement PLS-PM from any of Herman's papers (I challenge you to do that).

Likewise, Herman's writings tend to be accompanied by philosophical notes about causality in social sciences, something that is totally absent in the work of Svante. On one hand, Herman's work is tightly embeded in his econometrics tradition, always seasoned with his ideological posture about the non-experimental nature of socioeconomic data, and the role statistical methods play on this regard. On the other hand, Svante's work is embeded in the chemometrics field, dealing with physico-chemical data of a more experimental nature, and reflecting a decisively pragmatic spirit. Right from the beginning, the language used by Svante Wold has nothing to do with Herman's interests about philosophy of science, causality, and the confrontation of Least Squares -vs- Maximum Likelihood.

Another major difference has to do with the areas of application. Herman Wold applied his methods mainly for economic-sociological applications. Svante Wold and colleagues applied their methods to chemical data and related industries, which deal with more pragmatic and practical problems. While Herman Wold applications were of more theoretical nature, Svante's were the opposite.

Although the PLS regression algorithms were developed from adapting the basic PLS-PM algorithm, the PLS regression branch has some contrasting and even extreme differences with the path modeling method. The notation changed radically when PLS regression was introduced. Svante and colleagues made a conscious decision of coming up with a fairly standard notation since the very early days. Pertinently, they employ more vector and matrix notation which greatly improves the reading of equations. In addition, there is almost no use of greek letters, which avoids having cluttered expressions so common in path models. Likewise, the overall structure of the PLS-PM algrithm disappears under the PLS Regression adaptation. Svante Wold and colleagues greatly simplify the algorithmic steps, coming up with a minimalist version that at first glance has no resemblance with any of the procedural steps in Herman Wold's publications. At the conceptual-algorithmic level, this is perhaps the trait that stands out the most, and impedes the reader to see PLS Regression as a slightly modified version of the PLS Path Modeling algorithm.

On the terminology side, Herman used a long series of terms and acronyms: NILES, NIPALS, NIPALS procedures, NIPALS modeling, PLS modeling, soft modeling, basic design, etc. And of course more technical terms like inner and outer models, estimation of weights, estimation modes A, B, and C, etc. Some of this terminology is adopted by the chemometricians, specifically the acronym PLS, and the concept of latent variables. The idea of Soft Modeling, or more precisely, the predictive-modeling purpose. But the rest of the jargon is discarded.

During all the development years, and across all publications by Wold, it is word mentioning the following remarks:

- heavy specification (very verbose)
- minimal use of vector and matrix notation
- abundant philosophical and ideological content
- often descriptions of model building
- opposition of OLS -vs- ML
- no pseudo code notation
- never used the term "Structural Equation Modeling"

In short, the real difference between PLS Regression and PLS Path Modeling is in the chemist trait of writing formulas and algorithms like cooking recipes of the son, versus the more verbose and epistemological oriented equations of the philosopher father. Reduced to their bare bones, both approaches are not as different as they might seem to be.

Communication and Channels of Distribution

Equally important was the way in which most of Herman Wold's publications about PLS appeared; in particular *where* they appeared. More than 50 percent of Herman Wold's publications about PLS are in the form of chapters in proceedings books or edited books, with a very limited reach. This means that the scope of his work was highly restricted. Only a few readers would have access to this body of literature.

Also unfortunate was not having a publication in an academic journal of prestige and mass scope; among the more than 20+ PLS-PM publications of Herman Wold, none of them appeared in a renowned journal. Jöreskog published in the famous *Psyhometrika*. Svante Wold was more clever and he published in his field journals *Journal of Chemometrics*, and later in *Chemometrics and Intelligent Laboratory Systems*. Why did Herman choose to publish his works the way he did it? In his autobiographical essay *"Models for Knowledge"*, Herman (H. Wold 1982a) writes on page 205:

> Among experienced researchers working at the periphery or outside the mainstream, there is some consensus that established learned journals adhere to the mainstream; their referees do not have the function of evaluating contributions that break away from this mainstream.

Obviously his research and work around PLS was outside the mainstream literature. Herman adds:

> Time-consuming and futile debates with referees can be avoided by going elsewhere to publish dissident research, say in Festschrifts or other occasional volumes.

This helps us explain why he so frequently used "other occasional volumes" to publish his work. At a time when there was no Web—with blogs, videos, podcasts, etc.—and when academic editorial houses were the official channel of academic mass communication, Herman remained using secondary publications. Clearly, he encountered opposition from referees and editors that didn't understand his PLS framework. Perhaps the only text that might have raised attention and attraction was the 2-volume book *Systems Under Indirect Observation*. But this was also a specialized publication, not addressed to a more general audience.

Another fundamental issue was the lack of text books. The only publication in the 1908s fully dedicated to PLS was Lohmöller's book *Latent Variable Path Modeling with PLS*. But this was a specialized manuscript based on his PhD dissertation, not a text book for teaching or demonstrating users how to use and apply PLS-PM. Even worse, it didn't help the style and structure chosen by Lohmöller. His book is written in such a dense style that its content is extremely hard to understand for the average PLS reader. It is definitely not a manuscript for beginners. And even for experienced PLS users, it is still not an easy reading. A review of the book by Charles Bayne is a good example of Lohmöller's rigid style:

> The author describes his goals for his book as threefold: (1) to make mathematical reasoning the core of this monograph, (2) to explain the computer program PLS, and (3) to provide insights from new applications. The author states: "Personally, I don't mind if this monograph is all of 'too' mathematical, computerized, applied and didactic at the same time" (p. 5). The author does not completely achieved his goals, however.

A more favorable opinion of Lohmöller's book is given by Pieter Kroonenberg (1990):

> It is unique and a similar book is unlikely to appear on the market in the next five to ten years. It is a scholarly treatise on a subject that deserves to be more a part of the mainstream (especially in psychometrics) that it is at present.

The book of Lohmöller is definitely something to study and read for anyone interested in doing research about Partial Least Squares. It is not, however, a book for beginners, or for inexperienced users and practitioners. Not because its content is uninteristing or lacks quality, but because of the difficulties in its reading with complicated notation, symbology, and dense content.

Herman Wold's Personality

In addition to the points previously discussed, there's one more element without which it one cannot entirely explain the way in which the original PLS framework was developedL the personality of Herman Wold. Among his personal characteristics, there are three fundamental traits concerning the story of PLS:

- His relentless obsession with the principle of Least Squares.

- His wide interdisciplinary interests.
- His tendency to work on his own.

Herman Wold was passionated about Least Squares. In his autobiography (H. Wold 1982b), when reviewing his research trajectory divided in four main periods (time-series 1932-1938; economic analysis of consumer demand, 1938-1952; econometrics and path models with directly observed variables, 1945-1970; and systems analysis and path models with indirectly observed variables, 1966-onwards), he makes it very clearly:

> Throughout these 50 years, my contributions have been based on the least squares (LS) principle.

The attention that Maximum Likelihood attracted, and its mainstream place on the statistical literature, was something that Wold never felt comfortable with. Even at the end of Wold's academic life, in his last publications he often wrote lines reflecting his mixed feelings respecting but rejecting at the same time the method of Maximum Likelihood. Here's a revealing excerpt from Herman Wold's interview for the *Econometric Theory* (Hendry and Morgan 1994):

> (ET) Would it be fair to see the least-squares principle as the consistent link through your research contributions in statistics, econometrics, and time series? (HW) Yes, I think it is. I agree with that very much. (ET) And was this conscious—that you felt least squares was right and you kept applying it as you tackled a problem? (HW) Yes, yes, yes, . . .

Herman was as much a statistician and econometrician as a philosopher and thinker. His thoughts often crossed the thin line, not always clearly marked, that divides mathematics-statistical concepts from theoretical concepts of a more philosophical and epistemological nature. Playing both roles of protagonist and witness, Herman couldn't scaped to the revolutionary events hapenning at the time. He had not only taken an active role in econometrics, but he had also have being interested in psychometrics, causal analysis, and path analysis. However, instead of joining the efforts of the mainstream with linear structural models, maximum likelihood, covariance structure analysis, and the like, he decided to follow his own path. That decision was not random, it was part of who Herman was—his character and his way of thinking.

One of the most captivating texts about Herman Wold is the obituary written by Petter (Wold 1992)—former PhD student of Herman during the 1950s. In this touching document, Whittle portraits Wold's work style in the following terms:

> His tendency to work on his own, using his own methods, was a strength, but also a dissadvantage in that the advances that he was making were less known and recognized than they should have been.

In the same obituary Whittle writes:

> "Herman's very striking trait was his determination and single mindedness. Once he had embarked on an interest or a course of action he would pursue it for a matter of years, ignoring obstacles, which sometimes also meant ignoring susceptibilities. This characteristic made him something of a byword; it was said that whenever two or more Scandinavian statisticians were gathered together then the conversation would inevitably soon turn to the subject of Herman Wold."

So what can we say about Wold's personality influence on PLS? My assessment is this one. Taking a different route, Herman Wold saw the opportunity to come up with his own interpretation of model-building with latent variables, making use of his toolbox of iterative procedures based on least squares. Above all, he followed a vision radically different to the one adopted by most other researchers. In regards of structural equation models, Wold was dealing with almost the same type of models, at least visually speaking (using common graphical representations), and with the use of latent variables and systems of linear relationships. But conceptually, operationally, and ideologically, Herman Wold went on an another direction, away from the mainstream proposals. This is something that was not fully appreciated while Herman was alive—and even today after more than 30 years that PLS has been around us.

Chapter 10

Conclusions and Final Remarks

What we know today as Partial Least Squares is the result of a long period of evolution, with a vast range of methods and techniques proposed since the late 1960s / early 1970s. They come from different disciplines and fields of application, motivated to solve a number of multivariate problems—involving relations among one or more blocks of variables. This means that PLS methods have grown organically over several decades, gradually mutating both in form and substance, and migrating from one field of study to another.

Even though PLS methods have been built on the NIPALS computational foundations of the mid-1960s, historically—and ideologically—we can distinguish two main branches of Partial Least Squares approaches: 1) the "Path Modeling" branch, and 2) the "Regression" branch—both based on works originally introduced by Herman Wold. The recognition of these two large categories has to do with the way in which they have subsequently unfolded. By taking different directions, the branches have produced two major *movements* that for the most part, have grown apart, remained disconnected, and even unaware of one another in some areas of application.

Overall, the two PLS branches, accompanied with their respective analytical movements, reflect the works emerged from Herman Wold (the father) and Svante Wold (the son). So far I've tried to offer an account that allows us to get a better understanding of the differences between both works. I firmly believe that additional comprehension can be gained if we compare them specifically from a generational perspective, jointly with the respective expertise of each author.

The way and style in which the two Wolds presented their works have left a profound footprint in their ulterior developments. At first glance, the Path Modeling movement, emerged from systems of equations with latent variables, seems to have insurmountable disparities with the Regression Models stemmed from chemometrics. Both branches, with its various subdivisions, show contrasting differences at various levels, notably at the area of application, but also at a philosophical, ideological, conceptual, language, technical, spread and diffusion levels. Albeit their common mathematical and operational elements, and even their shared genes, the contrasting physical appearance between path and regression models puts an illusory divide between them that can easily mislead all inexperienced users, and even some well versed PLS *connoiseurs*.

While it is true that such differences are not negligible, most of them are at the format level (i.e. the presentation). It is surprinsing how their outer layers pull them apart more than bringing them closer together. Despite all the differences, there are still strong ties and similarities. The most important common denominator is the mathematical and algorithmic principles. These common traits can be exploited to link them back together and see them both under the same glass. Fortunately, this separation has been reduced considerably in the last years, thanks to the organization of PLS symposiums, and the active work of researchers who have taken care of bridging the existing gaps.

Despite all their differences and constrasts, it seems a bit unfair to keep isolated PLS-PM from PLS-R. This should not be the case anymore, specially when it can be shown the connection between both approaches. Similarly, there is also a drawback with the way PLS Regression methods haven presented in the chemometrics-based literature. As much attractive as it can be their minimal style, it misses so much of the richness of PLS-PM.

A Third PLS Culture

Today we can say that a contemporary conciliatory culture has taken the best of both schools, keeping the path modeling roots of Herman Wold, the very pragmatic side of the chemometricians, and adding a refreshing view with a strong multivariate data analysis flavor.

Regardless of whether you are a newcomer or an experienced user of some PLS methods, I think it is important to talk about these issues. They help us be aware of the differences in the literature, be conscious about difficulties, but also be open to new opportunities and advantageous analytical possibilities.

Enough talking. It's time to have some pudding!

Bookstein, Fred L. 1980. "Data Analysis by Partial Least Squares." In *Evaluation of Econometric Models*, edited by Jan Kmenta and James Bernard Ramsey, 75–90. Academic Press.

Fornell, Claes. 1982. *A Second Generation of Multivariate Analysis*. Praeger.

Hendry, David F., and Mary S. Morgan. 1994. "The eT Interview: Professor H.O.A. Wold: 1908-1992." *Econometric Theory* 10: 419–33.

Kaergard, Niels. 2012. "Haavelmo and Denmark." *Nordic Journal of Political Economy* 37: 1–20.

Lohmöller, Jan-Bernd. 1989. *Latent Variable Path Modeling with Partial Least Squares*. Heidelberg: Physica-Verlag.

Sanchez, Gaston. 2009. "Pathmox Approach: Segmentation Trees in Partial Least Squares Path Modeling." PhD thesis, Universitat Politècnica de Catalunya.

———. 2013. *PLS Path Modeling with R*. Trowchez Editions.

Tenenhaus, Michel. 1998. *La Régression PLS. Théorie et Pratique*. Paris, France: Editions Technip.

Wold, Herman. 1963. "On the Consistency of Least Squares Regression." *The Indian Journal of Statistics, Series A* 25 (2): 211–15.

———. 1966a. "Estimation of Principal Components and Related Models by Iterative Least Squares." In *Multivariate Analysis. Proceedings of an International Symposium Held in Dayton, Ohio, June 14-19, 1965*, edited by P. R. Krishnaiah, 391–420. Academic Press.

———. 1966b. "Nonlinear Estimation by Iterative Least Square Procedures." In *Research Papers in Statistics. Festschrift for J. Neyman*, edited by F. N. David, 411–44. Wiley.

———. 1973. "Nonlinear Iterative Partial Least Squares (NIPALS) Modelling: Some Current Developments." In *Multivariate Analysis III. Proceedings of* *the 3rd International Symposium on Multivariate Analysis. Dayton, Ohio*, edited by P. R. Krishnaiah, 383–407. Academic Press.

———. 1975a. "Soft Modelling with Latent Variables: The Non-Linear Iterative Partial Least Squares (NIPALS) Approach." In *Perspectives on Probability and Statistics, Festschrift (65th Birthday) for M. S. Bartlett*, edited by Joseph Gani, 117—142. Academic Press.

———. 1975b. "Model Construction and Evaluation When Theoretical Knowledge Is Scarce." In *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling*, edited by H. M. Blalock, A. Aganbegian, F. M. Borodkin, Raymond Boudon, and Vittorio Capecchi, 307–57. Academic Press.

———. 1977a. "Mathematical Economics and Game Theory: Essays in Honor of Oskar Morgenstern." In, edited by R. Hehn and O. Moeschlin, 536–49. Springer-Verlag.

———. 1977b. "Quantitative Wirtschaftsforschung: Wilhelm Krelle Zum 60. Geburtstag." In, edited by Horst Albach, Ernst Helmstädter, and Rudolf Henn, 729–54. J.C.B. Mohr.

———. 1978. "Ways and Means of Multidisciplinary Studies." In *The Search for Absolute Values in a Changing World*, 2:1023–33. Sixth International Conference on the Unity of the Sciences (San Francisco 1977); The International Cultural Foundation Press.

———. 1979. "Causal-Predictive Analysis of Problems with High Complexity and Low Information: Recent Developments of Soft Modeling." In *The Re-Evaluation of Existing Values and the Search for Absolute Values*, 2:1023–33. International Conference on the Unity of the Sciences (7th, Boston 1978); The International Cultural Foundation Press.

———. 1980a. "Model Construction and Evaluation When Theoretical Knowledge Is Scarce." In *Evaluation of Econometric Models*, edited by Jan Kmenta and James Bernard Ramsey, 47–74. Academic Press.

———. 1980b. "Soft Modelling: Intermediate Between Traditional Model Building and Data Analysis." In *Mathematical Statistics*, 6:333–46. Polish Scientific Publishers.

———. 1982a. "Soft Modeling: The Basic Design and Some Extensions." In *Systems Under Indirect Observation: Causality, Structure, Prediction*, edited by Karl Gustave Jöreskog and Herman Wold, 2:1–54. North-Holland.

———. 1982b. “Models for Knowledge.” In *The Making of Statisticians*, edited by Joseph Gani, 190–212. Applied Probability Trust, Academic.

———. 1985. “Partial Least Squares.” In *Encyclopedia of Statistical Sciences*, edited by S. Kotz and N. L. Johnson, 6:581–91. Wiley.

———. 1989a. “Introduction to the Second Generation of Multivariate Analysis.” In *Theoretical Empiricism: A General Rationale for Scientific Model-Building*, edited by Herman Wold, vii–xl. Paragon House.

———. 1992. “Obituary: Professor Herman Wold.” *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 155 (3): 466–69.

Wold, Herman, and Ejnar Lyttkens. 1969. “Non-Linear Iterative Partial Least Squares (NIPALS) Estimation Procedures.” *Bulletin of the International Statistical Institute* 43: 29–51.

———, ed. 1989b. *Theoretical Empiricism: A Rationale for Scientific Model-Building*. Paragon House.

Wold, Svante, Harald Martens, and Herman Wold. 1983. “The Multivariate Calibration Problem in Chemistry Solved by the PLS Method.” In *Proceedings of the Conference on Matrix Pencils*, edited by A.B. Ruhe and Kastrom, 286–93. Springer Verlag, Heidelberg.