The Saga of PLS

Gaston Sanchez

March 2015

Contents

Chapter 1

Introduction

According to the dictionary a saga is “a medieval Icelandic or Old Norse prose narrative of achievements and events in the history of a personage, family, etc.” A saga is also any narrative or legend of heroic exploits.

I’m using the word Saga to play with its meaning and convey the ideas of history, story, and narrative of events (in this case) associated to Partial Least Squares methods.

This work is about the history behind Partial Least Squares methods. 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. The main motivating trigger behind this ebook has been my insatiable desire to understand the historical development of PLS in order to find the who’s, why’s, what’s, when’s, and how’s.

I’ve written the text without having a particular reader in my 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. However, I’m convinced that this book can be enjoyed by anyone interested in PLS, regardless of whether you’re just starting on this topic, or if you’ve already traveled a long way around PLS territory.

This is NOT a technical book. It doesn’t cover theory, methodological aspects, nor technical details of how the various PLS methods work (no math, algebra, computation, interpretation, etc). Instead, this book can be used as a companion reading for any course, workshop, or seminar about PLS methods.

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, debunking myths, and clarifying 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 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 Wold (1982), 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 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, 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 of the various options in which the algorithm can be specified. Depending on the 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 tweaked the algorithm in a such a way that we are certain of what's going on. 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: The content of this text is divided in three major parts:

1. Introductory Review
2. Historical Narrative
3. Overall Remarks

Part I covers a general introduction describing general aspects about Partial Least Squares methods. Part II 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. Instead of using a writing style listing various events and polluting the text with dozens of citations, I've preferred to use a more fluid narrative style. This narrative is based on a number of references that directly or indirectly talk about the lifes of Herman Wold, Svante Wold, and the development and evolution of PLS methods. Likewise, 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). 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

Many thanks to Michel Tenenhaus for being so gentle to answer to my many emails full of questions and inquiries. *Merci beaucoup*.

Equally important has been the contribution of Wynne Chin, who has also patiently answered to all my questions, providing responses full of details and milimetric precision. *Thanks a lot*.

I would also like to *ringraziare* Vincenzo Esposito Vinzi. Despite his super busy agenda, Vincenzo was able to dedicate me a couple of skype meetings, answering my questions, and providing valuable feedback.

Likewise, I would like to thank Christian Ringle and Jörg Henseler for responding to my emails and questions, and for sharing several of their papers.

Also, many thanks to my various friends, colleagues, and, for the lack of a better term, *followers* around the world for 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

1. 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 (pronounced *Tenen-os*). 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 the 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 beginner. Luckily, *google translate* was already available at that time, providing a considerable help in deciphering some of the book's content.

It was 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. He introduced the technique with an example, and then he went 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 are the PLS regressions in the PLS-PM algorithm? I raised my hand and asked Tomàs about such regressions. 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."

All I was 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, 2008. 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 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. Why there was almost nothing about PLS Path Modeling?

One of the very first things I knew I wanted to do was 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 and I was happy to find that the Economics and Math libraries of the University of Barcelona had copies of Lohmöller's book, Wold's System Under Indirect Observation, The Making of Statisticians (Wold autobiography), and Encyclopedia of Statistical Sciences with the entry of PLS (vol 6, page xxx).

All I wanted was to be able to answer the *who*, *when*, *how*, and *why*. Who created PLS? When was it developed? How it was the development process? And why? How does something such as PLS methods is developed? What are the circumstances that give birth to that framework? In this sense, the quest becomes a quest not only about the man, but also about his mind, his ideas, and surrounding context. I always tell my students that statistical methods are not created out of the blue. They don't appear spontaneously or miraculously—even though many text books and the way we teach, may give that false impression.

Little by little I started to dig and uncover fragments of the history. But one question was always in the back burner of my head: why did PLS Path Modeling is not that well known? Think about it. By 2005, the first works about PLS had been already present for almost 40 years! Yet, having access to those references was ridiculously difficult and scarce. NIPALS-PCA and PLS regressions seemed to be closely connected, and they tend to appear together in the chemometrics literature. However, PLS Path Modeling wasn't mentioned whatsoever. It was a different beast, seemed to be made from a different material. Why?

Chapter 3

2. 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 methods 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
- the graphical possibilities they open for data visualization

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, the ability it gives us to connect a number of seemingly unrelated methods. Such a platform, encompassing both its computational side, and its flexible analytical spirit, 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. There is an amazing variety of methods and approaches, so diverse and plural, that it seems practically impossible to view them 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.

Path Modelism and Regressionism

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.

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.

Some words of caution

Algorithmic approaches

Despite the lack of a unique source, there are some common elements which help tight everything up together. 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.

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.

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.

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 Wold1966a, Wold1966b, WoldLyttkens1969. 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 PLS methods 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 switching 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 the 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.

Chapter 4

3. Timeline

Today PLS methods are well recognized. Symposiums exclusively dedicated to discuss advancements and state-of-the-art. Especial publications such as the PLS (Esposito et al, 2010), PLS New Advancements (Abdi, 2013). On the computational side a wide range of software programs are available both commercially and free. A dozen R packages, plug-ins for MS Excel, Smart-PLS, Umetrics, Matlab, Python, etc. Also important is the amazing amount of learning resources with tutorials, slides, youtube 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.

Just ten years ago, in 2005—when I was a grad student—, if you did an online search about PLS, the only retrieved results were about PLS Regression. 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. Its 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 never saw applications and interest before ending up in the archives of some dusty libraries.

Before presenting the historical narrative, I would like to point out some of the historical highlights that have paved the way for today's place of PLS methods. This brief timeline features events that have guided the path traveled until our current days.

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 to solve a series of ad-hoc data analysis problems (Wold 1966). Under the name **NILES**, short for "Nonlinear Iterative Least Squares," Herman Wold discussed a set of methods solved by means of iterative procedures based on steps of least squares regressions. Among the ad-hoc treated problems, NILES were used to perform Principal Components Analysis (Wold1966a) and Canonical Correlation Analysis.

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, WoldLyttkens1969 replaced the term "NILES" by "NIPALS" (Nonlinear Itearive PArtial Least Squares), consequently shifting from NILES procedures to NIPALS procedures.

Because the 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 will use the term NIPALS-PCA for this special case.

The truth is 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 Wold1973, p.387 "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.

Historically, the first type of PLS algorithm was a power method Wold 1966b for computing Principal Components. This procedure was almost immediately extended to a series of procedures among which there was a version to calculate Canonical Correlations Wold1966a. These first precursors were introduced in the mid-1960s under the name of "NILES procedures," and they changed to "NIPALS procedures," at the end of the 1960s decade WoldLyttkens1969. 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.

Early 1970s: NIPALS Modeling

In the early 1970s, the so-called NIPALS procedures experienced a wave of modifications. The most impacting factor was the inclusion of (at the time recent) simple path models with latent variables (two blocks). In 1973 Wold changed again the label for his methods: from "NIPALS procedures" to "NIPALS modeling," clearly reflecting a more mature—but still incomplete—modeling framework Wold1973.

NIPALS and Basic Path Models

The origins of the PLS Path Modeling branch date back to the early 1970s soon after the breakthrough synthesis resulting from the merging of econometric simultaneous equations models, psychometric latent variable models, sociology causal analysis, and biometric path analysis. Inspired by the modeling synthesis work achieved by Karl Joöreskog—one of Herman's former PhD students—Wold started to brew what he initially called *NIPALS modeling*, a more formal framework based on his NIPALS procedures.

Seeing and seizing the opportunity to develop his own modeling approach, Wold's team refined and polished different versions of a general methodology for estimating path models with observed and unobserved variables. Throughout the 1970s, Herman Wold and his group developed PLS Path Modeling as a methodological framework to handle different types of socio-econometric models that could be estimated by applying iterative algorithms of least squares regressions. 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 (Wold1980a,Wold1980b,Wold1982b,Wold1985a).

Mid 1970s: NIPALS Soft Modeling

The mid-1970s is a time of extensive work and exprimentation. In first place there is the extension of the algorithms from handling two blocks to handling three blocks; secondly there is the extension of handling one between-block relation to more than one between-block relation (Wold1974,Wold1975a,Wold1975b).

It is in the second half of the 1970s when Herman Wold introduces the notion of "Soft Modeling" Wold1975a. This notion is more of ideological nature and will also evolve with time. 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" Wold1975a.

In the late 1970s, after a long simmering and cooking, Wold and his team arrive to a more defined framework. The set of algorithms have been polished and refined. Finally, the acronym "NIPALS" is shortened by "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.

Late 1970s: PLS Soft Modeling

The end of the 1970s decade sees the official presentation of the so-called **Basic Design** for PLS path models. This is the what can be considered to be the *stable* version and further discussed in (Wold1980a,Wold1982a,Wold1985a). This is the basic algorithm for PLS Path Analysis with Latent Variables, and it is the algorithm on top of which all extensions and modifications are based on. Paraphrasing in marketing terms, the basic design is Wold's *minimal viable product*.

Late 1980s: Theoretical Empiricism

In the last of his works, the book "Theoretical Empiricism: A Rationale for Scientific Model-Building", edited by Herman Wold1989b, 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" Wold1989a 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.

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 dub the "basic design", wrapped around the notion of "soft modeling." H-Wold's "PLS Soft Modeling" represents his culminating work on its own terms, in all of its conceptual, intellectual, philosophical, scandinavian glory. His version is surrounded by Wold's aura of scientific causal model-building in social sciences that he always wrote about.

H-Wold envisioned a modeling approach for analyzing systems of linear relations with observed and unobserved variables. His methodology appeared almost simultaneously to the LISREL models of Karl Jöreskog. Both PLS and LISREL-based 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. It is inevitable to compare both approaches. One of the classical references for PLS is a two volume jointly edited by Jöreskog and H-Wold, in which both approaches are compared.

"Soft Modeling is the name for the methodology for PLS estimation of path models with latent variables indirectly observed by multiple indicators" (Wold, 1985). There is not only a difference in estimation procedure for both approaches (PLS, vs, ML), there is also a major ideological difference, heavily influenced in his posture regarding econometric simultaneous equation models (H-Wold had a hard time accepting simultaneity relations). Most contemporary works ignore this fact.

Herman Wold never used the term "Structural Equation Models" in his works. Instead he used a variety of terms such as:

- Quantitative systems analysis (1983)
- Systems analysis by Partial Least Squares
- Path models with latent variables
- Systems under indirect observation using PLS (1985)

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 version appeared in WoldMartensWold1983. This is the begining 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 WoldMartensWold1983.

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

Herman Wold

Herman Ole Andreas Wold was born on December 25th, 1908, as the sixth child of Edvard and Betsy Wold. His birthplace was the small town of Skien, the administrative center of Telemark county, 133 kilometers (82 miles) south of Oslo in Norway. 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 go to 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 so deep that 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, 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. In 1938 he received his doctorate with his thesis *A Study in the Analysis of Stationary Time Series*. For his dissertation, Wold did his research on stationary stochastic processes using the principle of Least Squares 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. 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_{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 $a_1, \dots, a_r$ are obtained by least squares regression of x_t onto $x_{t-1}, \dots, x_{t-r}$. 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." 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.

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. 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. In 1940, Herman Wold and Lars Juréen wrote a booklet with material that would later be used for a textbook on econometrics *Demand Analysis: A study in econometrics* for 1953—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.

Chapter 6

Simultaneity or Recursivity?

At the same time around the early 1940s, many other mathematicians, statisticians, and economists were working on econometric models. 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 figure 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.

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 in introducing 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:

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

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*.

The Stockholm school, strongly based on the works of Dutch economist Jan Tinbergen—Economics Nobel in 1969—thought of an economic system not in terms of simultaneity, but in terms of time-lagged periods, better known as _causal chain} systems:

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

Herman Wold had been formed under the Stockholm school tradition, in which econometric models included an *ex ante—ex post* component. In other words, 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. This type of model—dubed *causal chain* by Tinbergen—can be better described with a graphical illustration using what are known as *arrow schemes*: IMAGE!!!

"The income in one period determines the consumption in the following period and perhaps the investment two periods ahead. The consumption, investment and income are determined in a process with different lags and consequently there will normally be some sort of equilibrium," but not a simultaneity relation.

Before Haavelmo's contributions, the errors ϵ_t and δ_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. In addition, econometricians used to estimate similar models using a series of different approaches among which there was the method of Least Squares, the tool used by Tinbergen and Wold. With Haavelmo's approach, his advice was to avoid using Ordinary Least Squares, since this method provided inconsistent results when applied to simultaneous equation models. Instead, he favored and promoted the use of Maximum Likelihood.

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.

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. Herman was a prominent figure in favor or recursivity, a stance that he would maintain for the rest of his life. However, it became Haavelmo's point of view the one that became dominating. The Maximum Likelihood convenience and elegance offered the mechanism of hypothesis testing, too tempting and irresistible to let it go.

Observational and Experimental Data

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. "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 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 192X, 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.

Expanding Interests

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 *latentness*. 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.

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 quatitative analysis among social sciences disciplines was enormous after the works of Duncan and colleagues (REF!!!), 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.

Chapter 7

A Causality Crisis

Furthermore, the topics of *causality* and *model-building* were among his 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, researchers needed to be able to apply a scientific method of proposing hypotheses and test them.

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 it was 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."

"There was the question of causal interpretation of OLS regressions", Wold recalled, although "causal arguments were still practically taboo in the socioeconomic 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*. 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."

Causality and Recursivity

"A general category of theoretical models. This category, which is of fundamental importance, will be referred to as recursive systems with properties:"

- they are recursive in the twofold sense already indicated
- each equation of the system expresses a unilateral causal dependence

A system is recursive: (a) if the development of the three variables is known up to time $t-1$, the system gives us the variables at time t; (b) the variables at time t are obtained one by one. The second point is that each equation of (1) may be interpreted as a hypothesis of unilateral causal dependence.

A Causality Crisis

"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."

The problem to Wold was that of the well known *correlation does not imply causation*. "The trouble begins," Wold observed, "when we formulate 'laws' of causality, for example, 'every event is causally related to something else' "

“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 $y = f(x_1, \dots, x_p)$ is causal “if it is theoretically permissible to regard the variables as involved in a fictive controlled experiment with $x_1, \dots, x_p$ 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.”

“The logical interpretation of causal relationships is the same in experimental and nonexperimental situations. From a statistical point of view, on the other hand, there are deep-going differences between the two cases.”

“Regression relations may serve purposes of explanation or of description... Generally speaking a regression relation may or may not involve a causal hypothesis.”

“If the regression relation is based upon a causal hypothesis the relationship not only can be used as before to predict y for given $x_1, \dots, x_p$, but can now also be used for other purposes, just as a genuine causal relation.” (Wold 1955, pp 167)

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 on one of the research topics that had lately been absorbing great part of his energy—econometric models based on 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."

Chapter 8

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 avoiding 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 turmoil, it was his blind faith on Least Squares, which he was not willing to abandon. 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:

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

then:

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

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* (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. 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.

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. Using Least Squares and his theoretical argument of *predictor specification*, he soon came up with a new approach called the **Fix Point** (FP) method. Roughly speaking, the FP method consisted in an iterative procedure of least squares regressions for causal chain systems. More important, the Fix-Point method seemed to offered good promises for solving the thorny problem around interdependent systems.

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 PCA and CCA. 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. Still, 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.

Chapter 9

Bibliography

Wold (1966) NIPALS procedures

Wold (1973) NIPALS Modeling

Wold, H., Lyttkens, E. (1969) First Partial Least Squares

Wold, Herman. 1966. "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.