

THREE ESSAYS ON STATISTICAL METHODS FOR COMPLEXITY ECONOMICS

by

John S. Schuler
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Economics

Committee:

_____	Director

_____	Department Chairperson
_____	Program Director
_____	Dean, College of Humanities and Social Sciences
Date: _____	Summer Semester 2020 George Mason University Fairfax, VA

Three Essays on Statistical Methods for Complexity Economics

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy at George Mason University

By

John S. Schuler
Master of Arts
George Mason University, 2017
Master of Science
American University, 2017
Bachelor of Arts
St. John's College, 2009

Director: Richard E. Wagner, Professor
Department of Economics

Summer Semester 2020
George Mason University
Fairfax, VA

Copyright © 2020 by John S. Schuler
All Rights Reserved

Dedication

I dedicate this dissertation to my parents.

Acknowledgments

My first acknowledgements are due to my parents who have always supported me notwithstanding my often idiosyncratic choices. Warm thanks are due to the many scholars in several schools and departments who have supported me over my long graduate education building to this: Robert Axtell for his advice and tolerance of endlessly shifting research priorities, Richard Wagner for first driving home the immensity of the task of understanding an entangled economy, Carlos Ramirez for his low threshold for entertaining an idea and high threshold for its acceptance, and David Levy for his interesting suggestions. Garrett Jones, while not on my committee, was generous with advice and encouragement. Acknowledgement is also due to my American University professors, Michael Baron and Stephen Casey, for helping me use that program as a jumping off point for my subsequent studies. Finally, I would like to acknowledge my current and former PhD program colleagues for their friendship and willingness to entertain ideas: Andrew Humphries, Steven Monaghan, and Jon Murphy.

Table of Contents

	Page
List of Tables	vii
List of Figures	viii
Abstract	ix
Introduction	1
1 The Econometrics of Prices in a Network Economy	2
1.1 Introduction	2
1.2 Literature Review	3
1.3 A Simple Economic Model	4
1.4 The Data	6
1.5 The Statistical Problem	6
1.5.1 An Empirical Test of Assumptions	7
1.5.2 The Statistical Power of the Test	8
1.6 An Alternative Model	11
1.7 A Simulation Model	13
1.8 The Statistical Validity of the Model	17
1.9 Conclusion	20
2 Statistical Learning for Non-Parametric Copula Estimation	22
2.1 Technical Preliminaries	23
2.1.1 Copulas	23
2.1.2 Sklar's Theorem	23
2.1.3 Advantages of Copula Models	24
2.2 The Model	25
2.2.1 Model Preliminaries	25
2.2.2 Model Statement	26
2.2.3 Estimation	27
2.2.4 Derivations from the Model	28
2.3 Modeling Pulp Price Indices	28
2.3.1 The Data	28
2.3.2 Estimating the Univariate Marginals	29

2.3.3	Model Performance	29
2.4	Conclusion	29
3	Recursion, Praxeology, and the Redemption of Econometrics	32
3.1	Introduction	32
3.2	Klein and the Transition from Ancient to Modern Science	35
3.3	Notions of Probability	39
3.4	Ragnar Frisch and the Vision of Econometrics	41
3.5	Tryg Haavelmoe and the Structure of Econometric Counterfactuals	43
3.6	Judea Pearl and Haavelmoe's Idea of Causality	44
3.7	History, Praxeology, and Fixed Points	47
3.8	The Place of Such a Science in the History of Science	48
3.9	Conclusion: Recursive vs Logical Atomism	48
4	Appendix 1	50
5	Appendix 2	85
6	Appendix 3	93
7	Appendix 4	144
8	Appendix 5	186
	Bibliography	cxc

List of Tables

Table	Page
4.1 Network Node In-Degrees	50
4.2 Network Node Out-Degrees	67

List of Figures

Figure	Page
1.1 Distance Correlation of VAR Residuals (significant at ≤ 0.01)	8
1.2 Network Diagram of the Adjacency Matrix	9
1.3 Partial Distance Correlation of Bootstrapped Residuals	10
1.4 Histogram of In-Degrees	13
1.5 Histogram of Out-Degrees	13
1.6 Simulated Coefficient of Variation	17
1.7 Two Pulp Prices	18
1.8 Sampling Distributions of Two Pulp Prices	19
1.9 All Pulp Prices	21
2.1 In-Sample Fit	30
2.2 Out-of-Sample Fit	30

Abstract

THREE ESSAYS ON STATISTICAL METHODS FOR COMPLEXITY ECONOMICS

John S. Schuler, PhD

George Mason University, 2020

Dissertation Director: Richard E. Wagner

The research program commenced in this dissertation has three pillars: economics, complexity science, and statistics. Notwithstanding the frequency with which the three interact, they have very different foundations and thus are often in methodological tension. It is the goal of this research program to study concrete problems at the intersection of these fields with the goal of both their partial solution and also a better explication of their foundations.

Economics is, above all, the science of interdependence. The “division of labor is limited by the extent of the market” as Adam Smith famously said. Even in a simplified general equilibrium setting, a change in the extent of the market alters the situation immensely. Statistics, on the other hand, is the science of independence. Even exchangeability, a concept basic to practically all of statistics, is at its core, conditional independence; which is to say, structured independence. Complexity science provides the framework that can bridge this gap. All scientific theories ultimately involve the arrangement of facts within a set of counterfactuals; distinguishing realizations, or data, from other possible realizations and also those regarded to be impossible. Complex systems display a history. Different historical trajectories of complex systems then, can be considered independent. In order

to build an economic science that is aware of complexity then, the scope of existing economic models must be broadened.

This is where the basic tension comes to the fore. Statistics does have a tool set for dealing with data that has history through the use of exchangeability. What is needed is a richer class of models such that the agent-based concepts map onto the statistical constructs easily with explicit modeling of the counterfactuals. For this sort of work, more flexible statistical methods are necessary. Leo Brieman famously distinguished “two cultures” in data analysis: traditional statistical data modeling and predictive modeling. Traditional statistical modeling allows for ease of interpretation while statistical learning is optimized for prediction. It seems though that a good data model ought to be good at predicting as well. Thus, a logical line of inquiry is whether the two cultures may be bridged. I believe the answer is yes.

The key is attempt to use statistical learning techniques to estimate not the data itself but rather a model that could have generated it. Thus, the outcome of such a modeling exercise is *not* a series of point estimates but rather a joint probability distribution that allows for simulation. The strength of such an approach is that all the traditional econometric tools defined in terms of conditional moments may be easily adapted to this more flexible modeling framework. An additional advantage is that the concepts of Judea Pearl’s *Causal Inference* map very easily onto such a model as they are expressed in terms of Bayesian networks and conditional independence.

These advantages together allow such methods to be powerful tools for the statistical emulation of agent-based models. If we have shown that such a model works well as a reduction of a more complex agent-based model, we can describe the properties of the agent-based model in terms of the joint probability distribution or the Bayesian network. This also makes calibration of the agent-based model to data a more straightforward task. To this end, this dissertation offers three chapters: first, I attempt to adapt traditional time series methods to the study of the Cantillon Effect and articulate the several reasons why this does not work well. In the second chapter, I offer a possible solution to this problem:

a novel statistical method of the sort described above. Finally, I argue that a bottom up / agent-based economics enhanced with Austrian ideas, is actually more faithful to Ragnar Frisch's original vision for econometrics than the developmental path that it actually took.

Of course, a great deal remains to be done. This dissertation raises many more questions than it solves. Of particular interest is the ontic and epistemic status of so-called "heavy-tailed" distributions in economics. The first chapter demonstrates that, as was well-known to Benoit Mandelbrot, price changes exhibit heavy-tailed distributions, rendering traditional regression techniques unreliable for the study of price dynamics. An open question is to what extent these heavy-tailed distributions bear on the Hayekian notion of spontaneous order. Clearly prices convey information and yet, this information does not wish to yield traditional statistical treatments. An interesting question for further research will be whether complexity science can give an account of this.

A final theme of this dissertation is the universality of recursion as a basic intellectual construct. Much of traditional science was based instead on classical set theoretic mathematics. If we wish economics to be a process-oriented science, then perhaps the modeling framework ought to proceed according to procedural rather than class-based notions of abstractions. Thus, constructive mathematics and theoretical computer science are the mathematical sciences to which we must appeal.

Introduction

Economics has long been criticized for both the strength of its assumptions and their lack of realism. Their status as tractability assumptions accounts for their staying power but these criticisms will persist until they are weakened. Common assumptions in economic models are differentiability / smoothness, linearity, equal probability of interaction and equilibrium. These assumptions are likely impossible to relax within classical mathematical economics. A new approach is needed. This dissertation is a contribution to the development of *complexity economics*, an emerging paradigm with foundations in constant flux. In spite of this flux, certain themes reliably come up: non-linearity, far from equilibrium systems, and interactions through networks of various kinds. These properties generalize standard assumptions; and in many cases are precisely those required to generate rich economic behavior such as Cantillon effects and realistic market movements.

New mathematical, computational and statistical methods are needed to build out this complexity economics. The following chapters focus on relaxing the assumptions of equilibrium and equal probability of interaction and showing that classical statistical methods fail. An alternative is then offered based on newer techniques.

Chapter 1: The Econometrics of Prices in a Network Economy

1.1 Introduction

The quantity theory of money $MY = MV$ predicts that a fall in circulating money will cause prices to fall. As a long run matter, there is very little dispute about this. There is, however, some evidence of so-called *Cantillon* effects where some prices move, in the short run, in the “wrong” direction and further, that in the aftermath of a monetary shock, the overall variation among prices increases. This is of interest because disruption in relative prices may reduce the efficiency of economic activity. There are several questions:

1. What is the evidence that Cantillon effects exist?
2. If they exist, what is a mechanism?
3. Is there evidence for this mechanism?
4. Are the statistical methods used to answer these questions appropriate to both the purpose and the data?

[1] provides a theoretically plausible mechanism. Our goal is to study the extent to which this model agrees with the data. This paper has nine sections. We begin with a literature review to determine what the literature has to say about these questions. Following that is a description of the data. We then state the econometric problem and offer two attempted solutions to it. The first of these attempts to estimate directed network information. The second attempts to attack the multivariate dependency. Finally, there is a conclusion.

1.2 Literature Review

So what is the empirical evidence of Cantillon effects? Most importantly for our purposes is [2] which gives evidence for both of the effects. They do this with a vector autoregression following ideas in [3]. According to [2], some 50% of prices change in the “wrong” direction and monetary shocks are followed with an increase in the coefficient of variation. [1] gives an agent-based model that produces both of these effects. So far, it would seem we have no issue. However, a closer look reveals some difficulties.

At the core of [1] is the idea that money circulates through an economic network of suppliers. There is then a path dependence of the effects of that money. [2] must deal with a challenging statistical problem. A market price vector is a very high dimensional system. Further, these time series have many gaps and these gaps are themselves not statistically independent. As always, simplifying assumptions must be made. In brief, [2] assumes that having conditioned the price series on macroeconomic or aggregate variables assumed to be important, the individual price series are orthogonal and many be estimated using univariate methods and then stacked; following a method introduced in [3]. They defend this decision as a standard assumption for macroeconomic models, ie. the causation runs from the aggregate to the individual but not the other way. Also relevant is [4] which makes a similar assumption. Indeed, this assumption may be reasonable in many situations. In this paper, I aim to show that here, it is questionable on both theoretical and empirical grounds and further, if the network model of [1] is true, the simplifying assumption cannot well-approximate reality. In other words, the theoretical model is at loggerheads with its empirical basis. Thus, the question is whether we can render this statistical problem tractable and examine the reasonability of the theoretical model in light of this alternative statistical model.

This paper draws from several strands of literature. Firstly, there is the Cantillon effect literature just mentioned, both theoretical and empirical. It also bears a relationship to the developing use of statistical learning in economics as I find that traditional regression

methods are ill-suited to this purpose. See [5], [6], [7] for examples.

1.3 A Simple Economic Model

In *Human Action*, von Mises introduced the idea of an evenly rotating economy as an heuristic device [8, p 246]. It will prove useful here. In an evenly-rotating economy, technology and population are fixed and there is no growth. Thus, relative prices \vec{P} are fixed. If we allow money to vary in such an economy, then we also have a concept of a price level Π and thus, nominal prices are $\Pi\vec{P}$ for positive P_i . As is traditional, we represent the logarithm of a value with a lower case letter. Thus, $p_i = \log P_i$. Thus, in an economy of n commodities we have n prices where

$$p_{i,t} = \pi_t + r_i$$

where r_i is the real price and does not depend on time. Now, since output and velocity are fixed: $\pi_t = m_t - y + v$. Therefore, we can write the series as:

$$p_{i,t} = r_i + m_t - y + v.$$

By assumptions, the nominal price is a function of the real price and the relevant macroeconomic variables. In such a situation, any observed variation in prices is due either to a failure to distinguish subtly different goods, for example, grocery store milk available most of the day vs convenience store milk available all the time at a premium or the simple fact that some price differentials are smaller than the cost of arbitrage and thus, even in equilibrium, we do not expect the law of one price to hold exactly. In any case, it is obvious that in this case, it is quite reasonable to assume as [2] does that controlling for the macroeconomic variables is sufficient to render the error term orthogonal. Obviously, this is a toy model. It is introduced as a basis on which to build.

We now, consider a slightly more complex economy. We add growth and allow velocity

to vary. All this does is render the equations:

$$p_{i,t} = r_i + m_t - y_t + v_t.$$

The same logic applies. The macroeconomic variables and real prices contain all relevant information. We can even allow technology to change, resulting in changes of real prices:

$$p_{i,t} = r_{i,t} + m_t - y_t + v_t.$$

This is the simplest model that requires something not directly observable. Now, consider the first difference:

$$\Delta p_{i,t} = \Delta r_{i,t} + \Delta m_t - \Delta y_t + \Delta v_t.$$

It is true that changes in real prices must be correlated, sometimes positively and sometimes negatively. Nonetheless, so long as real prices do not change much or change quickly, the term $\Delta r_{i,t}$ is going to account for a small portion of the variance and thus any covariances it induces will remain small. Thus, we come to the core assumption: *so long as the economy is in equilibrium, the orthogonality assumption should be expected to hold approximately.* This means that we expect it to hold best for “medium term” time series which we define circularly as those that sample with a lag longer than the economy takes to equilibrate but shorter than which there are changes in real prices. What time frame this is and whether these even exist is a potentially empirical question.

The basic concern should now be evident. Cantillon effects are a *monetary disequilibrium*. If the network hypothesis offered in [1] is true then, even for fixed real prices, output, and velocity, the partial correlation matrix of current prices, regressed on their lags will differ, perhaps substantially, from the proposed identity matrix. Therefore, if the model is true, the empirical inspiration for the model is suspect. This is particularly problematic because the point is made on the basis of coefficient of variation. The empirical coefficient of variation is defined as: $\frac{1}{\bar{p}_t(n-1)} \sqrt{\sum_{i=1}^n (p_{i,t} - \bar{p}_t)^2}$ Recall that in general, where ϵ_i is the

error term for p_i , $\text{Var} [\sum_{i=1}^n \epsilon_i] = \sum_{i=1}^n \sum_{j=1}^n \text{Cov} [\epsilon_i, \epsilon_j]$. Under the orthogonality assumption, $\text{Cov} [\epsilon_i, \epsilon_j] = 0$ where $i \neq j$. If this is false, there is the potential for the correlations to introduce bias into the estimator of the coefficient of variation and further, this bias may grow with respect to the number of series modeled.

1.4 The Data

We use the producer price series provided by the Bureau of Labor Statistics. We make use of the price series for which we have at least three years of data and which are at least 95% populated. The producer price indices come in varying levels of granularity indicated by the length of the code. We make use only of the most granular 8 digit codes. Also, we use the series that are *not* seasonally adjusted. This leaves us with 379 price series. The BLS also provides an overall price index which we use in the model. These data are all reported monthly. The Federal Reserve also reports M2 monthly. The tricky part is the Federal Funds rate which is reported more than monthly. We use the reported rate closest but prior to the beginning of the month as the rate for that month.

1.5 The Statistical Problem

Fortunately, the question of the dependence structure of the error term is empirical. My tool will be so called “energy statistics” introduced in [9]. In particular, I will use *distance correlation*, a statistic with a support on $(0, 1)$. It has the property that it is zero asymptotically if and only if the two random variables are independent [9]. The statistic is derived from the concept of energy distance which is a true metric between probability distributions. Distance correlation can be used as the basis of a nonparametric T-test analogue for independence [10]. This test relies on the existence only of first moments. Thus, I begin by fitting the univariate models. We use $L_k = \sum_{i=1}^k L^i$ where L is the lag operator.

To ensure stationarity, I take the first difference of the logarithm of the BLS price series.

I fit the following model:

$$(1 - L_{12}) p_{i,t} = (1 + L_{12}) (1 - L_1) (M_t + F_t + \Pi_t)$$

where $p_{i,t}$ is the logarithm of the price index i at time t , M_t is the M2 measure of the money supply at t , F_t is the last reported federal funds rate prior to time t , and Π_t is the overall price index at time t . As the purpose of this is heuristic rather than to build the best possible model, I do not consider seasonal effects or attempt to orthogonalize the macroeconomic variables. Thus, this is a somewhat simplified version of the model used in [2] with the exception that they do not take the first difference. I chose to model the first difference because it gave the best performance. A few series failed to converge and were thrown out. Plots of these series showed them to be extremely erratic and thus not easily modeled in this way. These turned out to be merely particularly extreme examples of a broader problem to be discussed later. Given the number of series, it was not possible to test all of them for stationarity but for those tested, it was possible to reject the null hypothesis of integration of order 1 using the Perron-Philips test.

1.5.1 An Empirical Test of Assumptions

I then fit 379 models using univariate time series methods. These models yield residuals. Thus, I calculate statistics on these residuals to test the independence of the corresponding innovation terms and thus the validity of assuming them to be independent and thus uncorrelated. I build what one might call a partial distance correlation matrix using the univariate residuals. For sparsity, I test the significance of each coefficient at the 0.01 level. This matrix is sparse but it is far from identically 0. I set the entries where I fail to reject the null hypothesis of independence to 0. The distance correlogram is pictured in figure 1.1. If one applies the rule that a distance correlation of 0 means disconnected nodes and a positive one means connected nodes, one may think of this as the adjacency matrix of a network and as a very crude statistical reduction of the network of monetary flow. Figure 1.2, displays the corresponding network diagram. While most nodes are not connected to

most other nodes, no two nodes are more than a few hops apart and in this sense, the network is highly connected. This makes the multivariate process very difficult to model.

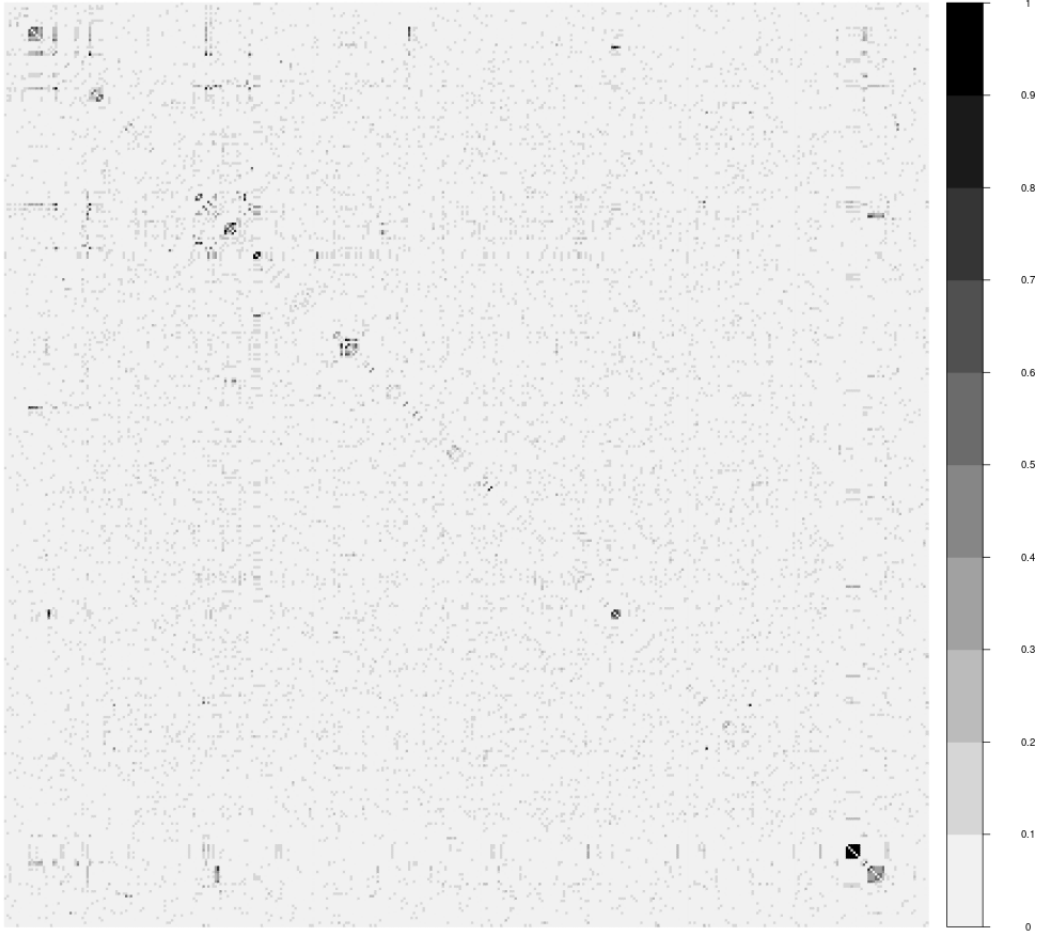


Figure 1.1: Distance Correlation of VAR Residuals (significant at ≤ 0.01)

1.5.2 The Statistical Power of the Test

There is an additional issue to be concerned about. The null hypothesis of the T-test is independence, that is, a distance correlation of 0. The null hypothesis lies at the boundary

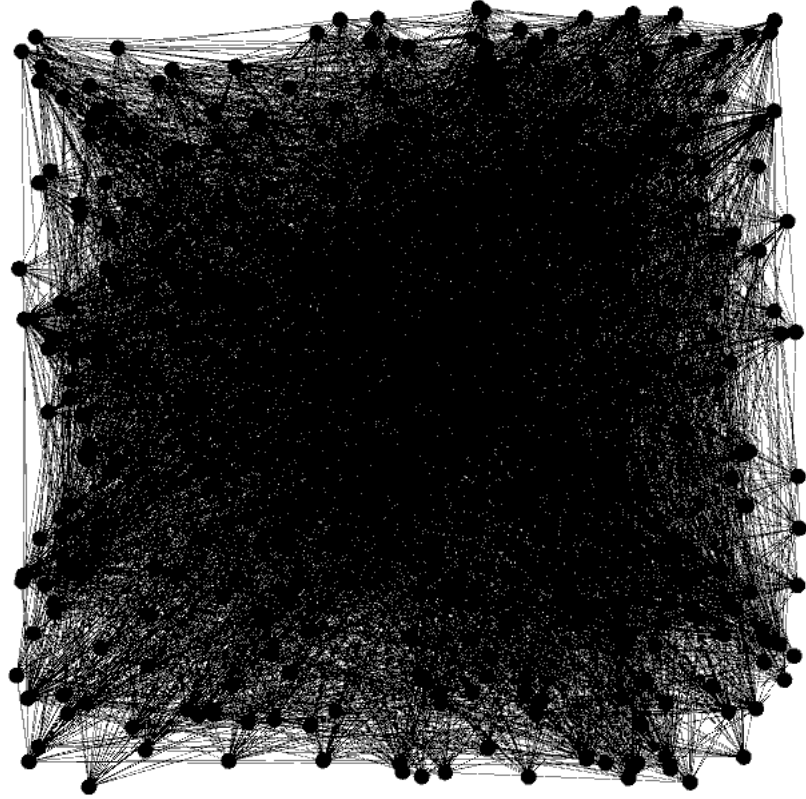


Figure 1.2: Network Diagram of the Adjacency Matrix

of the support of the statistic and thus, the estimator is necessarily biased upward. Is this is driving the rejection of independence and therefore orthogonality. Since the distance correlation is not a parameter of a statistical model, it is not possible to perform explicit power calculations. It is, however, possible to simulate independent time series and repeat the exercise many times using a bootstrapping approach. Bootstrap each series of residuals independently. Thus, the new series will have identical univariate statistical properties but will be independent by construction. If the upward bias in the estimator of distance

correlation is resulting in many type 2 errors at these sample sizes, we expect the resulting partial distance correlation matrix to strongly depart from being identically 0. As expected, while there are a few estimates that pass the rejection threshold, they are only very few as displayed in figure 1.3. Therefore, biased estimators are not driving these results.

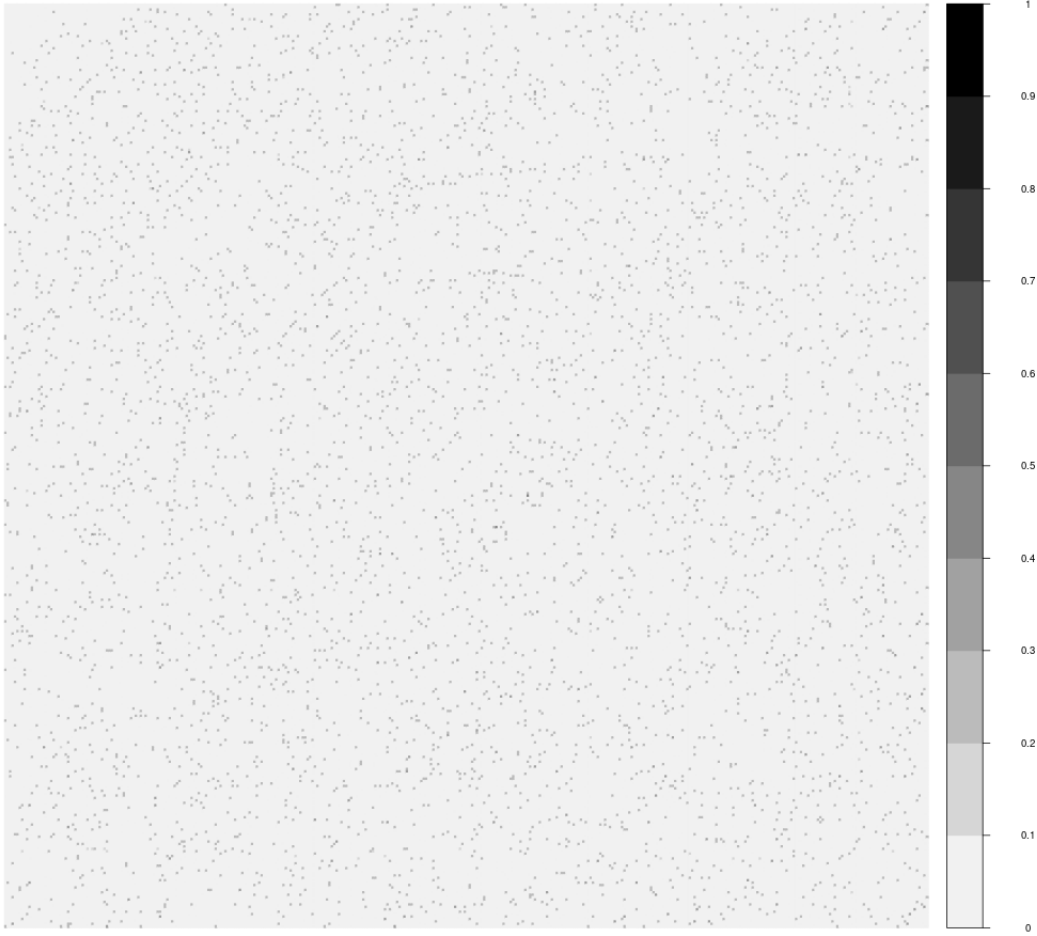


Figure 1.3: Partial Distance Correlation of Bootstrapped Residuals

1.6 An Alternative Model

The distance correlation matrix estimate gives us explicit information on the dependency structure among the various price series. This is still too crude as a model of the supplier network or even a statistical reduction thereof since this network is an undirected graph. Can we possibly model a directed graph? In order for this to work, it is necessary for price dependencies to be able to exist in only one direction. Given the limited granularity of the price indices, it is not obvious that rich directional information is discoverable. This is an empirical question. Again, the idea is to attempt to leverage our independence information to assemble univariate models into multivariate models. Consider the following assumptions:

1. I assume the same basic VAR framework on the first difference as before.
2. Each price is potentially a function of its own lags.
3. Each price is potentially a function of the lags of those prices it depends on but not on the contemporary prices.
4. Each price is potentially a function of contemporary or lagged macroeconomic variables.
5. Finally, once we have controlled for all relevant prices, *then the resulting series are orthogonal*.

There is another statistical problem to solve. Traditional regression models do not perform variable selection and thus, simply regressing a price difference on its own lags and the lags of all price differences on which it depends will result, at best, in a very complex model with many small coefficients. Further, the fact that the series are of different lengths and there is missing data poses a greater challenge. This is where this paper links up with the statistical learning literature. I make use of a LASSO regression to perform variable selection. For a description of the LASSO, see [11]. The basic idea is that I minimize the sum of squared errors subject to a penalty term of some quantity λ multiplied by the sum

of the absolute values of the β 's. The greater the λ parameter, the greater the penalty and the simpler the model. This technique will force many slope parameters to 0. The question is, how do we choose λ ?

It is common in statistical learning to choose λ based on the data and I follow this approach here using a technique called cross-validation. I reserve a portion of the data as a testing set, in this case, 30% of the available data. The remainder of the data is the so called training set used to fit the model. A possible vulnerability of this technique is temporal effects for which I have not accounted. This does lean somewhat heavily on the admittedly strong assumption of multivariate stationarity in arbitrary dimensions. This vulnerability is reduced though insofar as I choose the λ where the model fit from the training data best predicts the validation data. In other words, I take $(1 - L)p_{i,t}$ to depend on its own lags, other lags, or macroeconomic variables, if and only if it improves predictions of $(1 - L)p_{i,t}$. Where it does, its marginal effect is represented by a non-zero β parameter. This has the further advantage of giving us information that is directional and therefore potentially causal. As I have modeled prices on the basis of lags of other prices, I have directional information which is better than before. Further, the dependency information above was merely bivariate. Since the lasso model takes into account all possible dependencies and rejects those that are not useful for prediction, I can address some of these issues. This allows the estimation of statistical reduction of the producer network. In figures 1.4 and 1.5, I display the in and out-degree distributions of the network reduction.

The difficulty with the LASSO model is that it remains a collection of univariate models. It is suitable for multivariate modeling if and only if the final assumption is approximately true. Even a simple covariance matrix shows that it is not. Thus, the question remains: how to build a model that potentially contains the full set of effects when the dimensionality is too high for traditional methods to be tractable?

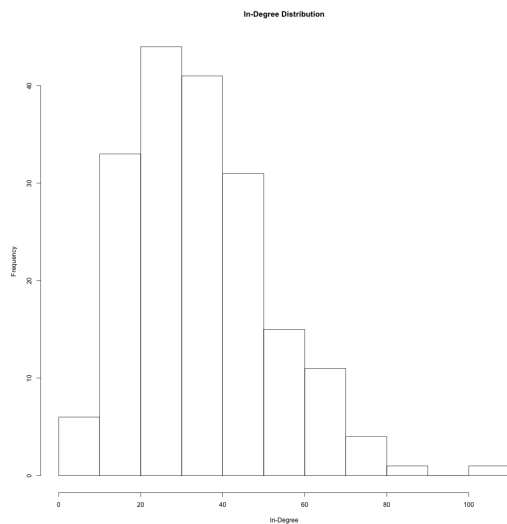


Figure 1.4: Histogram of In-Degrees

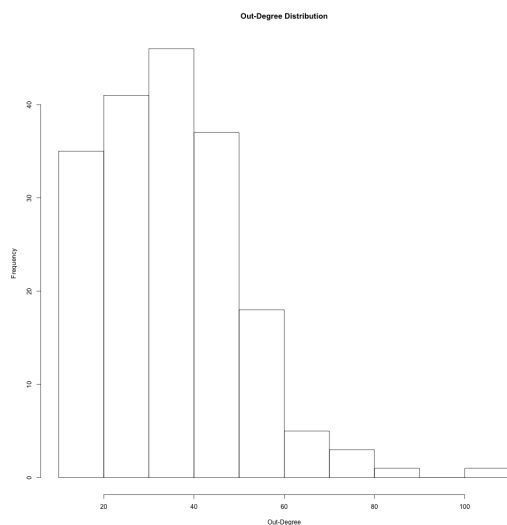


Figure 1.5: Histogram of Out-Degrees

1.7 A Simulation Model

The goal is to simulate a stochastic process and see if shocking the federal funds rate gives us Cantillon effects. For this purpose, I again begin with the collection of univariate VAR

models from the first section. We can now consider this as a vector:

$$(1 - L_{12})(1 - L_1)\vec{p}_t = (1 + L_{12})(1 - L_1)(M_t + \Pi_t + F_t) + \epsilon(M_t, \Pi_t, F_t)$$

There are two terms that need to be modeled in order to simulate the stochastic process: the macroeconomic process of $[M_t, \Pi_t, F_t]$ and the innovation term ϵ . For reasons given above, this innovation term is very far from a white noise process. I also allow it to depend on the macroeconomics variables. Thus, while I model each price only on its own lags and the macroeconomic variables, I also explicitly model how the innovation term tends to move certain ways. A complex innovation term is consistent with a network model. The difficulty I must solve again is that the dimension of this problem is high relative to the data. How can I model the innovation term in a way that respects its complex interdependencies?

One way to proceed is to attempt to take advantage of the network structure rather than be defeated by it. The partial distance correlation matrix on the residuals gives us a dependency structure. I can use this network structure to model the error term while only considering a subset of the dependent dimensions at once. My technique for doing this is quasi-agent-based. This sort of technique may be the basis for something reasonably called agent-based estimation; possibly worth exploring. The idea is to let a small number of agents walk around the dependency network, fit a model for each node depending on a subset of the nodes connected to it and then calculate an estimate. Now, since each node is regressed on a subset of the other nodes, this will introduce bias. The hope is that the randomization the starting nodes at each realization of the innovation process will reduce this bias. The algorithm in more detail is:

1. Select a row at random from the residuals of the VAR models.
2. Choose some k % of the nodes to value with the residuals from the chosen row.
3. Now, randomly assign n agents to nodes adjacent to some of the previously valued nodes.

4. Each agent takes stock of the connected nodes which are valued and fits a model for its current node based on the valued nodes connected to it. The model has the current residual as its y -variable and the current and lagged values of the connected nodes as x -variables *along with the current and lagged values of the macroeconomic variables* where the distance correlation indicates dependence. Thus, we allow the error term to depend directly on macroeconomic conditions and its own lagged values but only when the data supports it.
5. This model is stored in case we want to use it later. This model spits out a prediction for the current node and the node is valued with this.
6. The agents jump to an adjacent unvalued node. If there isn't such a node, they can "teleport" to a disconnected node.
7. When all nodes are valued, terminate the process. This gives a realization of the error term.

These models are fit using *projection pursuit regression*. This technique is potentially a universal estimator. For details, see [11, p 389]. Now that I have a model for the error term, I can focus on modeling the macroeconomic variables. The federal funds rate, when considered as a time series fails tests of stationarity even when differenced many times. This is consistent both with the federal funds rate being considered crudely as an exogenous control variable, and also with the outcome of a complex system whereby it is determined by private agents working in concert with the central bank. In any case, the lack of stationarity is likely due to a variety of institutional conditions through the history of our data. I will *not* attempt to model it as random but rather as simply given. In the simulation, which is 120 time steps, I begin with the lowest recorded federal funds rate and gradually build up to the highest at the final time step while following a roughly sigmoidal function. The step to step differences are actual month to month observed changes in the rate so these changes are realistic. This addresses, in part, some concerns about the non-linearity of Cantillon effects and therefore, the inappropriateness of classical impulse response techniques.

Given the fixed federal funds rate path, I need to determine a reasonable path for M2 and the price index. Now, the joint probability distribution of the macroeconomic variables can be factored.

$$P\left(\vec{M2}, \vec{\Pi} \mid \vec{FF}\right) = P\left(\Pi \mid M2, FF\right) P\left(M2 \mid FF\right)$$

Analogously, I can estimate the joint probability distribution of Π and $M2$ conditional on their lagged values and on current and past values of the federal funds rate by fitting two autoregressive models:

$$M_t = A_M L_{12} M_t + B_M (1 + L_{12}) F_t + \epsilon$$

$$\Pi_t = A_\Pi L_{12} \Pi_t + (1 + L_{12}) M_t + (1 + L_{12}) F_t + \epsilon$$

where M_t is M2, π_t is the BLS price index, and F_t is the federal funds rate and making standard white noise assumptions on the error term. Then, I can simulate the $M2$ process by plugging in 12 actual values of $M2$ along with our hypothetical federal funds path and sampling from the residuals. Given the simulated path for $M2$, the price index may be simulated in a similar manner. Thus, I simulate the joint probability distribution from the conditional marginals. Now, I have a representation of both the multivariate innovation term and the path of the macroeconomic variables. As the federal funds rate is increasing by construction, I can simulate the stochastic process of the prices and see how their coefficient of variation responds to an increase in the federal funds rate.

This simulation is consistent with the existence of Cantillon effects though the large jump suggests the model could use some additional calibration. It is also interesting that the coefficient of variation actually flattens out at higher federal funds rates. It is important to note that the simulation is of differenced values which are then concerted back into levels by simply summing them. It is also interesting that this outcome is consistent with Cantillon effects that are non-linear with respect to the federal funds rate. How reliable this is remains a question of model calibration. It does seem that whatever Cantillon effects that exist are

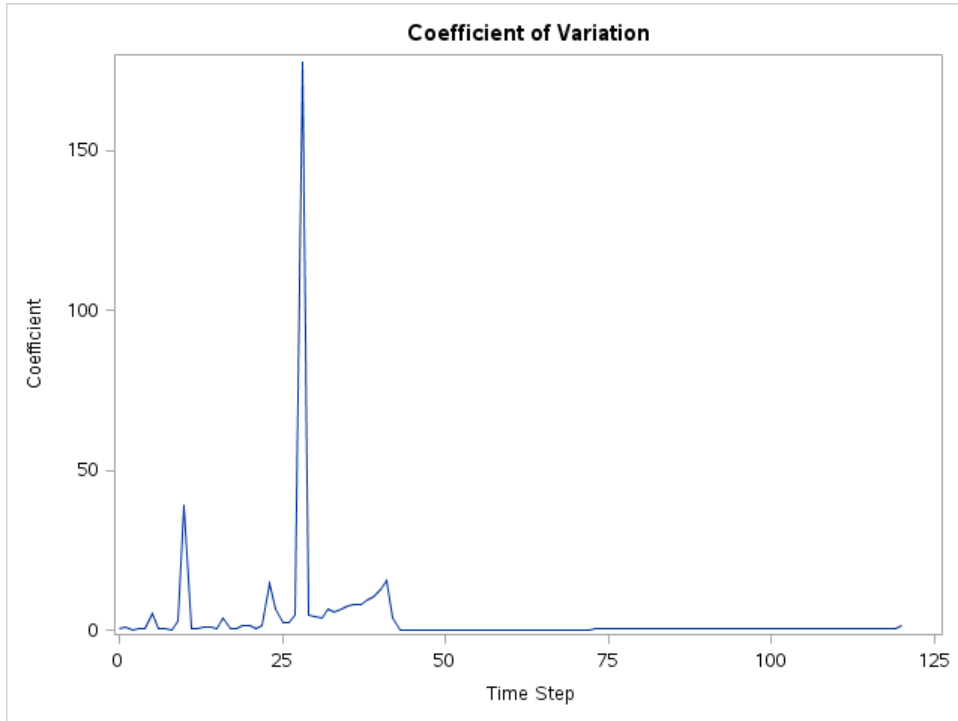


Figure 1.6: Simulated Coefficient of Variation

usually subtle. A potential avenue for future research is similar analysis done in countries that have recently experienced rapid inflation.

1.8 The Statistical Validity of the Model

Both multiple time series methods and projection pursuit regression rely on the efficient estimation of covariances and therefore cross-moments. Thus, it is necessary to ask whether these very common statistical assumptions hold. As it turns out, they do not. Using the BLS pulp price indices as an example, consider two of them and examine their histograms, the scatter plot and the histogram of the raw second moment. These are plotted in 1.7. Like a correlogram, the diagonal contains the histograms. The off-diagonal histogram plots the raw cross-moment. Thus, commodity prices too display heavy tails. Heavy-tailed

phenomena has been noted in a financial context in [12]. To drive the point home better, consider a bootstrapping to estimate the distribution of means. An analogous plot is given in 1.8. Now, the diagonal histograms may be interpreted as sampling distributions of the mean. The off-diagonal histogram is the sampling distribution of the raw cross moment and the plot is the sampling distribution of the bivariate mean.

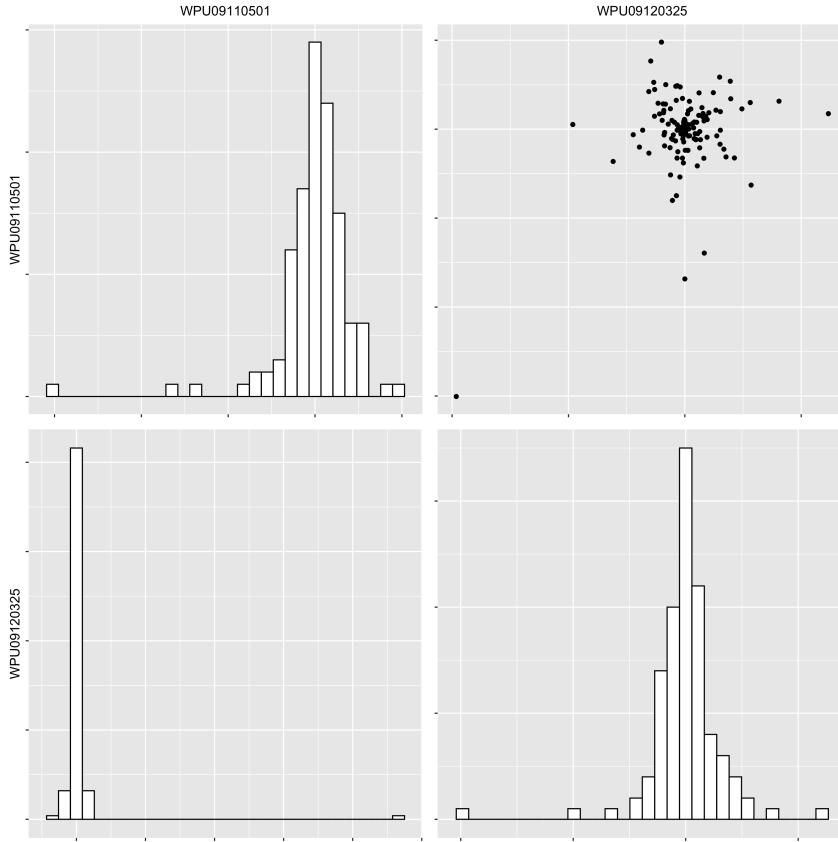


Figure 1.7: Two Pulp Prices

The raw cross moment has several modes which suggests that our estimators of covariances will be highly unstable. Even the first moment displays great variation considering the sample size. An easy way to see the extent to which the tails are super-Gaussian is to define the *deleted sample mean* $D_i(\vec{x}) = \frac{1}{n} \sum_{j \neq i}^n x_j$. Now, let $X_{(i)}$ represent the i th order statistic of a given variable. Standardize the pulp price log difference and plot its deleted

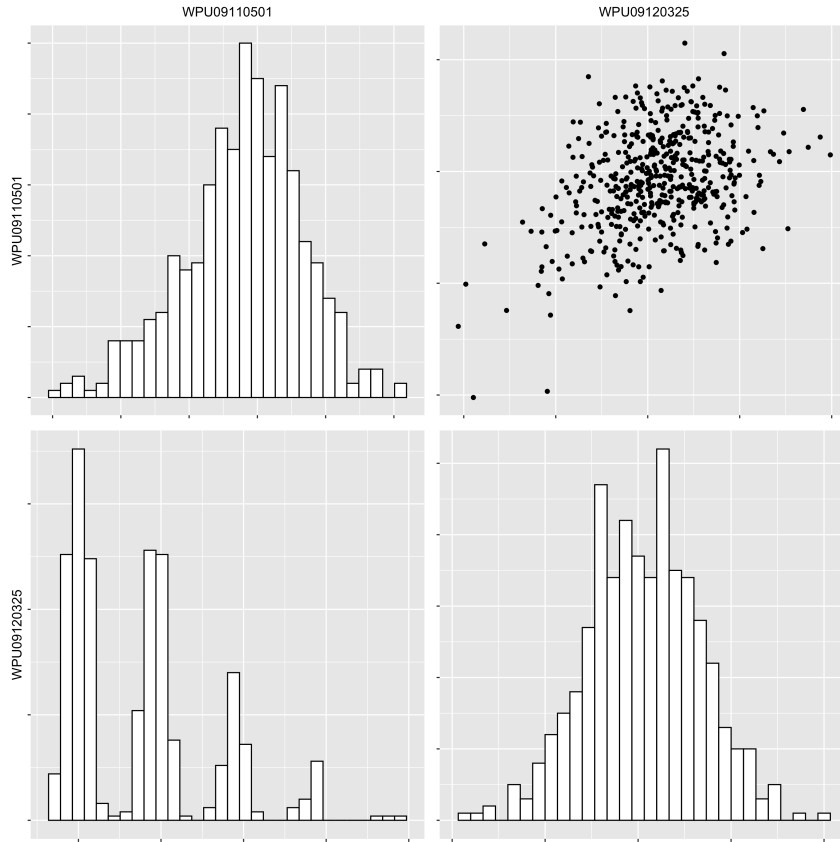


Figure 1.8: Sampling Distributions of Two Pulp Prices

sample mean against the ordinal position of the deletion. Plotting, as in figure 1.9 the other pulp prices this way reveals the same pattern. In red is the deleted sample mean of the month on month difference of logarithms. In blue is the sample calculation from simulated standard normal data. As this data is standardized, the deviation between the two shows deviation from a Gaussian model. Judging by these samples, not all of these distributions are necessarily symmetrical though spurious imbalance is expected behavior from such heavy tailed distributions.

1.9 Conclusion

Some consideration of what this all means is necessary. Practically all classical econometrics has been based on regression and therefore on cross-moments. These results suggest the necessity of developing techniques that do not depend on cross-moments, or second moments. Better yet would be the development of techniques that do not depend on first moments. One possible line of inquiry would involve the feasibility of adapting statistical learning methods for the purpose of modeling rather than simple prediction. This is the purpose of the next chapter.

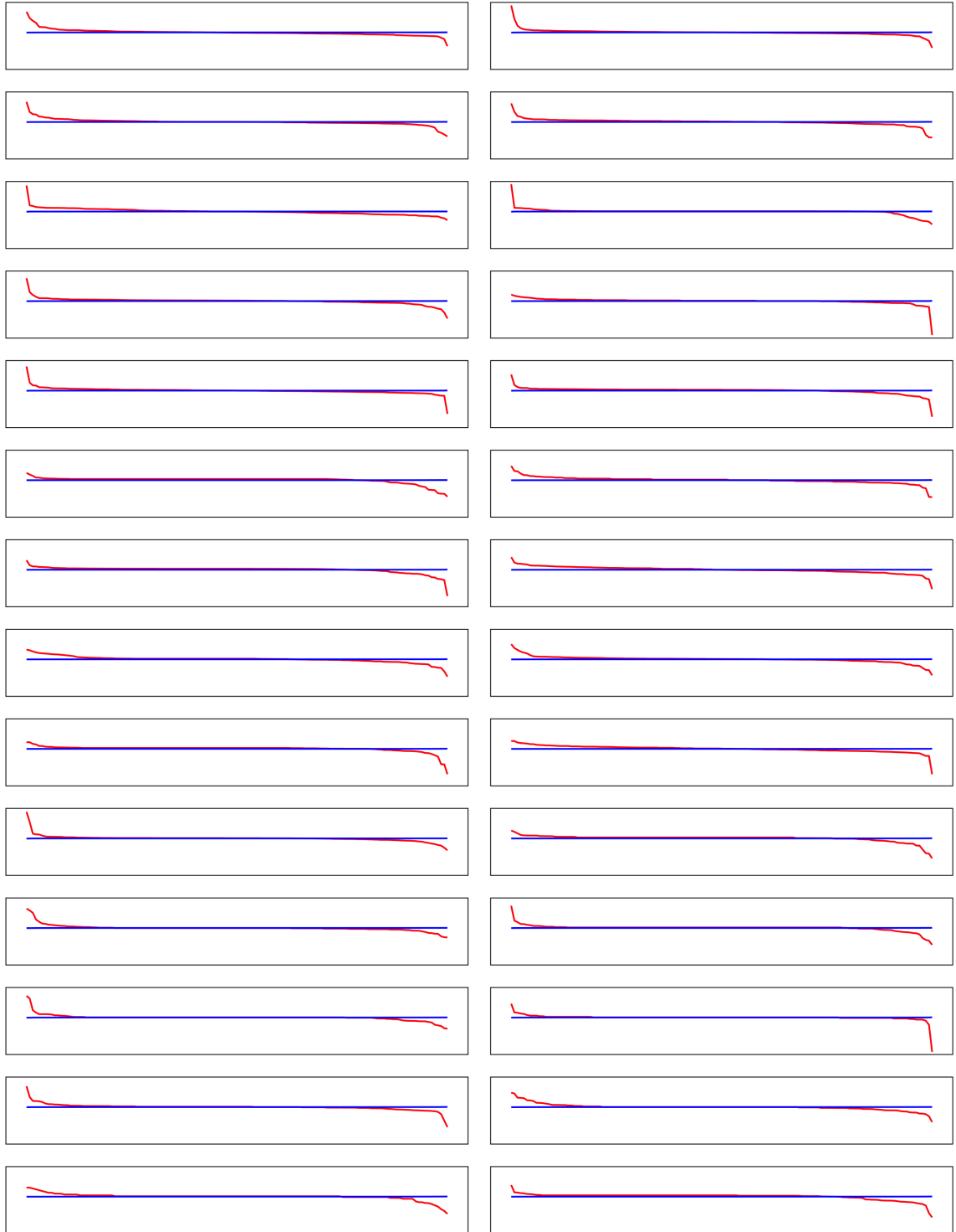


Figure 1.9: All Pulp Prices

Chapter 2: Statistical Learning for Non-Parametric Copula Estimation

Introduction

Leo Breiman famously distinguished “two cultures” of reaching conclusions from modeling data: data models and algorithmic models [13]. With the rise of statistical learning, this has developed into a culture of prediction in contradistinction to a culture of data modeling. As data sets grow larger, it is ever easier to find ways in which traditional parametric models do not fit well. On the other hand, some questions are inherently probabilistic. As useful as it is to have an algorithm that accurately makes predictions, often it is our goal to model the uncertainty. For this purpose, traditional tools such as conditional moments retain their value. This raises the question of whether the two cultures can be bridged.

This paper presents not a general attempt to model the two cultures but rather a technique to use statistical learning to build a data model. The basic probabilistic tool is the *copula*. The use of copulas, while common in finance for some time, is fairly new in statistics and econometrics. For examples see [14] and [15]. The basic idea behind a copula is that it allows us to model the univariate marginal distributions separately from their dependence. Stated very briefly, the univariate marginal distributions can be modeled using any technique that generates a cumulative distribution function. Applying this CDF to the data will map it into the unit interval and thus, if we apply each CDF to its column, our transformed dataset lies within the unit d -cube. Then, the joint CDF can be modeled as a function from this unit d -cube to the unit interval. I begin with a technical description of copulas and their advantages. I then propose a statistical learning method to fit a copula model. Finally, I apply the method to pulp price indices from the Bureau of Labor Statistics before concluding the paper.

2.1 Technical Preliminaries

2.1.1 Copulas

Following [16, p9], a d -dimensional copula $\mathcal{C} : [0, 1]^d \rightarrow [0, 1]$ is a function such that:

1. $\mathcal{C}(1, 1, \dots, u_i, \dots, 1) = u_i$.
2. \mathcal{C} is isotonic; that is $\mathcal{C}(u_1, u_2, \dots, u_d) \leq \mathcal{C}(v_1, v_2, \dots, v_d)$ provided $u_i \leq v_i$ for $i \in \{1, 2, \dots, d\}$.
3. \mathcal{C} is d -increasing within each hyper-rectangle.

From this, it follows that if any single entry is 0, the entire copula is 0. There are many known parametric copulas. For my present purposes, it suffices to consider the simplest; the *independence copula*. $\mathcal{C}_{\text{ind}} := \prod_i^d u_i$. We will use this as a foundation on which to build our technique.

An additional necessary concept is the *empirical copula*. For a data set with n independent observations and d dimensions, it is defined as:

$$\mathcal{C}_{\text{emp}} := \frac{1}{n} \sum_{i=1}^n I(U_1 \leq u_i, U_2 \leq u_2, \dots, U_d \leq u_d)$$

where I is the indicator function. The law of large numbers implies that the empirical copula will converge pointwise to the true copula.

2.1.2 Sklar's Theorem

The considerable modeling flexibility of copulas is due to *Sklar's Theorem*. Briefly, Sklar's theorem states that any multivariate CDF may be represented using a copula. I will adapt the statement given in [16, p12]. More formally:

Let F be a d -dimensional CDF with univariate marginals F_i for $i \in \{1, 2, \dots, d\}$. For each marginal, let $A_i \subset [0, 1] = F_i(\mathbb{R})$ be the range of F_i . Then, there exists a copula \mathcal{C} such

that:

$$F(x_1, x_2, \dots, x_d) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)).$$

\mathcal{C} is unique when the univariate marginal CDFs are continuous. It is also desirable that the univariate CDFs be differentiable since this can then be differentiated to yield a joint density.

2.1.3 Advantages of Copula Models

Copula-based models have many advantages. Obviously, they are extremely flexible and work well in high dimensions [15, Preface]. Another important advantage is how seamlessly they handle missing data. Since the univariate and multivariate estimation problems are separated, a missing value in one column does not render the entire row unusable. Further, when calculating empirical copulas, a missing value can be plugged with a 1 and the empirical copula calculated from the remaining columns. This is equivalent to plugging in the maximal possible value of X_i into the joint CDF. Additionally, if the researcher has any knowledge of independence in the data or is willing to assume it, then copulas can be used strategically in concert with the probability chain rule. The independence information reduces the number random variables on which each random variable depends. Then, as each conditional random variable can be represented as a quotient of joint distributions, the problem of estimating a high-dimensional random variable can be reduced to that of fitting much lower dimensional copulas. The Vines approach is related to this. See [15] for more details. Finally, given an estimate of $F(\vec{X})$, simply plugging in a 1 integrates out a dimension. In much the same way, when estimating empirical copulas, we can re-use rows of data by calculating the empirical copula in some smaller dimension and plugging in 1's for the dimensions left out. Of course, given that these rows are not statistically independent, this may pose some inferential challenges.

2.2 The Model

We begin by calculating the empirical copula. We denote it C_{emp} . In addition, it is necessary to estimate the univariate CDFs somehow. For illustrative purposes, an empirical CDF for each is simple and sufficient. In actual modeling applications, it is highly desirable that the univariate CDFs be differentiable as this allows the calculation of a joint density. Let \hat{F}_i be the empirical CDF for X_i . Now, our goal is to model C_{emp} from $\vec{U} = [\hat{F}_1(X_1), \hat{F}_2(X_2), \dots, \hat{F}_d(X_d)]$. Ideally, our model will be constrained to follow all of the mathematical requirements of a copula. This poses a bit of a mathematical challenge as not all of these constraints may be represented as convex constraints.

2.2.1 Model Preliminaries

The basic building block for the model will be Kumaraswamy cumulative distribution functions. This distribution is useful here not for probabilistic but rather for geometric purposes. For present purposes, it can be thought of as an alternative parametrization of the β -distribution to which it is closely related. The Kumaraswamy CDF is defined as: $K(x; \alpha, \gamma) = 1 - (1 - x^\alpha)^\gamma$ where α and γ are in \mathbb{R}_+ . Thus, considered as functions, they map $[0, 1]$ into $[0, 1]$ where for all α and γ , $K(0) = 0$ and $K(1) = 1$.

Rather than modeling a copula directly, it is useful to model a multiplicative “correction” on the independence copula $\mathcal{C}_{\text{emp}} = \prod_{i=1}^d u_i$. Call this correction $\Psi(\vec{u})$. Then:

$$\mathcal{C}_{\text{emp}} = (\mathcal{C}_{\text{ind}})^{W(\vec{u})} (\Psi(\vec{u}))^{1-W(\vec{u})}$$

where $W(\vec{u})$ is a weighting function which is 1 on the boundaries (where all copulas behave identically) and tends toward 0 as we move to the center. The purpose of this weighting function is to force our model to obey the boundary requirements of the copula so as to remove constraints on the correction term. W is actually a function on the order statistics of \vec{u} . Without loss of generality, relabel the dimensions such that $0 \leq u_1 \leq u_2 \leq \dots \leq u_d \leq 1$.

Now, $W(\vec{u}) = \left(\prod_{i=2}^d u_i\right)^2$. Notice that if all but one entries are 1, this function is 1. It falls quickly in any direction toward the interior of the set. The idea here is that at the boundary, our model is the independence copula. In the interior of the set, we correct the independence copula to better fit the empirical copula.

2.2.2 Model Statement

The basic idea is to model the correction as a product of exponents of kumaraswamy CDFs. Required are three tuning parameters $k \in \mathbb{N}$, $\nu_0 \in \mathbb{R}_+$ and $\delta \in \mathbb{R}_+$. The parameters k and δ together decide how finely we explore the space of these CDFs. The primary constraint here is memory as I will add in sparsity constraints later. ν_0 must be positive but should be small and can be arbitrarily close to 0. Let $T = \nu_0 + j\delta$ for $j \in \{1, 2, \dots, k\}$. Form the set $B = T^2$ where this is a Cartesian product. Then, denote:

$$K_b(u) = K(u, b)$$

where b is one of the ordered pairs of which B consists. The first and second parameter in the Kumaraswamy CDF come from this ordered pair. The idea is to approximately span the space of Kumaraswamy CDFs.

Next, I must decide the depth of our model. By this I mean that we have to decide how many interaction terms to consider. For instance, if $D = 1$, I consider u_1, u_2, \dots, u_d . If $D = 2$, I consider the all products of two of these and so on. Greater depth might make sense for large data sets but so long as sparse estimation techniques are used, memory is the primary constraint. To simplify the notation, I will re-index using t . The set I am ranging over is the set of Kumaraswamy CDFs K_b evaluated at each n -way product up to the depth. Thus, each b element corresponds to D t -elements.

Thus, the model:

$$\mathcal{C} = [\mathcal{C}_{\text{ind}}(\vec{u})]^{W(\vec{u})} \left(\prod_t [K_b(\vec{u})]^{\beta_t} \right)^{1-W(\vec{u})}$$

for all \vec{u} where no $u_i = 0$. If any $u_i = 0$, the model is 0.

2.2.3 Estimation

The form of the model above can be made highly tractable by taking the logarithm of both sides and isolating the correction and then estimating a model of that.

$$\log \mathcal{C} = W(\vec{u}) \log [\mathcal{C}_{\text{ind}}] + (1 - W(\vec{u})) \left[\sum_t \beta_t \log (K_b(\vec{u})) \right].$$

So substituting $\hat{\mathcal{C}}$ in for \mathcal{C} and solving for the rightmost term we get:

$$\hat{y} = \frac{\hat{\mathcal{C}}(\vec{u}) - W(\vec{u}) \log [\mathcal{C}_{\text{ind}}]}{1 - W(\vec{u})} = \sum_t \beta_t \log [K_b(\vec{u})].$$

As the left hand side of this can be easily calculated from data, we can estimate the right hand side using constrained optimization. The approach is similar to the LASSO.

$$\min_{\beta} \left[y - \sum_t \beta_t \log [K_b(\vec{u})]^2 \right]$$

subject to:

$$\sum_t \beta_t^2 \leq m$$

$$\sum_t \beta_t \log [K_b(\vec{u})] \leq 0.$$

The constraints are both convex. The parameter m controls sparsity. This parameter is,

in some sense, more important than the previously mentioned ones as so long as sparsity is controlled, it is tractable to consider a finer mesh of the space of Kumaraswamy CDFs. The second constraint forces the entire quantity to be negative which means it ranges from 0 to 1 when exponentiated. I can plug this model for the correction back into the initial equation and this yields a true copula and thus a model of the joint CDF.

2.2.4 Derivations from the Model

Further, we can represent the joint density using the Kumaraswamy pdfs along with the derivatives of the estimated CDFs of the univariate marginals. Thus, finding a joint density is tractable even in high dimensions. While the formula admits of a closed form, it is too complex to be usefully displayed but is tractable in a computational setting. Logarithmic differentiation is necessary here. As mentioned above, integrating variables out can be performed by simply entering a one into the copula. Thus, it is straight forward to build conditional pdfs and simulate conditional moments. Here, it is useful to think about the tuning parameter m . There are two desirable properties of a model based on m . As is traditional, model fit as represented by the sum of squared errors is useful and so m could be chosen using cross-validation. An additional consideration is the behavior of the estimated density. Greater sparsity should smooth the estimated density, possibly at the cost of model fit.

2.3 Modeling Pulp Price Indices

2.3.1 The Data

To test the formalism, I have fitted this model on an actual data set. I am using some pulp price indices supplied by the Bureau of Labor Statistics. The data product is the producer price index which has been produced since the early 20th century. In general, this is a challenging data set to work with. It changes dimension over time as price series are added or discontinued. In addition, the price series themselves tend to fail tests of stationarity while

the first difference displays moments large enough to render the finite sample properties of all techniques depending on cross-moments suspect. It is especially in these heavy-tailed situations where it is useful to separate out the modeling of the univariate marginals from the modeling of the dependence structure. Further, given the changing dimensionality of the data, the robustness of this technique to missing data is valuable.

2.3.2 Estimating the Univariate Marginals

As my present purpose is to estimate the copula, I can simply use empirical CDF's to estimate the univariate marginals. This is not ideal since these are not differentiable and thus I cannot produce densities. Further, given the heavy-tailed aspect of the data, extreme values are not well-modeled. Modeling the heavy-tailed univariate aspects is a distinct problem that is beyond my present scope.

2.3.3 Model Performance

To check the model performance, I compare the empirical copula to the model copula in a plot. The closer to the 45-degree line the plot falls, the better the fit. In figure 2.1, we see the performance in sample. The out-of-sample performance is in figure 2.2. The model fits well but, as expected, it does somewhat less well at higher values of the response variable where data is scarce.

2.4 Conclusion

This technique may turn out to be a powerful complement to more traditional parametric copula methods; especially on larger data sets. What is lacking so far is a means of performing inference on these models. In addition, this technique makes no use of the asymptotic results available in empirical process theory. Another substantial unanswered question has to do with the importance of distributional assumptions as against structural ones. This technique makes no distributional assumptions at all but it does make some structural ones. Further, any error modeling must be worked out separately for the joint CDF and the joint

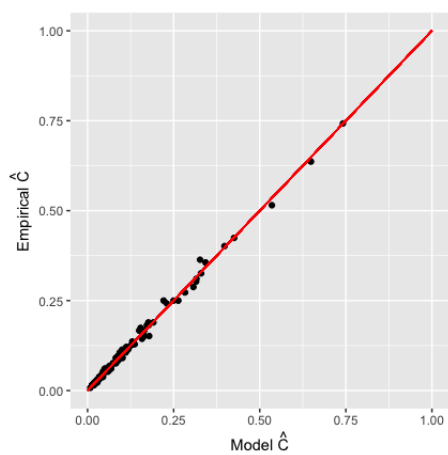


Figure 2.1: In-Sample Fit

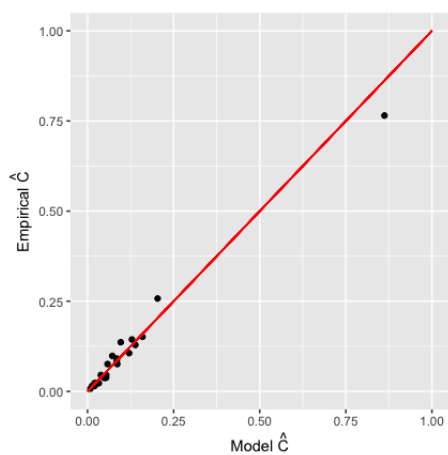


Figure 2.2: Out-of-Sample Fit

density. There are likely trade offs here that affect the choice of the tuning parameter m . Overall though, combined with dimension reduction techniques, this technique may allow the building of high dimensional probability models and thus bridge the “two cultures”. In addition, the capacity to calculate conditional moments rather than mere predictions allows this technique to work seamlessly in scientific contexts where the common language is based around regression models; as is the case, for instance, in much of econometrics.

Chapter 3: Recursion, Praxeology, and the Redemption of Econometrics

3.1 Introduction

Economics always seems to be in some kind of methodological crisis. While it might seem that methodological crises are common to all sciences and indeed they are, it is difficult to avoid the impression that in the human sciences, they seem to be more frequent and more acute. Economists have long taken pride in what they consider the greater rigor and precision of our own science in comparison to the other human sciences even if we are not always sure what we mean by this. Rigor, after all, is only the Latin for rigidity. It is a good thing only when we are sure of the end of our science and sure of the path. All too often, models, though perhaps deductively rigorous have no simple relationship to the phenomenon under study. If we are proud of our accomplishments, it is nonetheless equally obvious that the accomplishments of economics never seem to match up to the hopes placed in it. Models are initially oversold as having greater predictive power than they do. Further, there seems to be no agreement on the precise use to which economic models are put. This is exacerbated by the fact that within any given economic model, the actions open to an agent are ad hoc. The arbitrary long run- short run distinction as taught in introductory microeconomics is a simple example of this. Proper accounting for this fact should remove the dogmatism from a great many policy arguments.

It is likely due to this continual disappointment that efforts to make economics more like physics have remained popular for a century. Whatever interpretive quandaries physics may have, and it has many, there seems to be agreement on the end of physics and that which it intends. Weve come a long way from Marshalls famous dictum though whether our core formalism has advanced beyond his remains an open question. Thus, in determining

the way forward, it is necessary to have a conversation with the past. Thus, the questions requiring at least the beginnings of an answer are broad and I will be forced to paint in broad strokes. The basic questions are: I) what is and ought to be the relationship of economic science to the physical sciences, II) what is the place of quantification within economics, and III) what is a tentative way forward for economic science?

Economists take the scientific status of economics for granted. It is useful to pose the question as to what sorts of human reasoning are scientific notwithstanding the rarity with which this, often polemically intended, question is asked in good faith. A science is a set of propositions with meaning only in relationship to each other and further, a means of distinguishing some observable fact from some set of counterfactuals. In other words, in the context of a science, as currently understood, truth is the same thing correct predictions by means of a validated method. Those concerns the philosopher of science calls “methodological” involve this means of validation. Therefore, the rules of scientific reasoning ultimately reduce to rules for validation of relationships among propositions. Science requires basic objects as logical atoms and these atoms must be somehow self-similar. Thus, scientific theorizing involves compression of the immense complexity of the world. In modern economics, the atom is generally the neoclassical agent complete with an endowment and a utility function representing some preferences. Thus, this agent is a parametric atom. Given such an atom, it is clear enough that economics models itself ultimately on mathematics though filtered through statistics and physics. The understanding and critique of the philosophy and history of this filtration process is the primary end of this paper.

Before we go on, though, it is useful to raise the question: is this rootedness in mathematics inevitable? Is it not the case that mathematics is the king of sciences? It is certainly the case that mathematics is the oldest independent discipline having already been separated from philosophy by the time of Plato. We shall find out that the answer to this question is by no means simple. Any economic science will have strong mathematical elements but there is reason to believe that in order to exploit gains from trade with mathematicians, mathematics itself will have to change. Indeed, there is some work in this direction as the

rise of computation incentivizes mathematicians to reexamine fundamental assumptions and reopen old debates thought long settled.

The rise of computation, ironically a side effect of a failed effort at a universal axiomatization is of prime importance here. Computation poses questions of interpretation to scientists and philosophers and is both a tremendous opportunity and a philosophical burden. For now, it suffices to emphasize that the theory of computation has an entirely different mode of abstraction than classical mathematics. I believe that this mode of abstraction is ultimately the more fruitful one for economic science but defending this will take time. For now, it suffices to mention Abelson and Sussmans understanding of computation as outlined in the *Structure and Interpretation of Computer Programs*.

Underlying our approach to this subject is our conviction that “computer science” is not a science and that its significance has little to do with computers. The computer revolution in the way we think and the way we express what we think. The essence of this change is the emergence of what might best be called *procedural epistemology* —the study of the structure of knowledge from an imperative point of view, as opposed to the more declarative view taken by classical mathematical subjects. Mathematics provides a framework for dealing precisely with notions of “what is.” Computation provides a framework for dealing precisely with “How to” [17].

The essentially imperative nature of computation will concern us later. Indeed, it is this structure that dovetails nicely with the very procedural theory of market processes.

To place the shift from the existential logic of mathematics to the imperative logic of computation, it is necessary to travel back in time long before the rise of modern science. We must first examine the notion of a science from the point of view of the ancient Greeks.

3.2 Klein and the Transition from Ancient to Modern Science

Our guide on this brief tour of ancient science will be Jacob Klein through his magnificent *Greek Mathematical Thought and the Origin of Algebra*. We find already in Greek antiquity a highly developed system of geometry and number theory much of it with largely Pythagorean roots. The Greek conception of the world was intimately related to the Greek conception of number especially through geometry, the manifestation of number in space. Indeed, geometry formed the basic language of Greek mathematics which had as its ambition the reduction of all other concepts to geometrical ones. This is exemplified in the fact that more algebraic notions were overlaid on equivalent but originally geometrical notions [18, p 373].

As developed as these sciences were, they differed substantially from the science of high modernism in a fundamental way. Often in the course of the historical study of Greek mathematics, the fact of symbolic mathematics was taken for granted and “far too much as a matter of course” [19, p 5]. The Greek concept of “arithmos” only ever means a definite number of definite things [19, p 7]. Number, the basis of geometry, was not an abstraction. Generality was not ascribed to the mathematical objects themselves but rather the procedure for their construction. Thus, the mathematical objects in themselves were not logical atoms of the sense of modern symbolic mathematics. It is in this way that the basic mode of abstraction was fundamentally different in ancient science. A triangle was not a general triangle but rather a particular triangle constructed using the recipe for such constructions of triangles. As geometrical reason was the basis for Greek mathematics and geometry is a constructive science, its mode of generality was one of the variety of circumstances leading to the construction, not the abstraction of the thing constructed [20, p 17].

We see the beginnings of algebra in the work of the Greek mathematician Diophantus. In terms of its purely mathematical content, it expresses things known much earlier. [19, p 126]

What is of primary interest is rather its form which was symbolic rather than literary [19, p 126]. This formalism filtered down to Arabic sources before being rediscovered in the early modern period by, Francois Vieta among others. Klein tells us that Vietas elaboration of the mathematical language of Diophantus amounted to nothing less than the creation of a formal mathematical language and thereby rendered Vieta the father of modern mathematics [19, p 5]. The treatment of number not as a definite number of definite things but rather as a symbol with an arbitrary referent allowed mathematicians and thus scientists the use of a new and powerful mode of abstraction; the abstraction of objects rather than of procedures and thereby the ability to make use of logical deduction within the confines of a formal model that had not necessarily been constructed.

Descartes adapted the algebra of Vieta to create his analytic geometry which paved the way for the creation of mathematical physics as it gave a formalism for relations of measurements in space. Descartes himself was interested in work in the physical sciences and clearly intended his analytic geometry as precisely a tool for this project. Nonetheless, Descartes does not appear to have believed in his geometry, he had advanced beyond the ancient sources in any fundamental way. Indeed, Descartes did not realize that the intentionality of his mathematics was fundamentally different from that of the ancient mathematicians [20, p 13]. Now that entire relationships could be encapsulated in a symbol, mathematicians over time grew interested in formal symmetries in the symbolism itself. This led to an explosion of mathematical research that still continues to this day.

If Descartes was one of the founders of modernity, Francis Bacon was the other; though, to Hayek's point his qualifications are not above suspicion [21]. Bacon bequeathed a particular attitude to all following generations of scientists. The notion of a science for the "relief of mans estate". He also advocated the use of experimentation in contradistinction to Cartesian a priorism. The very process of experimentation gave us a new kind of data of a not-necessarily periodic sort and therefore very different from the astronomical data that had been the subject of scientific reasoning since long before the Greeks.

It turned out that the calculus of Newton and Leibniz could be used to express mathematical relationships among physical measurements, indeed calculus, the science of the real number system, appeared to be the natural formalism for physics and thus the basis of an authentically natural science of nature. Given that mathematical physics had its root in deductive mathematics, it was acceptable to the Cartesians. Given that it was a relationship among measurable, it was acceptable to the Baconians. [20] This dual interpretation was sufficient to make mathematical physics the queen of the modern sciences and thus the foundation of modernism. As the various sciences declared their independence from philosophy, they turned their attention to mathematical physics as an exemplar. None of this would have been possible without the symbolic art of algebra and its capacity for the symbolic as opposed to the procedural encapsulation of the unknowns in the expression of a problem. Nonetheless, this power came at a cost which was only beginning to be understood when Klein was writing.

The calculus of Newton and Leibniz, monumental though it was, struck many as a logical mess. Indeed, the classical presentation of the calculus did not stabilize for centuries; the process was completed by 19th century mathematicians. An important step in this clarification process was the science of the real number system studied through the set theory of Geor Cantor. The creation of set theory and the notion of a function as a mapping between two sets gave us the calculus now taught but it presented certain problems for the treatment of infinite sets which were clearly necessary as the real numbers are just such a set. This set in motion a line of research ultimately culminating in Godels Incompleteness Theorem and the invention of computer science. Bertrand Russell demonstrated that a universal set could not be an internal construct within set theory. Thus, there were two options. The one followed by what is now called “naive set theory” assumes a particular concrete set and expresses only propositions about sets of subsets of this postulated sets. This is sufficient for most mathematical purposes. Alternatively, several axiomatic set theories was developed which came with odd consequences. In particular, mathematicians became aware of the existence of classes of objects where it can be proven impossible to

construct a concrete example of any of them [22, p xvi]. If this sounds paradoxical, recall that for an object to be “constructible”, there must exist some algorithm or *procedure* that can generate it. On the other hand, an existence proof can deal with the entailments of abstract classes. Thus, modern mathematics as founded by Vieta and elaborated by Cantor and others contains pathological objects which cannot be visualized.

Klein originally wrote this during the war when a number of sciences of now immense importance to contemporary sciences were in their infancy and not widely known. Those of primary interest to us are statistics and computer science. These relationship of these “data sciences” with mathematical physics is not clear. Indeed, while mathematical physics required ideas that were later understood as probabilistic or statistical, their development within physics was informal. Explicit statistical reasoning was not needed for the development of what is not called statistical mechanics. Rather, it was biology and economics that could not do without statistical reasoning.

Statistics traditionally uses a probability distribution to model data; that is, statistical models are data models. Machine learning does not fit into this paradigm but is still fundamentally based on probability. If machine learning values predictive accuracy above all else, the relatively simple data models of mathematical statistics are instead interpretable at the cost of simplicity. In any case, the notion of probability is the fundamental means of organizing facts with respect to counter-factuals. Indeed, it seems that both biological and economic systems work by a kind of bounded randomness; in other words, we do not need to know every detail of the situation to make sense of it. Thus, probabilistic or stochastic ideas are at the core of economics. Thus, the economic interpretation of probability and therefore of the relationship between facts and counterfactuals are germane to the present question. Unfortunately, the economic interpretation of probability has never been decided on. This will be central to our commentary on the econometric revolution. First, it will be useful to review various interpretations of probability. This will inform later commentary on the semantics of statistics and econometrics.

Computer science is the other fundamental science of great importance nowadays but

nascent when Klein was writing. It is effectively an offshoot of mathematics in the sense that it is an elaboration of recursion theory. It is of interest to us in that it utilizes a different mode of abstraction than classical mathematics. Like ancient mathematics, its abstraction is procedural. It takes as given not a set of principles but rather a set of rules and a data structure. This language for “how to” in Sussmans words may prove to be a natural logic for social science insofar as social scientific questions are frequently questions of how to.

3.3 Notions of Probability

The interpretation of probability has always been controversial. Since statistics is nothing more than an application of probability theory, this semantic uncertainty has permeated statistics as well and led to much confusion. Formally, a probability distribution is a possibly infinite set of events where to every subset of that set, one can attach a real number between 0 and 1. This mathematical definition gives no guide to its relationship to actual measurement which is the critical thing if it is to serve as a scientific tool. There are effectively three offered interpretations of probability: frequentist/ classical, Bayesian, and necessary. A brief description of each should suffice.

The classical or frequentist notion of probability conceives of probability as the quantity to which some ratio limiting process converges. If we consider an experiment that could hypothetically be repeated many times, then the probability of a specific outcome is the limit of the number of times it obtains divided by the number of experiments run. While empirically, these long run rates do appear to converge, as a purely logical matter, the assumption that this is a well-defined limiting process is ad hoc. Nonetheless, even ignoring this issue, this implies that probabilities can only correctly be assigned to repeatable events. In other words, to use a classical notion of probability is to put a restriction on precisely the set of counterfactuals you consider and to assume that the event under study is not the only one that can exist in principle and indeed not uniquely defined by its generation process.

The terms “classical” and “frequentist” must be used carefully in reference to a notion of probability since they are also used to refer to the related topic of Neyman-Pearson statistical inference which is also based on the control of long run error rates. While Neyman-Pearson inference requires it be possible to express long run error rates in terms of a probability, it does not require that we restrict our notion of probability to these. In other words, Neyman-Pearson *does not* depend on the adoption of a frequentist stance on probability but it does require that one consider the plausible counterfactuals in a particular scientific situation to be outcomes of a repeated or repeatable experiment and thus has important consequences for the interpretation of any models based on it. We will meet this issue again in the course of considering a dispute between Ragnar Frisch and Tryg Haavelmoe.

The frequentist approach to probability and inference was dominant for a long time though other notions have always coexisted with it; rarely peacefully. Now popular is the Bayesian approach to both. Bruno DeFinetti, among others, pioneered this notion of probability that was then popularized in the English speaking world through LJ Savages *The Foundation of Statistics* [23, p 3]. Obviously, Savage sought a new basis on which to build statistics in this notion of probability and used it to construct statistics as a special case of decision theory. According to a Bayesian point of view, probability is a degree of belief and is subjective and therefore applicable to anything about which there is uncertainty. Critically, to this way of thinking, probability is never a property of any physical system.

Like in the previous case, the relationship between the foundational view of probability and the mode of inference based on it is not as tight as is sometimes represented. Savage himself regarded probability as an extension of logic in the case of uncertainty [23, p 7]. This is related to a rather old notion of probability as a generalization of binary logic that goes back at least to George Booles book on the laws of thought [24].

It is interesting that Savage thought of probability as an extension of logic since this is most associated with a third school of probability that Savage himself refers to as the “necessary view”. To this way of thinking, probabilities are necessary logical objects and thus are not subjective but they are also not the outcome of a physical process. This notion

of a probability distribution as containing information was championed by the physicist turned statistician ET Jaynes in his book *Probability: the Logic of Science*. While Jaynes considered himself a Bayesian, he gave rise to the heterodox school of maximum entropy methods [25, p xxii]. This concept of probability was elaborated on, altered, and, in my view, completed by WM Briggs in his recent [26]. According to Briggs, all probabilities are conditional; that is once specific things are assumed then probabilities may be deducible from these assumptions. In other words, probabilities are neither ontic nor subjective [26, p 40]. This has a natural affinity with the maximum entropy methods which then relates it to complexity science. It is the correct notion of probability in the sense that the inevitably conditional nature of probability forces the modeler to be explicit in the assignment of probabilities to counterfactuals *and therefore explicit about the counterfactuals*. This will enforce modeling discipline. His presentation is very new and one can hope it permeates econometrics. In any case, it is the only notion of probability acceptable to complexity economics.

These differing notions of probability matter because they give a probability model rather different semantics. Any scientific model that makes use of that probability model will inherit these semantics and thus, restrict its attention to the set of counterfactuals and facts comprehended in the probability model. It was precisely over this notion of probability that there was a disagreement between Ragnar Frisch and Tryg Haavelmoe.

3.4 Ragnar Frisch and the Vision of Econometrics

Ragnar Frisch was motivated to develop econometrics by his experience of the depression. Thus, econometrics had a central planning flavor to it even from the very beginning. Nonetheless, even if the science Frisch was trying to build may have been vulnerable to Hayekian knowledge problems, there are nonetheless Hayekian elements to Frisch's own thinking; likely symptomatic of the fact that these ideas were much closer to the core of economics in the 30s. Frisch, like many of his era, believed that the key to a truly scientific economics was the emulation of physics at least along certain margins. Thus, he began a

search for an analogy. It is useful to recall that Frisch would not have recognized the now observed distinction between mathematical economics and econometrics [27]. A qualitative science of economics and the attendant methods of data analysis were for Frisch one and the same. The meaning of the contemporary separation of these two things is well worth pondering.

If Frisch desired to borrow the methods of physics, he needed an analogy. Force is a core physical concept. If you conceive of an agent maximizing utility in a specific action space, there is a clear analogy to the principle of least action and Hamiltonian mechanics. Thus, utility or “oophelimity” as Frisch called it in his first Poincaré lecture is the force analogue in Frischian econophysics to use an anachronistic term. The concept of force is thus the basic organizing concept in Frischian economics. The agent impeded by oophelimity is the logical atom. Given these logical atoms, Frisch wishes to build models using structural equations. Ideally, these models have microfoundations. Indeed, from the first of Frisch’s Poincaré lectures, it is clear that there is agent heterogeneity. In other words, not only does Ragnar Frisch have microfoundations but he also anticipated agent-based modeling [27, p 23-26].

This is not all Frisch anticipated. His final Poincaré lecture contains reflections on the philosophy of chaos [27]. Again showing the influence of physics, in particular, thermodynamics and the existence of information destroying processes, Frisch highlights some of the considerable drawbacks to the mathematical approach he nonetheless champions. He correctly articulates the notion that what is required is a philosophy of chaos [27, p 129]. This understanding of the complexity aspects of economics is critical and, indeed, I have found no indication the same caution was present in later generations of econometricians.

This understanding of complexity is likely related to Frisch’s rejection of the statistical methods of Ronald Aylmer Fisher. These methods are similar but also distinct from those developed by Egon Pearson and Jerzy Neyman. Fisher relied on experimental control to establish causation. From Fisher comes the traditional statistical attitude that without such experimental control, no causal claims may be made. As this was not feasible, clearly, the R A Fisher paradigm is not suited for economic research. For related reasons, Frisch

ultimately rejected the use of probability in econometrics at all as introduced by Tryg Haavelmoe [28, p 222] [28, p 98] .

3.5 Tryg Haavelmoe and the Structure of Econometric Counterfactuals

It is in the work of Tryg Haavelmoe that we begin to see recognizably modern econometrics emerge [29]. The leveraging of probability theory provides a justification for the use of regression methods in terms of mathematical relations among random variables. The trouble came from Haavelmoe's acceptance of the Neyman-Pearson framework of inference; the only one widely known at the time aside from the related system of R A Fisher. To understand the problem, it is useful to recall the precise meaning of the result of Neyman-Pearson inference. First, the data is supposed to follow a given probability distribution determined up to a parameter. The data is used to estimate the value of that parameter. If, given a specific null hypothesis referring to a particular parameter value, we require a 5% or lower p-value for rejection of the null hypothesis, this means that given the same mechanism of data generation, under the assumption that the null hypothesis is true, fewer than 5% of actual instances of generated data would produce results extreme enough to pass this threshold. The Neyman-Pearson method of inference *controls long run error rates*. In other words, the p-value rule establishes a property not of the data but *of that data generating process*. This is the only interpretation of these methods that is strictly correct. The common follow up assumption: something like the current data set was drawn at random from the infinite set of possible data sets requires additional assumptions which are hard to state explicitly and anyway do not accurately represent the research process.

So what then is the problem? If Neyman-Pearson controls long run error rates, then the particular set of counterfactuals it deals with involves the possible outcomes of some abstract data generating procedure which can only be made concrete with some sort of “many worlds” construct. In order to be made applicable, one must relate the observations

at hand to the set of counter-factuals by means of some random mechanism. In other words, an econometrician is forced to adopt the additional ad hoc assumption in order to anchor the Neyman-Pearson inferential results in the context of an actual economic model. As the results of the test at hand are not drawn at random from some hypothetical set of many worlds, we expect to see more or less what we see which is econometric estimates pointing every which way. The problem is exacerbated by the fact that even if one allows that each successful test statistic was drawn at random from a hypothetical population, it is never the case that the various tests treated jointly are drawn from such a hypothetical population precisely because research is not done at random. In other words, the application of Neyman-Pearson in support of particular propositions completely loses its validity with respect to joint propositions tested separately. Problems such as these bedevil econometric research to this day.

3.6 Judea Pearl and Haavelmoe's Idea of Causality

Haavelmoe was aware of these issues and did take steps to mitigate them [30]. Unfortunately, the aspect of his work that really stuck was the importation of Neyman-Pearson into econometrics and thus the aping by econometricians of statistical methods intended for the laboratory sciences. For reasons discussed above, the version of econometrics that caught on was in fact vulnerable precisely to the problems of which Ragner Frisch (and Haavelmoe) were aware of from the beginning [31]. It is also likely due to this importation that econometrics turned into merely statistics applied to economics and thus was introduced a division of labor between the econometrician and the mathematical economist that persists to this day and is strongly at variance with Frisch's original vision.

Judea Pearl points out the extent to which Haavelmoe anticipated his own work:

Haavelmo introduced three revolutionary insights in 1943. First, when an economist sits down to write a structural equation he/she envisions, not statistical relationships but a set of hypothetical experiments, qualitative aspects of

which are then encoded in the system of equations. Second, an economic model thus constructed is capable of answering policy intervention questions, with no further assistance from the modeler. Finally, to demonstrate the feature above, Haavelmo presented a mathematical procedure that takes an arbitrary model and produces quantitative answers to policy questions [31].

Haavelmo again has the experimental sciences in mind here.

What makes a piece of mathematical economics not only mathematics but also economics is, I believe, this: When we set up a system of theoretical relationships and use economic names for the otherwise purely theoretical variables involved, we have in mind some actual experiment, or some design of an experiment, which we could at least imagine arranging, in order to measure those quantities in real economic life that we think might obey the laws imposed on their theoretical namesakes [30].

Haavelmo is explicitly reckoning in terms of the set of counterfactuals, understood now to be manipulable variables. He shows that given a set of structural equations: for instance:

$$y = ax + \epsilon_1$$

$$x = by + \epsilon_2$$

it is *not* the case that $E[y|x] \neq ax$ [31]. In other words, a regression is not an unbiased estimator of the causal mechanism. Additionally, no amount of purely statistical controls will fix this contrary to the impression gathered from a great deal of econometric work. The problem is that economic data records measurable coincidences in the strict sense of that term. Such coincidences involve things that occur together but it is precisely when a variable is “controlled” that whatever laws more usually driving the coincidences are suspended. When considering controls, it matters who is doing the controlling. Haavelmo rejected the statistical interpretation of the parameter in favor of the causal interpretation

[31]. This requires a set of counterfactual assumptions again involving some hypothetical experiments but if it matters who is doing the controlling and what is being controlled, then any properly causal econometric model must be not a statistical model but rather an agent-based model. Haavelmoe's intuition was brilliant but it was too far ahead of its time and Ragnar Frisch's reservations were sound as well. The many worlds notion required to make Haavelmoe's causal models compatible with Neyman-Pearson inference was too vague and indeed completely invalidated any broader model of a sampling distribution of statistical models. Its independence assumptions alone often ensure this. The often contradictory nature of the available estimates in econometric literature is symptomatic of this problem; which stands in addition to the bias introduced by using Neyman-Pearson derived results as an estimator of some effect size in the presence of a publication filter.

Haavelmoe's many worlds notion needs to be sharpened into something Ragnar Frisch could accept. The current state of econometrics is symptomatic of the fact that rather than try to improve Haavelmoe's modeling of causation, the very notion of separating causal modeling from economic modeling was ignored in favor of hoping economic a priorism could determine cause and endow the merely coincidental results of traditional statistical methods with causal content. The irony here is that the a prioristic approach is on the right track and does not doom economics to anti-empiricism.

The problem rather is the notion that the joint probability distribution is the basic model. Indeed Fumio Hayashi actually defines an econometric model as a set of restrictions on a joint probability distribution [32]. This is a purely statistical notion. The difficulty is the very process of control actually invalidates the data model. To Judea Pearl's point, there is information contained in a causal model represented by structural equations that is not contained in a joint probability distribution even if the causal model does limit the number of compatible joint probability distributions.

The important thing to realize is that a causal theory may be tested *only once it has already been assumed*. The very process of estimation must follow from an assumption not only of a form of a probability distribution defined by a parametric family but also

from a structural model but a structural model *with microfoundations*. In other words, the redemption of econometrics is a job for Austrian economists. In order to do this, a new scientific paradigm, a new notion of probability, and a more generalized notion of inference is required.

3.7 History, Praxeology, and Fixed Points

In *Human Action*, Ludwig von Mises famously distinguished between a prioristic praxeology and observable history. [33] A great many of the debates and mutual misunderstandings between self-styled Austrian economists and others involve this distinction. It is, of course true, that there is no reason to expect particular economic relationships to be quantitatively similar across time. They are part of history. On the other hand, certain relationships do seem to last over time. Urban hierarchies remain stable, in some cases, for centuries. While these are part of history, that does not make them evanescent and the stability of these hierarchies is a proper subject of economic study precisely because stability is a desideratum for the construction and maintenance of institutions; not to say the design thereof.

Austrian economics has room within it for evolutionary phenomena which is why, while it can explain change, it can also legitimately explain stasis. This would be a welcome generalization over the ad hoc assumption of instantaneous determination between variables made by the Cowles commission. Those phenomena that last for a long time must themselves be driven by relatively fixed factors which themselves are heavily insulated from many influences. Indeed, it is one of the great accomplishments of the modern market economy that existing hierarchies may be so easily overturned without overturning more basic constitutional principles. An economics with an internal theory of institutions must have an evolutionary and praxeological theory of institutional stasis. Thus, it needs an explicit account of evolutionary counterfactuals. In other words, an Austrian inspired economics that takes up the project begun by Ragner Frisch will inevitably be a complexity economics. Its explicit micro foundations requires it be agent-based.

3.8 The Place of Such a Science in the History of Science

Now, at long last, we return to where we began. The initial descriptions of the scientific method had no place for probability which anyway was developed centuries later and well-understood only very late. Further, probability itself provides examples of some of the more bizarre things to fall out of mathematical analysis; in particular, the consideration of infinite sets without a well-defined limiting process leads down dark alley ways [25, p 672].

Since any economic model with microfoundations should produce a mechanism by which some change is accomplished, mathematical existence proofs are empty of economic content since they reflect symbolic rather than procedural generality. Thus, we need to make use of constructive mathematics. The way forward is a step “backward”, we must recover some the sensibilities and the wisdom of ancient science which was as obscured by modern interpretation as it was by scholastic systematizing. A concrete example will help here. F A Hayek made no secret of his reservations with the style of mathematical economics that culminated in the Arrow-Debreu theorem. It seemed to him that it gave a misleading impression of how markets work. More recently, [34] showed that the computational complexity of an Arrow-Debreu general equilibrium is exponential with respect to the number of goods in the economy. In less technical language, no computer could ever possible solve such a problem even given all the data. This result actually casts doubt on whether this Arrow-Debreu model is a true characterization of how markets actually clear. This example is instructive because it is based on Brauer’s fixed point theorem which is notably a non-constructive result. It is not necessarily a non-constructible result but given its computational complexity, it is not practically constructible and thus has no economic content.

3.9 Conclusion: Recursive vs Logical Atomism

The way forward for economic sciences is to combine the best aspects of modern and ancient science. We must replace set theoretic mathematics with constructive mathematics so even static models may hint at an account of how they came to be there. We need a richer

probability theory; that provided by Briggs and inspired by E T Jaynes. This alternative notion of probability resolves the tension between Haavelmoe's causal models and traditional limiting frequency definitions of probability and in the process, allows truly causal models to coexist with data uncertainty. In other words, the conditions on which Briggsian probabilities are defined can themselves contain the additional causal assumptions required to build and render testable causal models. This is because the causal models represented as directed acyclic graphs do constrain the compatible probability distributions and therefore can be included in the conditions of the probability.

So long as we are considering hypothetical experiments, we do not need to discard Neyman-Pearson. Indeed, there are circumstances where it is valid. It is not true, as seems widely believed that Neyman-Pearson depends on the limiting frequency *definition* of probability. This means that the best of both the contemporary econometrics and the original econometrics would remain a special case of the new econometrics. In particular, there is no reason to dismiss the physics analogy that Frisch entertained so long as its validity is not justified in positivist terms.

A combination of Pearl- Haavelmoe style causal models and the Jaynes-Briggs notion of probability is necessary but not sufficient. Indeed, if we want to build macroeconomic models, we must have an economically meaningful, and therefore, algorithmic, means of aggregation. The models must be agent-based because causal models require manipulation of the underlying situation but it matters who manipulates what. Manipulation is subjective. I mentioned earlier that Ragnar Frisch in his remarks in Paris anticipated agent-based modeling. It is my contention that, after all these years, Austrian inspired agent-based modeling is the rightful inheritor of his project for understanding if not for control.

Appendix 4: Network Information

Table 4.1: Network Node In-Degrees

Code	In Degree	Description
WPUSI089011	103	Special indexes-Machinery and equipment, except electrical
WPU10230205	81	Metals and metal products-Used beverage can scrap
WPUIP231000	75	Net inputs to new construction, excluding capital investment, labor, and imports
WPUSI019011	73	Special indexes-Copper and copper products
WPU10260314	72	Metals and metal products-Copper wire and cable
WPU12670141	71	Furniture and household durables-Cutlery, scissors, shears, trimmers, and snips
WPU11621201	70	Machinery and equipment-Textile machinery (except parts, attachments, and accessories)
WPU09150216	69	Pulp, paper, and allied products-Specialty bags, pouches and liners
WPU11840103	68	Machinery and equipment-Counting devices
WPU13950111	68	Nonmetallic mineral products-Dressed dimension granite (including gneiss, syenite, diorite, and cut granite)

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU11440378	66	Machinery and equipment-Parts and attachments for industrial trucks and tractors
WPU03830351	63	Textile products and apparel-All other fabricated textile products, excluding trimming and findings
WPU03940101	63	Textile products and apparel-Textile fibers, yarns, and fabrics, n.e.c.
WPU12410220	63	Furniture and household durables-Household laundry equipment and parts
WPU10230201	61	Metals and metal products-Solids and clippings, new aluminum base scrap
WPUIP230000	61	Net inputs to construction industries, excluding capital investment, labor, and imports
WPUIP231100	61	Net inputs to new residential construction, excluding capital investment, labor, and imports
WPUSI012011	60	Special indexes-Construction materials
WPUSI017011	58	Special indexes-Fabricated metal products
WPU09470102	57	Pulp, paper, and allied products-Magazine and periodical printing (lithographic)
WPU11840102	57	Machinery and equipment-Integrating and totalizing meters for gases or liquids

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU11440602	56	Machinery and equipment-Parts and accessories for bulk material handling conveyors and conveying systems
WPU10230206	54	Metals and metal products-Other old aluminum base scrap
WPU11411911	54	Machinery and equipment-Parts and attachments for pumps
WPU13210121	54	Nonmetallic mineral products-Crushed and broken stone
WPU03830321	53	Textile products and apparel-Rope, cordage, and twine
WPU10540211	53	Metals and metal products-Bath and shower fittings
WPU11382631	53	Machinery and equipment-Metalworking presses (except forging and die-stamping presses)
WPU10740803	52	Metals and metal products-Iron, steel, and aluminum stairs, staircases and fire escapes
WPU13990211	52	Nonmetallic mineral products-Dimension stone mining and quarrying
WPU11470143	51	Machinery and equipment-Dust collection and other air purification equipment for cleaning incoming air

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU11490538	51	Machinery and equipment-Other roller bearings, unmounted
WPU09471101	50	Pulp, paper, and allied products-Commercial screen printing, excluding on garments, apparel accessories, and other fabric articles
WPU12210112	50	Furniture and household durables-Wood office seating, including upholstered
WPU11490201	49	Machinery and equipment-Gates, globes, angles, and checks
WPU11490202	49	Machinery and equipment-Industrial ball valves, incl. manual and power operated
WPU13440131	49	Nonmetallic mineral products-Clay floor and wall tile, glazed and unglazed (including quarry tile and ceramic mosaic tile)
WPU13990101	49	Nonmetallic mineral products-Industrial glass sand
WPU1P232100	49	Net inputs to residential maintenance and repair, excluding capital investment, labor, and imports
WPU10270111	48	Metals and metal products-Nonferrous forge shop products
WPU10890811	48	Metals and metal products-Flexible packaging foil

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPUIP232200	48	Net inputs to nonresidential maintenance and repair, excl. capital investment, labor, and imports
WPU09220125	47	Pulp, paper, and allied products-Particleboard made from particleboard produced at the same location
WPU10230104	47	Metals and metal products-Other copper and brass scrap
WPU10890564	47	Metals and metal products-Metal powders, paste, and flake
WPUIP232000	47	Net inputs to maintenance and repair construction, excluding capital investment, labor, and imports
WPUSI134011	47	Special indexes-Selected textile mill products
WPU10260301	46	Metals and metal products-Electronic wire and cable
WPU11350243	46	Machinery and equipment-Precision measuring tools
WPU10230102	45	Metals and metal products-No. 2 copper scrap, including wire
WPU10810231	45	Metals and metal products-Cap, set, machine, lag, flange, and self-locking screws, except aircraft types

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU11790103	45	Machinery and equipment-Storage batteries, lead acid type, BCI dimensional group 8D or smaller
WPU13620121	45	Nonmetallic mineral products-Roofing asphalts, pitches, coatings, and cement
WPU13920201	44	Nonmetallic mineral products-Mineral wool for industrial, equipment, and appliance insulation
WPU11350501	43	Machinery and equipment-Other machine tool attachments and accessories
WPU13312101	43	Nonmetallic mineral products-Decorative concrete block (including screen, split, slump, shadowal block, etc.)
WPU13450199	43	Nonmetallic mineral products-All other structural clay products, excluding clay refractories
WPU09150301	42	Pulp, paper, and allied products-Corrugated shipping containers
WPU09470506	42	Pulp, paper, and allied products-Advertising printing (lithographic)
WPU10170710	42	Metals and metal products-Cold rolled steel sheet and strip
WPU11140611	41	Machinery and equipment-Harvesting machinery (except hay and straw) and attachments
WPU11784901	41	Machinery and equipment-Bare printed circuit boards

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU11790512	41	Machinery and equipment-Irradiation equipment
WPU11450201	40	Machinery and equipment-Plain bearings and bushings
WPU11730901	40	Machinery and equipment-Parts and supplies for motors and generators
WPU13130117	40	Nonmetallic mineral products-All other machine-made pressed and blown glass-ware
WPUSI006111	40	Special indexes-All commodities except farm products
WPU11410701	39	Machinery and equipment-Parts and attachments for air and gas compressors and vacuum pumps
WPU10230101	38	Metals and metal products-No. 1 copper scrap, including wire
WPUFD413121	38	Final demand-Private capital equipment for manufacturing industries
WPUSI018011	38	Special indexes-Special metals and metal products
WPU10260333	37	Metals and metal products-Fiber optic cable
WPU12210114	37	Furniture and household durables-Wood office files, storage units, and tables

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPUSISTEEL2	37	Special indexes-Finished steel mill products, excluding fabricated wire products
WPU09150218	36	Pulp, paper, and allied products-Shipping sacks and multiwall bags, all materials, except textiles
WPU113A0101	36	Machinery and equipment-Metalworking assembly machines
WPU13330101	36	Nonmetallic mineral products-Ready-mix concrete
WPUSI070011	36	Special indexes-Pharmaceutical preparations for human and veterinary use
WPUSI094011	36	Special indexes-Underwear and nightwear
WPU09150441	35	Pulp, paper, and allied products-Other corrugated and solid fiber products, including containers, pallets, pads, etc.
WPU10720122	35	Metals and metal products-Gas cylinders
WPU11440485	35	Machinery and equipment-Overhead traveling cranes and monorail systems
WPU11490534	35	Machinery and equipment-Mounted bearings, except plain
WPU11760302	35	Machinery and equipment-Radio station and wireless communication equipment
WPU15110154	35	Miscellaneous products-Other nonelectronic toys, including parts

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU10750103	34	Metals and metal products-Heat exchangers and steam condensers (except for nuclear applications)
WPU10880101	34	Metals and metal products-Ferrous wire rope, cable, forms and strand
WPU10890522	34	Metals and metal products-Precision turned products, except automotive
WPU11941301	34	Machinery and equipment-Parts and accessories for internal combustion engines, ex. aircraft and gasoline automotive engines
WPU12680101	34	Furniture and household durables-Stamped and spun kitchen utensils, aluminum
WPUSI095011	34	Special indexes-Processed foods
WPU09150337	33	Pulp, paper, and allied products-Paperboard fiber drums with ends of any material
WPU10890521	33	Metals and metal products-Precision turned products, automotive
WPU10890701	33	Metals and metal products-Metal job stampings, except automotive
WPU11490203	33	Machinery and equipment-Industrial butterfly valves, incl. manual and power operated

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU12220409	33	Furniture and household durables-Fixtures (bank, office, and store) except wood
WPU03THRU15	32	Industrial Commodities-Industrial commodities
WPU09220131	32	Pulp, paper, and allied products-Medium density fiberboard (MDF) made from MDF produced at the same location
WPU09470402	32	Pulp, paper, and allied products-Financial and legal printing (lithographic)
WPU11490537	32	Machinery and equipment-Tapered roller bearings (including cups and cones), unmounted
WPU15110155	32	Miscellaneous products-Models, science and craft kits/supply, and collectors miniatures
WPU10810206	31	Metals and metal products-Hex bolts, including heavy, tap-and-joint
WPU11382611	31	Machinery and equipment-Metal punching and shearing (power and manual), and bending and forming machines (power only)
WPU11440212	31	Machinery and equipment-Unit handling conveyors and conveying systems
WPU10150211	30	Metals and metal products-Pressure pipe and fittings, ductile iron
WPU10890424	30	Metals and metal products-Precision mechanical wire springs

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU12150112	30	Furniture and household durables-Metal porch, lawn, outdoor and casual furniture
WPU13314101	30	Nonmetallic mineral products-Concrete pavers (including grid, interlocking, etc.)
WPUSI092011	30	Special indexes-Hosiery
WPU09470609	29	Pulp, paper, and allied products-Other commercial and general job printing
WPU10830524	29	Metals and metal products-All other miscellaneous electric and nonelectric lighting equipment, incl parts and accessories
WPU11490535	29	Machinery and equipment-Parts and components for ball and roller bearings, incl. balls and rollers
WPU10710515	28	Metals and metal products-Metal combination screen, storm sash, and storm doors
WPU11382651	28	Machinery and equipment-Other metal-forming machine tools (except presses)
WPU11420221	28	Machinery and equipment-Parts and attachments for elevators and moving stairs (sold separately)
WPU11430406	28	Machinery and equipment-Fluid power hose and tube fittings
WPUSI023012	28	Special indexes-Farm and garden tractors, less parts

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPUSI024012	28	Special indexes-Agricultural machinery, excluding tractors
WPU04410132	27	Hides, skins, leather, and related products-Personal leather goods (excluding womens handbags and purses,sacks and variety and shopping bags
WPU09150636	27	Pulp, paper, and allied products-Envelopes
WPU11490204	27	Machinery and equipment-Industrial plug valves
WPUSI020011	27	Special indexes-Chemicals and allied products, including synthetic rubber and synthetic fibers and yarns
WPU11440601	26	Machinery and equipment-Parts and accessories for unit handling conveyors and conveying systems
WPU11782890	26	Machinery and equipment-Filters, crystals, and transducers
WPU09150999	25	Pulp, paper, and allied products-Other sanitary paper products and misc. converted paper and paperboard products
WPU09470302	25	Pulp, paper, and allied products-Catalog and directory printing (lithographic)
WPU10890425	25	Metals and metal products-Other light gauge wire springs

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU11450108	25	Machinery and equipment-Loose gear, pinions and racks
WPU11470144	25	Machinery and equipment-Dust collection and other air purification equipment for industrial gas cleaning systems
WPU11760121	25	Machinery and equipment-Telephone switching and switchboard equipment
WPU13990121	25	Nonmetallic mineral products-Hydraulic fracturing sand and all other industrial sand
WPUSISTEEL1	25	Special indexes-Steel mill products, including fabricated wire products
WPU09480104	24	Pulp, paper, and allied products-Blankbooks, looseleaf binders, and devices
WPU10790354	24	Metals and metal products-Panels, parts, and sections for prefabricated buildings, steel and aluminum
WPU11792902	24	Machinery and equipment-Rectifying apparatus
WPU11930700	24	Machinery and equipment-Office machinery, including mailing, letter handling, and addressing machines and parts

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU12110102	24	Furniture and household durables-Metal household dining room and kitchen furniture
WPU13130116	24	Nonmetallic mineral products-Machine-made pressed and blown table, kitchen, art, and novelty glassware
WPU10410311	23	Metals and metal products-Motor vehicle hardware
WPUSI023011	23	Special indexes-Total tractors
WPU11140711	22	Machinery and equipment-Haying machinery and attachments
WPU11670502	22	Machinery and equipment-Parts for packing, packaging, and bottling machinery
WPU11760301	22	Machinery and equipment-Broadcast, studio, and related electronic equipment
WPU12220407	22	Furniture and household durables-Storage racks and accessories, except wood
WPU09470202	21	Pulp, paper, and allied products-Label and wrapper printing (lithographic)
WPUFD413111	21	Final demand-Nondurable consumer goods less foods and energy
WPUSI024011	21	Special indexes-Agricultural machinery, including tractors
WPUSI093011	21	Special indexes-Machinery and motive products

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU03810441	20	Textile products and apparel-Apparel and accessories, n.e.c.
WPU10150501	20	Metals and metal products-Standard and pearlitic malleable iron castings
WPU11950501	20	Machinery and equipment-Other machine shop products
WPUSI004011	20	Special indexes-Lumber and plywood
WPU10280207	19	Metals and metal products-Aluminum die-castings
WPU12690101	19	Furniture and household durables-Window shades and window shade accessories and rollers
WPU12690102	19	Furniture and household durables-Venetian blinds
WPU13450101	19	Nonmetallic mineral products-Vitrified clay sewer pipe and fittings
WPU10890589	18	Metals and metal products-Other fabricated metal products
WPU11760303	18	Machinery and equipment-Intercommunications, alarm and traffic control systems
WPU11790551	18	Machinery and equipment-Electronic hearing aids
WPU13990209	18	Nonmetallic mineral products-Ground or treated minerals and earths

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU04410112	17	Hides, skins, leather, and related products-Luggage (including suitcases, travel bags, backpacks, and other types)
WPU09470203	17	Pulp, paper, and allied products-Label and wrapper printing (gravure)
WPU11490209	17	Machinery and equipment-Other industrial valves, including nuclear
WPU11720501	17	Machinery and equipment-Test equipment for electrical, radio, and communication circuits and motors
WPU12680102	17	Furniture and household durables-Stamped and spun kitchen utensils, except aluminum
WPUFD413112	17	Final demand-Durable consumer goods
WPU11490205	16	Machinery and equipment-Plumbing and heating valves (low pressure)
WPU12310101	16	Furniture and household durables-Carpet and rugs
WPUSI021012	16	Special indexes-Agriculture machinery and equipment, less parts
WPUFD413122	15	Final demand-Private capital equipment for nonmanufacturing industries
WPUSISTEEL3	15	Special indexes-Finished steel mill products, including fabricated wire products,
WPU09150322	14	Pulp, paper, and allied products-Setup (rigid) paperboard boxes

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU11410801	14	Machinery and equipment-Industrial spraying equipment
WPU12140114	14	Furniture and household durables-Mattresses, other types, including crib, foam, waterbed mattresses and mattress inserts
WPU12210113	14	Furniture and household durables-Wood office desks and extensions
WPU10880701	13	Metals and metal products-Ferrous wire cloth, other woven wire products
WPU13990111	13	Nonmetallic mineral products-Industrial molding sand
WPU09470603	12	Pulp, paper, and allied products-Other commercial and general job printing (gravure)
WPU11490208	12	Machinery and equipment-Solenoid valves
WPU11627701	12	Machinery and equipment-Textile machinery parts and attachments
WPU12220325	12	Furniture and household durables-Nonwood office seating
WPU09150901	10	Pulp, paper, and allied products-Pasted, lined, laminated, or surface-coated paper-board

Continued on next page

Table 4.1 – *Continued from previous page*

Code	In Degree	Description
WPU09480201	10	Pulp, paper, and allied products-Edition, library, and other hardcover binding of books printed elsewhere
WPU10830522	10	Metals and metal products-Outdoor lighting equipment (including parts and accessories)
WPU113A0201	10	Machinery and equipment-Other metal-working machinery
WPU12210116	8	Furniture and household durables-Wood office furniture, other types (incl. panel/desking systems)
WPU10890507	7	Metals and metal products-Automotive job stampings

Table 4.2: Network Node Out-Degrees

Code	In Degree	Description
WPUSI089011	110	Special indexes-Machinery and equipment, except electrical
WPU12670141	84	Furniture and household durables-Cutlery, scissors, shears, trimmers, and snips
WPUSI017011	77	Special indexes-Fabricated metal products
WPU11440378	73	Machinery and equipment-Parts and attachments for industrial trucks and tractors

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU10260314	72	Metals and metal products-Copper wire and cable
WPU10230205	70	Metals and metal products-Used beverage can scrap
WPU11621201	67	Machinery and equipment-Textile machinery (except parts, attachments, and accessories)
WPUSI019011	64	Special indexes-Copper and copper products
WPUIP231100	63	Net inputs to new residential construction, excluding capital investment, labor, and imports
WPU11840103	61	Machinery and equipment-Counting devices
WPU03830351	59	Textile products and apparel-All other fabricated textile products, excluding trimming and findings
WPU10890564	59	Metals and metal products-Metal powders, paste, and flake
WPU12410220	59	Furniture and household durables-Household laundry equipment and parts
WPUSI012011	59	Special indexes-Construction materials
WPU11440602	58	Machinery and equipment-Parts and accessories for bulk material handling conveyors and conveying systems

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU13950111	58	Nonmetallic mineral products-Dressed dimension granite (including gneiss, syenite, diorite, and cut granite)
WPU09150216	57	Pulp, paper, and allied products-Specialty bags, pouches and liners
WPU03940101	56	Textile products and apparel-Textile fibers, yarns, and fabrics, n.e.c.
WPU09470102	55	Pulp, paper, and allied products-Magazine and periodical printing (lithographic)
WPU10540211	55	Metals and metal products-Bath and shower fittings
WPU13990101	54	Nonmetallic mineral products-Industrial glass sand
WPUFD413121	54	Final demand-Private capital equipment for manufacturing industries
WPU11840102	53	Machinery and equipment-Integrating and totalizing meters for gases or liquids
WPU09150301	52	Pulp, paper, and allied products-Corrugated shipping containers
WPU10230201	52	Metals and metal products-Solids and clippings, new aluminum base scrap
WPUIP232100	52	Net inputs to residential maintenance and repair, excluding capital investment, labor, and imports
WPU12210112	51	Furniture and household durables-Wood office seating, including upholstered

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPUIP231000	51	Net inputs to new construction, excluding capital investment, labor, and imports
WPU13130117	50	Nonmetallic mineral products-All other machine-made pressed and blown glass-ware
WPU10740803	49	Metals and metal products-Iron, steel, and aluminum stairs, staircases and fire escapes
WPU11382631	49	Machinery and equipment-Metalworking presses (except forging and die-stamping presses)
WPU10260333	48	Metals and metal products-Fiber optic cable
WPU10810231	48	Metals and metal products-Cap, set, machine, lag, flange, and self-locking screws, except aircraft types
WPU09471101	47	Pulp, paper, and allied products-Commercial screen printing, excluding on garments, apparel accessories, and other fabric articles
WPU10890811	47	Metals and metal products-Flexible packaging foil
WPU11790103	47	Machinery and equipment-Storage batteries, lead acid type, BCI dimensional group 8D or smaller

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU11790512	47	Machinery and equipment-Irradiation equipment
WPU09220125	46	Pulp, paper, and allied products-Particleboard made from particleboard produced at the same location
WPU10260301	46	Metals and metal products-Electronic wire and cable
WPU11490538	46	Machinery and equipment-Other roller bearings, unmounted
WPU13312101	46	Nonmetallic mineral products-Decorative concrete block (including screen, split, slump, shadowal block, etc.)
WPUSISTEEL1	46	Special indexes-Steel mill products, including fabricated wire products
WPU10230101	45	Metals and metal products-No. 1 copper scrap, including wire
WPU10890521	45	Metals and metal products-Precision turned products, automotive
WPU13210121	45	Nonmetallic mineral products-Crushed and broken stone
WPUIP232200	45	Net inputs to nonresidential maintenance and repair, excl. capital investment, labor, and imports
WPUSI020011	45	Special indexes-Chemicals and allied products, including synthetic rubber and synthetic fibers and yarns

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU10170710	44	Metals and metal products-Cold rolled steel sheet and strip
WPU11760302	44	Machinery and equipment-Radio station and wireless communication equipment
WPU12220409	44	Furniture and household durables-Fixtures (bank, office, and store) except wood
WPU13314101	44	Nonmetallic mineral products-Concrete pavers (including grid, interlocking, etc.)
WPUSI134011	44	Special indexes-Selected textile mill products
WPU10230102	43	Metals and metal products-No. 2 copper scrap, including wire
WPU10230206	43	Metals and metal products-Other old aluminum base scrap
WPU11730901	43	Machinery and equipment-Parts and supplies for motors and generators
WPU09150337	42	Pulp, paper, and allied products-Paperboard fiber drums with ends of any material
WPU10270111	42	Metals and metal products-Nonferrous forge shop products
WPU10890522	42	Metals and metal products-Precision turned products, except automotive
WPU11350243	42	Machinery and equipment-Precision measuring tools

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU11470143	42	Machinery and equipment-Dust collection and other air purification equipment for cleaning incoming air
WPU11490201	42	Machinery and equipment-Gates, globes, angles, and checks
WPUSI018011	42	Special indexes-Special metals and metal products
WPU13330101	41	Nonmetallic mineral products-Ready-mix concrete
WPU13920201	41	Nonmetallic mineral products-Mineral wool for industrial, equipment, and appliance insulation
WPUSISTEEL2	41	Special indexes-Finished steel mill products, excluding fabricated wire products
WPU09150441	40	Pulp, paper, and allied products-Other corrugated and solid fiber products, including containers, pallets, pads, etc.
WPU11760121	40	Machinery and equipment-Telephone switching and switchboard equipment
WPU09470506	39	Pulp, paper, and allied products-Advertising printing (lithographic)
WPU10230104	39	Metals and metal products-Other copper and brass scrap
WPU10890701	39	Metals and metal products-Metal job stampings, except automotive

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU11490202	39	Machinery and equipment-Industrial ball valves, incl. manual and power operated
WPU11784901	39	Machinery and equipment-Bare printed circuit boards
WPU13440131	39	Nonmetallic mineral products-Clay floor and wall tile, glazed and unglazed (including quarry tile and ceramic mosaic tile)
WPU13990211	39	Nonmetallic mineral products-Dimension stone mining and quarrying
WPU15110154	39	Miscellaneous products-Other nonelectronic toys, including parts
WPU11140611	38	Machinery and equipment-Harvesting machinery (except hay and straw) and attachments
WPU11450201	38	Machinery and equipment-Plain bearings and bushings
WPU11941301	38	Machinery and equipment-Parts and accessories for internal combustion engines, ex. aircraft and gasoline automotive engines
WPU12210114	38	Furniture and household durables-Wood office files, storage units, and tables
WPU09150218	37	Pulp, paper, and allied products-Shipping sacks and multiwall bags, all materials, except textiles

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU09480104	37	Pulp, paper, and allied products-Blankbooks, looseleaf binders, and devices
WPU12680101	37	Furniture and household durables-Stamped and spun kitchen utensils, aluminum
WPU11420221	36	Machinery and equipment-Parts and attachments for elevators and moving stairs (sold separately)
WPU13990121	36	Nonmetallic mineral products-Hydraulic fracturing sand and all other industrial sand
WPUSI095011	36	Special indexes-Processed foods
WPU03830321	35	Textile products and apparel-Rope, cordage, and twine
WPU10150211	35	Metals and metal products-Pressure pipe and fittings, ductile iron
WPU10710515	35	Metals and metal products-Metal combination screen, storm sash, and storm doors
WPU11411911	35	Machinery and equipment-Parts and attachments for pumps
WPU11440212	35	Machinery and equipment-Unit handling conveyors and conveying systems

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU13450199	35	Nonmetallic mineral products-All other structural clay products, excluding clay refractories
WPUIP230000	35	Net inputs to construction industries, excluding capital investment, labor, and imports
WPUIP232000	35	Net inputs to maintenance and repair construction, excluding capital investment, labor, and imports
WPUSI094011	35	Special indexes-Underwear and nightwear
WPU09220131	34	Pulp, paper, and allied products-Medium density fiberboard (MDF) made from MDF produced at the same location
WPU09470402	34	Pulp, paper, and allied products-Financial and legal printing (lithographic)
WPU10750103	34	Metals and metal products-Heat exchangers and steam condensers (except for nuclear applications)
WPU11350501	34	Machinery and equipment-Other machine tool attachments and accessories
WPU11382651	34	Machinery and equipment-Other metal-forming machine tools (except presses)
WPU13620121	34	Nonmetallic mineral products-Roofing asphalts, pitches, coatings, and cement
WPU11430406	33	Machinery and equipment-Fluid power hose and tube fittings

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU11490535	33	Machinery and equipment-Parts and components for ball and roller bearings, incl. balls and rollers
WPU10810206	32	Metals and metal products-Hex bolts, including heavy, tap-and-joint
WPU11410701	32	Machinery and equipment-Parts and attachments for air and gas compressors and vacuum pumps
WPU09150636	31	Pulp, paper, and allied products-Envelopes
WPU10880101	31	Metals and metal products-Ferrous wire rope, cable, forms and strand
WPU11382611	31	Machinery and equipment-Metal punching and shearing (power and manual), and bending and forming machines (power only)
WPU11950501	31	Machinery and equipment-Other machine shop products
WPU15110155	31	Miscellaneous products-Models, science and craft kits/supply, and collectors miniatures
WPUSI023011	31	Special indexes-Total tractors
WPUSI070011	31	Special indexes-Pharmaceutical preparations for human and veterinary use
WPU04410132	30	Hides, skins, leather, and related products-Personal leather goods (excluding womens handbags and purses)

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU11490203	30	Machinery and equipment-Industrial butterfly valves, incl. manual and power operated
WPU13130116	30	Nonmetallic mineral products-Machine-made pressed and blown table, kitchen, art, and novelty glassware
WPUSI006111	30	Special indexes-All commodities except farm products
WPU10280207	29	Metals and metal products-Aluminum die-castings
WPU10720122	29	Metals and metal products-Gas cylinders
WPU10830524	29	Metals and metal products-All other miscellaneous electric and nonelectric lighting equipment, incl parts and accessories
WPU11490204	29	Machinery and equipment-Industrial plug valves
WPU11490534	29	Machinery and equipment-Mounted bearings, except plain
WPU12150112	29	Furniture and household durables-Metal porch, lawn, outdoor and casual furniture
WPUSI023012	29	Special indexes-Farm and garden tractors, less parts
WPU09470609	28	Pulp, paper, and allied products-Other commercial and general job printing

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU10890424	28	Metals and metal products-Precision mechanical wire springs
WPU10890589	28	Metals and metal products-Other fabricated metal products
WPU113A0101	28	Machinery and equipment-Metalworking assembly machines
WPU11440485	28	Machinery and equipment-Overhead traveling cranes and monorail systems
WPU11440601	28	Machinery and equipment-Parts and accessories for unit handling conveyors and conveying systems
WPU12220407	28	Furniture and household durables-Storage racks and accessories, except wood
WPU12680102	28	Furniture and household durables-Stamped and spun kitchen utensils, except aluminum
WPUSI024011	28	Special indexes-Agricultural machinery, including tractors
WPUFD413111	27	Final demand-Nondurable consumer goods less foods and energy
WPU11792902	26	Machinery and equipment-Rectifying apparatus
WPU11930700	26	Machinery and equipment-Office machinery, including mailing, letter handling, and addressing machines and parts

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU03THRU15	24	Industrial Commodities-Industrial commodities
WPU11470144	24	Machinery and equipment-Dust collection and other air purification equipment for industrial gas cleaning systems
WPU11670502	24	Machinery and equipment-Parts for packing, packaging, and bottling machinery
WPU11782890	24	Machinery and equipment-Filters, crystals, and transducers
WPUSI024012	24	Special indexes-Agricultural machinery, excluding tractors
WPUSI092011	24	Special indexes-Hosiery
WPU10790354	23	Metals and metal products-Panels, parts, and sections for prefabricated buildings, steel and aluminum
WPU11490537	23	Machinery and equipment-Tapered roller bearings (including cups and cones), unmounted
WPUSI021012	23	Special indexes-Agriculture machinery and equipment, less parts
WPU03810441	22	Textile products and apparel-Apparel and accessories, n.e.c.
WPU09150999	22	Pulp, paper, and allied products-Other sanitary paper products and misc. converted paper and paperboard products
WPU09470202	22	Pulp, paper, and allied products-Label and wrapper printing (lithographic)

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU09470302	22	Pulp, paper, and allied products-Catalog and directory printing (lithographic)
WPU12110102	22	Furniture and household durables-Metal household dining room and kitchen furniture
WPU12690101	22	Furniture and household durables-Window shades and window shade accessories and rollers
WPU12690102	22	Furniture and household durables-Venetian blinds
WPUSI093011	22	Special indexes-Machinery and motive products
WPU13990209	21	Nonmetallic mineral products-Ground or treated minerals and earths
WPU09150322	20	Pulp, paper, and allied products-Setup (rigid) paperboard boxes
WPU09470203	20	Pulp, paper, and allied products-Label and wrapper printing (gravure)
WPU11140711	20	Machinery and equipment-Haying machinery and attachments
WPU11450108	20	Machinery and equipment-Loose gear, pinions and racks
WPU11490205	20	Machinery and equipment-Plumbing and heating valves (low pressure)
WPU11490209	20	Machinery and equipment-Other industrial valves, including nuclear

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU13450101	20	Nonmetallic mineral products-Vitrified clay sewer pipe and fittings
WPU09470603	19	Pulp, paper, and allied products-Other commercial and general job printing (gravure)
WPU10150501	19	Metals and metal products-Standard and pearlitic malleable iron castings
WPU10410311	19	Metals and metal products-Motor vehicle hardware
WPU10890425	19	Metals and metal products-Other light gauge wire springs
WPUFD413112	19	Final demand-Durable consumer goods
WPU09480201	18	Pulp, paper, and allied products-Edition, library, and other hardcover binding of books printed elsewhere
WPU12140114	18	Furniture and household durables-Mattresses, other types, including crib, foam, waterbed mattresses and mattress inserts
WPUSI004011	18	Special indexes-Lumber and plywood
WPU113A0201	17	Machinery and equipment-Other metal-working machinery
WPU11410801	17	Machinery and equipment-Industrial spraying equipment

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU11760303	17	Machinery and equipment-Intercommunications, alarm and traffic control systems
WPU12210113	17	Furniture and household durables-Wood office desks and extensions
WPUFD413122	17	Final demand-Private capital equipment for nonmanufacturing industries
WPUSISTEEL3	17	Special indexes-Finished steel mill products, including fabricated wire products,
WPU11627701	16	Machinery and equipment-Textile machinery parts and attachments
WPU11790551	16	Machinery and equipment-Electronic hearing aids
WPU04410112	15	Hides, skins, leather, and related products-Luggage (including suitcases, travel bags, backpacks, and other types)
WPU09150901	15	Pulp, paper, and allied products-Pasted, lined, laminated, or surface-coated paper-board
WPU10880701	15	Metals and metal products-Ferrous wire cloth, other woven wire products
WPU12310101	15	Furniture and household durables-Carpet and rugs
WPU11490208	14	Machinery and equipment-Solenoid valves

Table 4.2 – *Continued from previous page*

Code	In Degree	Description
WPU11720501	14	Machinery and equipment-Test equipment for electrical, radio, and communication circuits and motors
WPU11760301	14	Machinery and equipment-Broadcast, studio, and related electronic equipment
WPU12220325	14	Furniture and household durables-Nonwood office seating
WPU12210116	13	Furniture and household durables-Wood office furniture, other types (incl. panel/desking systems)
WPU13990111	13	Nonmetallic mineral products-Industrial molding sand
WPU10830522	12	Metals and metal products-Outdoor lighting equipment (including parts and accessories)
WPU10890507	10	Metals and metal products-Automotive job stampings

Continued on next page

Appendix 5: R Code for Distance Correlation

```
setwd("/scratch/jschule4/DATA/RAW/")

# for testing purposes, subset the data

# load libraries
library(vctr, lib="/scratch/jschule4/rpkg")
library(withr, lib="/scratch/jschule4/rpkg")
library(ggplot2, lib="/scratch/jschule4/rpkg")
library(energy, lib="/scratch/jschule4/rpkg")
library(parallel)
library(matrixStats, lib="/scratch/jschule4/rpkg")
library(corrplot, lib="/scratch/jschule4/rpkg")
library(reticulate, lib="/scratch/jschule4/rpkg")
#library(corrgram)

# Detect # of machine cores
cores <- detectCores()

# we need two parameters, the data and the output file
args = commandArgs(trailingOnly=TRUE)

alpha <- 0.01
sparmat <- c()

distcorr <- function(lst)
```

```

{
  rows1 <- is.na(lst[[1]])
  rows2 <- is.na(lst[[2]])
  rows <- rows1 | rows2
  table(rows) -> absence
  if (is.na(absence["FALSE"])) {absence["FALSE"] <- 0}
  if (absence["FALSE"] < 30) {return(NA)} else
    {
      #print("Calculate Distance Correlation")
      #print(length(lst[[1]][!rows]))
      #print(length(lst[[2]][!rows]))
      dcorel <- dcor(lst[[1]][!rows], lst[[2]][!rows])
      #print(dcorel)
    }
  dcorT.test(lst[[1]][!rows], lst[[2]][!rows]) -> cortest
  if (cortest$p.value > alpha){outcorr <- 0 } else
  {outcorr <- dcorel}

  return(outcorr)
}

datalst<- list()
# Now, read in the data
res <- read.csv(args[1], na.strings = c('.'))
# subset data for rapid testing
#res <- res[,1:25]
ncols <- ncol(res)
for (i in 1:ncols)

```

```

{
for (j in 1:ncols)
  { if (j != i){ datalst [[ length(datalst)+1]]
    <- list(res[,i],res[,j],i,j)} }
}

# Now apply the distance correlation function

Dcorrs <- mclapply(datalst, distcorr, mc.cores=cores)

t <- 0

matrix(rep(0,ncols**2), ncol=ncols) -> dcormat
rownames(dcormat) <- names(res)
colnames(dcormat) <- names(res)
for (i in 1:ncols)
{
  for (j in 1:ncols){
    if (j != i)
    {
      t <- t+1
      currCorr <- Dcorrs[[t]]
      dcormat[i,j] <- currCorr
      dcormat[j,i] <- currCorr
    }
  }
}

# Now that we have the distance correlation matrix,
# we can begin figuring out what depends on what
# Now, let's generate a correlelogram

colobj <- c("red", "white", "black")
colobj=colorRampPalette(colobj)(20)

```

```

png(filename = paste("/scratch/jschule4/plots/cgram",args[3],
".png",sep=""),
width = 1000, height = 1000,
units = "px", pointsize = 12,
bg = "white")
corrplot(data.matrix(dcormat),
method="color",col=colobj,tl.pos='n',
cl.lim=c(0,1))

dev.off()

boolmat <- (dcormat==0)
rownames(boolmat) <- names(res)
colnames(boolmat) <- names(res)
print(dcormat)
print(boolmat)
# Check if there are no remaining dependencies

if (all(boolmat))
{
  print(paste("Halting on iteration:",args[3]))
  system('echo halt">>>/scratch/jschule4/ctrl.csv')
  quit()
} else
{
  print(paste("Running: iteration:",args[3]))
  system('echo run">>>/scratch/jschule4/ctrl.csv')
}

# Now, find the column dependent on the most others

```

```

c() -> degvec
#print("Chk")

for (j in 1:ncol(boolmat))
  {table(boolmat) -> degree
    if (is.na(degree["FALSE"])) {degree["FALSE"] <- 0}
    degvec <- c(degvec, degree["FALSE"])
  }

names(degvec) <- names(res)
# Now, sort this descending
sort(degvec, TRUE) -> degvec
returnreg <- names(degvec)[1]

dependency <- data.frame(boolmat[returnreg,])
dependency$X <- returnreg
dependency$Y <- rownames(dependency)
names(dependency) <- c("Bool", "X", "Y")

write.csv(dependency, file=args[2], na='.', row.names=FALSE)
# Now, we need to get a list of dependencies for the LASSO models
# We can store this in a list of lists

dep <- list()

for (i in 1:nrow(boolmat))
  {dep[[i]] <- list(rownames(boolmat)[i], c())
    for (j in 1:ncol(boolmat))
      {
        if (!boolmat[i, j]) {dep[[i]][[2]]
          <- c(dep[[i]][[2]], colnames(boolmat)[j]) }
      }
  }

```

```

}
}

# Now, we need to output these dependencies

yvar <- c()
xvar <- c()

for (i in 1:length(dep))
{
yvar <- c(yvar, dep[[i]][[1]])
xvar <- c(xvar, paste(unlist(dep[[i]][[2]]), sep=" ", collapse="_"))
print(xvar)
print(yvar)
}

outdep <- data.frame(yvar, xvar)

write.csv(outdep, file=args[2], na='.', row.names=FALSE)

# Now, we can begin building a copula model

# We need to build the dependency structure for the sampling

c(list(col(boolmat)[ncol(boolmat)], NA)) -> dep2

for (i in (ncol(boolmat)-1):1)
{
dep2 <- c(dep2, list(colnames(boolmat)[i],
colnames(boolmat)[i:ncol(boolmat)][!boolmat[i:ncol(boolmat)])))
}

print("Second_L_Dependencies")

print(dep2)

# we need nonparametric CDFs

origresnames <- names(res)

for (nm in names(res))

```

```

{
  eCDF(res[nm]) -> func
  res[paste0('U_',nm)] <- func(res[nm])
}

normres <- res[-origresnames]

# We need a function that calculates the
# percent of the data less than the current row
jCDF <- function(x,dt)
{
  vec <- c()
  for (i in 1:nrow(dt)){ vec <- c(vec,as.numeric(all(x>=dt[i,])))}
  return(mean(vec))
}

vjCDF <- function(dt)
{
  hld <- c()
  for (i in 1:nrow(dt)){ hld <- c(hld,jCDF(dt[i,],dt))}
  return(hld)
}

# Now, we can estimate conditional probabilities
for (i in dep2)
{
  jvec <- c(paste0('U_',i[[1]]),paste0('U_',unlist(i[[2]]))
  cvec <- paste0('U_',unlist(i[[2]]))
  Fulldat <- normres[jvec]
  Partdat <- normres[cvec]
  FCDF <- vjCDF(Fulldat)

```

```
PCDF <- vjCDF(Partdat)
P_hat_bar <- FCDF / PCDF
}
```

Appendix 6: SAS Code for Network Linkages

```
/*
Code to Estimate Multivariate Density
March 2020.
John S. Schuler
This Program also depends on R
*/
*OPTIONS COMPRESS=YES;
LIBNAME DOCS "/scratch/jschule4/DATA/DOCS";
LIBNAME PERM "/scratch/jschule4/DATA/PERM/";
LIBNAME SDOCS "/scratch/jschule4/DATASDOCS/";
LIBNAME MDOCS "/scratch/jschule4/DATA/MDOCS/";

*OPTIONS MACROGEN SYMBOLGEN MLOGIC;

* parameters;

* Now, set a minimum number of observations;
*to use a time series , let's try 30 years;
%LET_CTHRES=360;

*_and_set_a_min_percentage_populated;
%LET_PTHRES=.95;

*_Set_a_parameter_for_the_number_of_simulated_time_series;
%LET_SIMUL=1000;

*_Step_1;

*_Read_in_the_raw_data;
DATA_FILOC;
```

```

LENGTH_URL_$500.;
INPUT_URL;
DATA_LINES;
wp.data.1.AllCommodities
wp.data.10.Pulp
wp.data.11a.Metals10–103
wp.data.11b.Metals104–109
wp.data.12a.Machinery11–113
wp.data.12b.Machinery114–116
wp.data.12c.Machinery117–119
wp.data.12i.MachineryInputs
wp.data.13.Furniture
wp.data.14.Minerals
wp.data.14i.TransportEqInputs
wp.data.15.Transportation
wp.data.16.Miscellaneous
wp.data.17.TotalDurables
wp.data.18.SpecialIndexes
wp.data.19.SteelMill
wp.data.2.FarmProducts
wp.data.21.Aggregates
wp.data.22.FD–ID
wp.data.3.ProcessedFoods
wp.data.30.Services
wp.data.30i.TransportServInputs
wp.data.4.Textile
wp.data.5.Leather
wp.data.51i.HealthServInputs

```

```

wp.data.56 i . EntertnOtherInputs
wp.data.6. Fuels
wp.data.7. Chemicals
wp.data.7 i . ChemicalsInputs
wp.data.8. Rubber
wp.data.80. Construction
wp.data.80 i . ConstructnInputs
wp.data.8 i . RubberInputs
wp.data.9. Lumber
;

RUN;

PROC PRINT DATA=FILOC; _RUN;

X_ '
cd /scratch/jschule4/DATA/RAW/;

rm wp.*;
rm M2.csv;
rm FEDFUNDS.csv;

let path=https://download.bls.gov/pub/time.series/wp/

wget wp.series;

wget $path/wp.data.1.AllCommodities;
wget $path/wp.data.10.Pulp;
wget $path/wp.data.11a.Metals10-103;
wget $path/wp.data.11b.Metals104-109;
wget $path/wp.data.12a.Machinery11-113;
wget $path/wp.data.12b.Machinery114-116;
wget $path/wp.data.12c.Machinery117-119;

```

```

wget $path/wp.data.12 i . MachineryInputs ;
wget $path/wp.data.13 . Furniture ;
wget $path/wp.data.14 . Minerals ;
wget $path/wp.data.14 i . TransportEqInputs ;
wget $path/wp.data.15 . Transportation ;
wget $path/wp.data.16 . Miscellaneous ;
wget $path/wp.data.17 . TotalDurables ;
wget $path/wp.data.18 . SpecialIndexes ;
wget $path/wp.data.19 . SteelMill ;
wget $path/wp.data.2 . FarmProducts ;
wget $path/wp.data.21 . Aggregates ;
wget $path/wp.data.22 . FD-ID ;
wget $path/wp.data.3 . ProcessedFoods ;
wget $path/wp.data.30 . Services ;
wget $path/wp.data.30 i . TransportServInputs ;
wget $path/wp.data.4 . Textile ;
wget $path/wp.data.5 . Leather ;
wget $path/wp.data.51 i . HealthServInputs ;
wget $path/wp.data.56 i . EntertnOtherInputs ;
wget $path/wp.data.6 . Fuels ;
wget $path/wp.data.7 . Chemicals ;
wget $path/wp.data.7 i . ChemicalsInputs ;
wget $path/wp.data.8 . Rubber ;
wget $path/wp.data.80 . Construction ;
wget $path/wp.data.80 i . ConstructnInputs ;
wget $path/wp.data.8 i . RubberInputs ;
wget $path/wp.data.9 . Lumber ;
wget https://dl.dropbox.com/s/0sh5ol483bq5315/M2.csv ;

```

```
wget https://dl.dropbox.com/s/3dhhy87pkwdyj5d/FEDFUNDS.csv;
```

```
’;
```

```
FILENAME_DAT
```

```
(  
"/scratch/jschule4/DATA/RAW/wp.data.1.AllCommodities",  
"/scratch/jschule4/DATA/RAW/wp.data.10.Pulp",  
"/scratch/jschule4/DATA/RAW/wp.data.11a.Metals10-103",  
"/scratch/jschule4/DATA/RAW/wp.data.11b.Metals104-109",  
"/scratch/jschule4/DATA/RAW/wp.data.12a.Machinery11-113",  
"/scratch/jschule4/DATA/RAW/wp.data.12b.Machinery114-116",  
"/scratch/jschule4/DATA/RAW/wp.data.12c.Machinery117-119",  
"/scratch/jschule4/DATA/RAW/wp.data.12i.MachineryInputs",  
"/scratch/jschule4/DATA/RAW/wp.data.13.Furniture",  
"/scratch/jschule4/DATA/RAW/wp.data.14.Minerals",  
"/scratch/jschule4/DATA/RAW/wp.data.14i.TransportEqInputs",  
"/scratch/jschule4/DATA/RAW/wp.data.15.Transportation",  
"/scratch/jschule4/DATA/RAW/wp.data.16.Miscellaneous",  
"/scratch/jschule4/DATA/RAW/wp.data.17.TotalDurables",  
"/scratch/jschule4/DATA/RAW/wp.data.18.SpecialIndexes",  
"/scratch/jschule4/DATA/RAW/wp.data.19.SteelMill",  
"/scratch/jschule4/DATA/RAW/wp.data.2.FarmProducts",  
"/scratch/jschule4/DATA/RAW/wp.data.21.Aggregates",  
"/scratch/jschule4/DATA/RAW/wp.data.22.FD-ID",  
"/scratch/jschule4/DATA/RAW/wp.data.3.ProcessedFoods",  
"/scratch/jschule4/DATA/RAW/wp.data.30.Services",  

```

```

"/scratch/jschule4/DATA/RAW/wp.data.30 i .TransportServInputs"
"/scratch/jschule4/DATA/RAW/wp.data.4. Textile",
"/scratch/jschule4/DATA/RAW/wp.data.5. Leather",
"/scratch/jschule4/DATA/RAW/wp.data.51 i .HealthServInputs",
"/scratch/jschule4/DATA/RAW/wp.data.56 i .EntertnOtherInputs",
"/scratch/jschule4/DATA/RAW/wp.data.6. Fuels",
"/scratch/jschule4/DATA/RAW/wp.data.7. Chemicals",
"/scratch/jschule4/DATA/RAW/wp.data.7 i .ChemicalsInputs",
"/scratch/jschule4/DATA/RAW/wp.data.8. Rubber",
"/scratch/jschule4/DATA/RAW/wp.data.80. Construction",
"/scratch/jschule4/DATA/RAW/wp.data.80 i .ConstructnInputs",
"/scratch/jschule4/DATA/RAW/wp.data.8 i .RubberInputs",
"/scratch/jschule4/DATA/RAW/wp.data.9. Lumber"
);

```

```

DATA.DOCS.COMMODITIES(DROP=PERIOD_MONTH_YEAR_FLAG_SEASONAL)
  ____DOCS.PLEVEL(DROP=PERIOD_MONTH_YEAR_FLAG_SEASONAL);
  ____INFILE _DAT_ _DLM='09'X _DSD
  _____LRECL=4096 _MISSOVER
  _____FIRSTOBS=1
  _____TERMSTR=LF;
  ____LENGTH_CODE_$11.
  _____YEAR_3.
  _____PERIOD_$3.

```

```

.....VALUE_8.

.....FLAG_$1.;

.....INPUT_CODE_$YEAR_PERIOD_$VALUE_FLAG$;


*****_Clean_up_the_data_a_bit;

****IF _PERIOD='M13' _THEN_DELETE;

.....MONTH=INPUT(COMPRESS(PERIOD, , 'kd' ), _2.);

****IF _SUBSTR(CODE,1,3) _NOT_IN_( 'WPS' , 'WPU' ) _THEN_DELETE;

****IF _SUBSTR(CODE,1,3)= 'WPS' _THEN_SEASONAL=1;

.....ELSE_SEASONAL=0;

.....GRANULARITY=LENGTH(COMPRESS(SUBSTR(CODE,4) ));

.....DATE=MDY(MONTH,1 ,YEAR);

.....IF _SEASONAL_THEN_DELETE;


.....IF _GRANULARITY_NE_8 _THEN_DELETE;

.....IF _COMPRESS(CODE)="WPU00000000" _THEN_OUTPUT_DOCS.PLEVEL;

.....ELSE_OUTPUT_DOCS.COMMODITIES;


.....RUN;


PROC_SORT_DATA=DOCS.COMMODITIES_NODUPKEY;

.....BY_CODE_DATE;

RUN;


PROC_SORT_DATA=DOCS.PLEVEL_NODUPKEY;

.....BY_CODE_DATE;

RUN;

```

```

*_now, _diagnostic _freq;

PROC_FREQ_DATA=DOCS.COMMODITIES;

_____TITLE_ " Granularities ";

_____TABLES _GRANULARITY;

RUN;

TITLE_ " ";

*_Now, _pull _down _the _description _data;

FILENAME _DESCNAME

"/scratch/jschule4/DATA/RAW/wp.series";

DATA_DOCS.DESCRDAT(RENAME=(CODE=SERIES _LINK ));

INFILE _DESCNAME _dlm='09'x

dsd _lrecl=4096 _MISSOVER_____firstobs=2 _termstr=LF;

LENGTH

CODE_____ $11.

GROUP_CODE_____ $2.

ITEM_CODE_____ $6.

SEASONAL_____ $1.

BASE_DATE_____ $6.

SERIES _TITLE_____ $500.

FOOTNOTE_CODES_ $10.

BEGIN _YEAR_____ $4.

BEGIN _PERIOD_____ $3.

```

END_YEAR.....\$ 4.

END_PERIOD.....\$ 3.;

INPUT

CODE.....\$

GROUP_CODE.....\$

ITEM_CODE.....\$

SEASONAL.....\$

BASE_DATE.....\$

SERIES_TITLE....\$

FOOTNOTE_CODES_\$

BEGIN_YEAR.....\$

BEGIN_PERIOD....\$

END_YEAR.....\$

END_PERIOD.....\$ _;

IF _BEGIN_PERIOD= 'M13' _THEN_BEGIN_PERIOD= 'M12' ;

IF _END_PERIOD= 'M13' _THEN_END_PERIOD= 'M12' ;

BEGDATE=MDY(INPUT(SUBSTR(BEGIN_PERIOD,2,2),2.),

.....1,INPUT(BEGIN_YEAR,2.));

ENDDATE=MDY(INPUT(SUBSTR(END_PERIOD,2,2),2.),

.....1,INPUT(END_YEAR,2.));

RUN;

*_Now, _we _need _the _macro _data;

```
FILENAME_M2_"/scratch/jschule4/DATA/RAW/M2.csv";  
FILENAME_FF_"/scratch/jschule4/DATA/RAW/FEDFUNDS.csv";
```

```
PROC_IMPORT_DATAFILE=FF_DBMS=CSV_REPLACE_OUT=DOCS.FF;  
    GETNAMES=NO;  
RUN;
```

```
PROC_CONTENTS_DATA=DOCS.FF;  
    TITLE_"FF_Contents";  
RUN;
```

```
TITLE_";
```

```
DATA_DOCS.FF(KEEP=DATE_VALUE);  
SET_DOCS.FF;  
    VALUE=INPUT(VAR2,8.2);  
    DATE=INPUT(VAR1,YYMMDD10.);  
RUN;
```

```
PROC_PRINT_DATA=DOCS.FF(OBS=25);  
TITLE_"Federal_Funds";  
RUN;
```

```
TITLE_ " " ;
```

```
PROC_IMPORT_ DATAFILE=M2_ DBMS=CSV_ REPLACE_ OUT=DOCS.M2;
```

```
    _GETNAMES=NO;
```

```
RUN;
```

```
DATA_ DOCS.M2(KEEP=DATE_ VALUE) ;
```

```
    _SET_ DOCS.M2;
```

```
VALUE=INPUT (VAR2, 8.2) ;
```

```
DATE=INPUT (VAR1, YYMMDD10. ) ;
```

```
RUN;
```

```
PROC_CONTENTS_ DATA=DOCS.M2;
```

```
RUN;
```

```
PROC_PRINT_ DATA=DOCS.M2(OBS=25) ;
```

```
TITLE_ "M2" ;
```

```
RUN;
```

```
TITLE_ " " ;
```

```
PROC_CONTENTS_ DATA=DOCS.FF ; _RUN;
```

```
PROC_CONTENTS_ DATA=DOCS.M2 ; _RUN;
```

```
*_Step_2;
```

```

*Prep that data for analysis;

*We have four data files , the overall price index ,
*the granular prices indices , MS and Federal funds;


*Deduplicate;

PROC SORT DATA=DOCS.FF NODUPKEY; BY DATE; RUN;

PROC SORT DATA=DOCS.M2 NODUPKEY; BY DATE; RUN;

PROC SORT DATA=DOCS.PLEVEL NODUPKEY; BY DATE; RUN;


*This is slightly more complex as
  it has both a code and a date;

PROC SORT DATA=DOCS.COMMODITIES NODUPKEY;
  BY CODE DATE;
RUN;


*Now, since the M2 variable is more than monthly;
*we will need to find the reported M2 closest
  to the beginning of the next month;

DATA DOCS.M2;
  SET DOCS.M2;
  NEXTMON=INTNX(" month",DATE,1,"B");
  *Now, how many days is between the
  beginning of the next month and this report date;
  LAGVAL=INTCK("DAY",DATE,NEXTMON,"D");
RUN;


PROC CONTENTS DATA=DOCS.M2; RUN;

```

```

PROC MEANS DATA=DOCS.M2 NOPRINT;
  CLASS NEXTMON;
  OUTPUT OUT=DOCS.M2_DT_FIX MIN(LAGVAL)=LAGVAL;
RUN;

```

```

PROC SORT DATA=DOCS.M2_DT_FIX (WHERE=( _TYPE_ =1)
  KEEP=NEXTMON LAGVAL _TYPE_ );
  BY NEXTMON LAGVAL;
RUN;

```

```

PROC SORT DATA=DOCS.M2 (DROP=DATE);
  BY NEXTMON LAGVAL;
RUN;

```

```

PROC CONTENTS DATA=DOCS.M2; RUN;
PROC CONTENTS DATA=DOCS.M2_DT_FIX; RUN;
DATA DOCS.M2 (KEEP=DATE VALUE);
MERGE DOCS.M2
  DOCS.M2_DT_FIX (IN=A DROP=_TYPE_ );
  BY NEXTMON LAGVAL;
  IF A;
  DATE=NEXTMON;

RUN;

```

```

* Now, we need missing values for gaps;
PROC MEANS DATA=DOCS.FF NOPRINT;
  OUTPUT OUT=DOCS.FF_SMRY

```

```

.....MIN (DATE)=MN_DT
.....MAX (DATE)=MX_DT;
RUN;

```

```

PROC_MEANS_DATA=DOCS.M2_NOPRINT;

....OUTPUT_OUT=DOCS.M2_SMRY

.....MIN (DATE)=MN_DT
.....MAX (DATE)=MX_DT;
RUN;

```

```

PROC_MEANS_DATA=DOCS.PLEVEL_NOPRINT;

....OUTPUT_OUT=DOCS.PLEVEL_SMRY

.....MIN (DATE)=MN_DT
.....MAX (DATE)=MX_DT;
RUN;

```

```

PROC_MEANS_DATA=DOCS.COMMODITIES_NOPRINT;

....CLASS_CODE;

....OUTPUT_OUT=DOCS.COMM_SMRY

.....MIN (DATE)=MN_DT
.....MAX (DATE)=MX_DT;
RUN;

```

```

*_identify_gaps;

DATA.DOCS.FF_SMRY(KEEP=DATE);

....SET_DOCS.FF_SMRY;

....INTV=INTCK(" month",MN_DT,MX_DT," D");

....DO_J=0_TO_INTV;

```

```

.....DATE=INTNX("MONTH" ,MN_DT, J , " B" );
.....OUTPUT;
.....END;
RUN;

```

```

DATA.DOCS.M2_SMRY(KEEP=DATE);
.....SET DOCS.M2_SMRY;
.....INTV=INTCK(" month" ,MN_DT,MX_DT, " D" );
.....DO _J=0 _TO _INTV;
.....DATE=INTNX("MONTH" ,MN_DT, J , " B" );
.....OUTPUT;
.....END;
RUN;

```

```

DATA.DOCS.PLEVEL_SMRY(KEEP=DATE);
.....SET DOCS.PLEVEL_SMRY;
.....INTV=INTCK(" month" ,MN_DT,MX_DT, " D" );
.....DO _J=0 _TO _INTV;
.....DATE=INTNX("MONTH" ,MN_DT, J , " B" );
.....OUTPUT;
.....END;
RUN;

```

```

DATA.DOCS.COMM_SMRY(KEEP=CODE_DATE);
.....SET DOCS.COMM_SMRY(WHERE=( _TYPE_ =1 ));
.....INTV=INTCK(" month" ,MN_DT,MX_DT, " D" );
.....DO _J=0 _TO _INTV;
.....DATE=INTNX("MONTH" ,MN_DT, J , " B" );

```

```

*****OUTPUT;

****END;

RUN;

*Now that we have a list of dates ,
we can merge these in to yield missing
values with no date gaps;

PROC_SORT_DATA=DOCS.FF_SMR; _BY_DATE; _RUN;
PROC_SORT_DATA=DOCS.M2_SMR; _BY_DATE; _RUN;
PROC_SORT_DATA=DOCS.PLEVEL_SMR; _BY_DATE; _RUN;
PROC_SORT_DATA=DOCS.COMM_SMR; _BY_CODE_DATE; _RUN;

PROC_SORT_DATA=DOCS.FF; _BY_DATE; _RUN;
PROC_SORT_DATA=DOCS.M2; _BY_DATE; _RUN;
PROC_SORT_DATA=DOCS.PLEVEL; _BY_DATE; _RUN;
PROC_SORT_DATA=DOCS.COMMODITIES; _BY_CODE_DATE; _RUN;

DATA_DOCS.FF;
****MERGE_DOCS.FF
*****DOCS.FF_SMR(IN=A);
****BY_DATE;
****IF _A;
RUN;

*Now, shock_FF sometimes;

DATA_DOCS.M2;
****MERGE_DOCS.M2

```

```

.....DOCS.M2_SMRY(IN=A);
....BY DATE;
....IF _A;
RUN;

```

```

DATA_DOCS.PLEVEL;
....MERGE_DOCS.PLEVEL
.....DOCS.PLEVEL_SMRY(IN=A);
....BY DATE;
....IF _A;
RUN;

```

```

DATA_DOCS.COMMODITIES;
....MERGE_DOCS.COMMODITIES
.....DOCS.COMM_SMRY(IN=A);
....BY CODE DATE;
....IF _A;
RUN;

```

```

DATA_DOCS.STACK(KEEP=CODE DATE VALUE DESC VAL POP);
.....SET_DOCS.COMMODITIES(IN=A)
.....DOCS.PLEVEL(IN=B)
.....DOCS.M2(IN=C)
.....DOCS.FF(IN=D);

.....IF _B THEN DO;

```

```

.....CODE="Index";
.....DESC="Price_Index";
.....END;
.....IF _C_THEN_DO;
.....CODE="M2";
.....DESC="M2";
.....END;
.....IF _D_THEN_DO;
.....CODE="FEDFUNDS";
.....DESC="Federal_Funds_Rate";
.....END;

.....VAL_POP=NOT_MISSING(VALUE);

RUN;

*Now, we will transpose and output to R
to analyze coverage;
PROC SORT DATA=DOCS.STACK; BY DATE; RUN;

PROC TRANSPOSE DATA=DOCS.STACK
.....OUT=DOCS.TRANS_STACK(DROP=_NAME _INDEX);
.....VAR VALUE;
.....BY DATE;
.....ID CODE;
RUN;

```

```

ODS_CSV_FILE="/scratch/jschule4/DATA/RAW/transstack.csv";
PROC_PRINT_DATA=DOCS.TRANS_STACK_NOOBS;

RUN;

ODS_CSV_CLOSE;

```

```

PROC_FORMAT;

VALUE_BOOL
    ____="Missing"
    ____OTHER="Present";

RUN;

```

```

PROC_MEANS_DATA=DOCS.STACK_NOPRINT;

____CLASS_CODE;

____OUTPUT_OUT=DOCS.STACK_SMRY

____MIN (DATE)=MN_DT
____MAX (DATE)=MX_DT
____N=CNT
____MEAN (VAL_POP)=PCTPOP;

RUN;

```

```

DATA_DOCS.STACK_SMRY;

____SET DOCS.STACK_SMRY(WHERE=( _TYPE_ =1 ));

____DIFF=INTCK("MONTH",MN_DT,MX_DT,"D");

RUN;

```

```
PROC_SORT_DATA=DOCS.STACK_SMRY;
BY_DESCENDING_CNT;
RUN;
```

```
ODS_PDF_FILE="DataReport.pdf" STYLE=MINIMAL;
```

```
PROC_SORT_DATA=DOCS.STACK_SMRY;
BY_DESCENDING_CNT;
RUN;
```

```
PROC_PRINT_DATA=DOCS.STACK_SMRY;
TITLE "By Total Count";
VAR_CODE_MN_DT_MX_DT_CNT_PCTPOP;
LABEL_CODE="Commodity"
MN_DT="First Appearance"
MX_DT="Last Appearance"
DIFF="Elapsed Time"
CNT="Count"
PCTPOP="% Populated";
FORMAT_MN_DT_MX_DT_MONYY.
PCTPOP_PERCENT8.;
RUN;
```

```
PROC_SORT_DATA=DOCS.STACK_SMRY;
BY_DESCENDING_PCTPOP;
RUN;
```

```

PROC PRINT DATA=DOCS.STACK_SMRY;
  TITLE "By % Populated ";
  VAR CODE MN_DT MX_DT CNT PCTPOP;
  LABEL CODE="Commodity"
         MN_DT="First Appearance"
         MX_DT="Last Appearance"
         DIFF="Elapsed Time"
         CNT="Count"
         PCTPOP="% Populated ";
  FORMAT MN_DT MX_DT MONYY.
         PCTPOP PERCENT8.;
RUN;

```

```

TITLE " ";

```

```

ODS PDF CLOSE;

```

```

PROC SORT DATA=DOCS.STACK_SMRY (WHERE=(PCTPOP >=&PTHRES AND
  CNT >=&CTHRES))
  OUT=DOCS.KEEPERS (KEEP=CODE);
  BY CODE;
RUN;

```

```

* Now, we add a step to limit the
data to 50 observations for testing purposes;

```

```
*_This_should_be_commented_out_for_real_runs;
```

```
DATA_DOCS.KEEPERS;
```

```
____SET_DOCS.KEEPERS;
```

```
____*_IF _N_>_50 THEN_DELETE;
```

```
____RUN;
```

```
*_Now,_drop_the_series_with_insufficient_data;
```

```
PROC_SORT_DATA=DOCS.STACK;
```

```
____BY_CODE;
```

```
RUN;
```

```
DATA_DOCS.STACK;
```

```
____MERGE_DOCS.KEEPERS( IN=A)
```

```
____DOCS.STACK(DROP=VAL_POP);
```

```
____BY_CODE;
```

```
____IF _A;
```

```
RUN;
```

```
PROC_SORT_DATA=DOCS.STACK;
```

```
____BY_CODE_DATE;
```

```
RUN;
```

```
*_Now,_log_transform_everything_but_FF;
```

```
PROC_EXPAND_DATA=DOCS.STACK(WHERE=(CODE_NE_"FEDFUNDS"))
```

```
____OUT=DOCS.LOGSERIES(DROP=TIME);
```

```
____BY_CODE;
```

```

_____CONVERT_VALUE=VALUE_/ _TRANSFORMOUT=(LOG);
RUN;

*_We_will_need_lags_of_the_macro_variables
_____for_later_modeling;
PROC_EXPAND_DATA=DOCS.STACK(WHERE=(CODE=_ "FEDFUNDS" ))
_____OUT=DOCS.LAGFED(KEEP=DATE_FEDFUNDS_L:);
_____BY_CODE;

_____CONVERT_VALUE=FEDFUNDS_L_1_/ _TRANSFORMOUT=(LAG_1_);
_____CONVERT_VALUE=FEDFUNDS_L_2_/ _TRANSFORMOUT=(LAG_2_);
_____CONVERT_VALUE=FEDFUNDS_L_3_/ _TRANSFORMOUT=(LAG_3_);
_____CONVERT_VALUE=FEDFUNDS_L_4_/ _TRANSFORMOUT=(LAG_4_);
_____CONVERT_VALUE=FEDFUNDS_L_5_/ _TRANSFORMOUT=(LAG_5_);
_____CONVERT_VALUE=FEDFUNDS_L_6_/ _TRANSFORMOUT=(LAG_6_);
_____CONVERT_VALUE=FEDFUNDS_L_7_/ _TRANSFORMOUT=(LAG_7_);
_____CONVERT_VALUE=FEDFUNDS_L_8_/ _TRANSFORMOUT=(LAG_8_);
_____CONVERT_VALUE=FEDFUNDS_L_9_/ _TRANSFORMOUT=(LAG_9_);
_____CONVERT_VALUE=FEDFUNDS_L_10_/ _TRANSFORMOUT=(LAG_10);
_____CONVERT_VALUE=FEDFUNDS_L_11_/ _TRANSFORMOUT=(LAG_11);
_____CONVERT_VALUE=FEDFUNDS_L_12_/ _TRANSFORMOUT=(LAG_12);

_____CONVERT_VALUE=D_FEDFUNDS_L_0
_____/_TRANSFORMOUT=(DIF_1);
_____CONVERT_VALUE=D_FEDFUNDS_L_1
_____/_TRANSFORMOUT=(DIF_1_LAG_1_);
_____CONVERT_VALUE=D_FEDFUNDS_L_2
_____/_TRANSFORMOUT=(DIF_1_LAG_2_);
_____CONVERT_VALUE=D_FEDFUNDS_L_3

```

```

...../_TRANSFORMOUT=(DIF _1 _LAG_3_);
...._CONVERT_VALUE=D_FEDFUNDS_L_4
...../_TRANSFORMOUT=(DIF _1 _LAG_4_);
...._CONVERT_VALUE=D_FEDFUNDS_L_5
...../_TRANSFORMOUT=(DIF _1 _LAG_5_);
...._CONVERT_VALUE=D_FEDFUNDS_L_6
...../_TRANSFORMOUT=(DIF _1 _LAG_6_);
...._CONVERT_VALUE=D_FEDFUNDS_L_7
...../_TRANSFORMOUT=(DIF _1 _LAG_7_);
...._CONVERT_VALUE=D_FEDFUNDS_L_8
...../_TRANSFORMOUT=(DIF _1 _LAG_8_);
...._CONVERT_VALUE=D_FEDFUNDS_L_9
...../_TRANSFORMOUT=(DIF _1 _LAG_9_);
...._CONVERT_VALUE=D_FEDFUNDS_L_10
...../_TRANSFORMOUT=(DIF _1 _LAG_10);
...._CONVERT_VALUE=D_FEDFUNDS_L_11
...../_TRANSFORMOUT=(DIF _1 _LAG_11);
...._CONVERT_VALUE=D_FEDFUNDS_L_12
...../_TRANSFORMOUT=(DIF _1 _LAG_12);
RUN;

```

```

PROC_EXPAND_DATA=DOCS.STACK(WHERE=(COMPRESS(CODE) _=_”M2”))
.....OUT=DOCS.LAGM2(KEEP=DATE_M2_L:);
....BY_CODE;
...._CONVERT_VALUE=M2_L_1_/_TRANSFORMOUT=(LOG_LAG_1_);
...._CONVERT_VALUE=M2_L_2_/_TRANSFORMOUT=(LOG_LAG_2_);
...._CONVERT_VALUE=M2_L_3_/_TRANSFORMOUT=(LOG_LAG_3_);

```

```

_____CONVERT_VALUE=M2_L_4___/_TRANSFORMOUT=(LOG_LAG_4_);
_____CONVERT_VALUE=M2_L_5___/_TRANSFORMOUT=(LOG_LAG_5_);
_____CONVERT_VALUE=M2_L_6___/_TRANSFORMOUT=(LOG_LAG_6_);
_____CONVERT_VALUE=M2_L_7___/_TRANSFORMOUT=(LOG_LAG_7_);
_____CONVERT_VALUE=M2_L_8___/_TRANSFORMOUT=(LOG_LAG_8_);
_____CONVERT_VALUE=M2_L_9___/_TRANSFORMOUT=(LOG_LAG_9_);
_____CONVERT_VALUE=M2_L_10___/_TRANSFORMOUT=(LOG_LAG_10);
_____CONVERT_VALUE=M2_L_11___/_TRANSFORMOUT=(LOG_LAG_11);
_____CONVERT_VALUE=M2_L_12___/_TRANSFORMOUT=(LOG_LAG_12);

_____CONVERT_VALUE=D_M2_L_0___/_TRANSFORMOUT=(LOG_DIF_1_);
_____CONVERT_VALUE=D_M2_L_1
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_1);
_____CONVERT_VALUE=D_M2_L_1
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_1);
_____CONVERT_VALUE=D_M2_L_2
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_2);
_____CONVERT_VALUE=D_M2_L_3
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_3);
_____CONVERT_VALUE=D_M2_L_4
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_4);
_____CONVERT_VALUE=D_M2_L_5
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_5);
_____CONVERT_VALUE=D_M2_L_6
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_6);
_____CONVERT_VALUE=D_M2_L_7
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_7);
_____CONVERT_VALUE=D_M2_L_8

```

```

...../_TRANSFORMOUT=(LOG_DIF_1_LAG_8);
...._CONVERT_VALUE=D_M2_L_9
...../_TRANSFORMOUT=(LOG_DIF_1_LAG_9);
...._CONVERT_VALUE=D_M2_L_10
...../_TRANSFORMOUT=(LOG_DIF_1_LAG_10);
...._CONVERT_VALUE=D_M2_L_11
...../_TRANSFORMOUT=(LOG_DIF_1_LAG_11);
...._CONVERT_VALUE=D_M2_L_12
...../_TRANSFORMOUT=(LOG_DIF_1_LAG_12);
RUN;

```

```

PROC_EXPAND_DATA=DOCS.STACK(WHERE=
.....(COMPRESS(CODE)="_" Index"))
.....OUT=DOCS.LAGPDEX(KEEP=DATE_PDEX_L:);
....BY_CODE;
...._CONVERT_VALUE=PDEX_L_1_/_TRANSFORMOUT=(LOG_LAG_1_);
...._CONVERT_VALUE=PDEX_L_2_/_TRANSFORMOUT=(LOG_LAG_2_);
...._CONVERT_VALUE=PDEX_L_3_/_TRANSFORMOUT=(LOG_LAG_3_);
...._CONVERT_VALUE=PDEX_L_4_/_TRANSFORMOUT=(LOG_LAG_4_);
...._CONVERT_VALUE=PDEX_L_5_/_TRANSFORMOUT=(LOG_LAG_5_);
...._CONVERT_VALUE=PDEX_L_6_/_TRANSFORMOUT=(LOG_LAG_6_);
...._CONVERT_VALUE=PDEX_L_7_/_TRANSFORMOUT=(LOG_LAG_7_);
...._CONVERT_VALUE=PDEX_L_8_/_TRANSFORMOUT=(LOG_LAG_8_);
...._CONVERT_VALUE=PDEX_L_9_/_TRANSFORMOUT=(LOG_LAG_9_);
...._CONVERT_VALUE=PDEX_L_10_/_TRANSFORMOUT=(LOG_LAG_10);
...._CONVERT_VALUE=PDEX_L_11_/_TRANSFORMOUT=(LOG_LAG_11);
...._CONVERT_VALUE=PDEX_L_12_/_TRANSFORMOUT=(LOG_LAG_12);

```

```

_____CONVERT_VALUE=D_PDEX_L_0___/_TRANSFORMOUT=(LOG_DIF_1);
_____CONVERT_VALUE=D_PDEX_L_1
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_1);
_____CONVERT_VALUE=D_PDEX_L_2
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_2);
_____CONVERT_VALUE=D_PDEX_L_3
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_3);
_____CONVERT_VALUE=D_PDEX_L_4
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_4);
_____CONVERT_VALUE=D_PDEX_L_5
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_5);
_____CONVERT_VALUE=D_PDEX_L_6
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_6);
_____CONVERT_VALUE=D_PDEX_L_7
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_7);
_____CONVERT_VALUE=D_PDEX_L_8
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_8);
_____CONVERT_VALUE=D_PDEX_L_9
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_9);
_____CONVERT_VALUE=D_PDEX_L_10
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_10);
_____CONVERT_VALUE=D_PDEX_L_11
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_11);
_____CONVERT_VALUE=D_PDEX_L_12
_____/_TRANSFORMOUT=(LOG_DIF_1_LAG_12);

RUN;

```

```
*_Now, _combine _these _together _for _later _use;
```

```
PROC_SORT_DATA=DOCS.LAGFED; _BY _DATE; _RUN;
```

```
PROC_SORT_DATA=DOCS.LAGM2; _BY _DATE; _RUN;
```

```
PROC_SORT_DATA=DOCS.LAGPDEX; _BY _DATE; _RUN;
```

```
DATA_DOCS.LAGMACROS;
```

```
_____MERGE_DOCS.LAGFED
```

```
_____DOCS.LAGM2
```

```
_____DOCS.LAGPDEX;
```

```
_____BY _DATE;
```

```
RUN;
```

```
PROC_SORT_DATA=DOCS.STACK(WHERE=(COMPRESS(CODE) _="FEDFUNDS" )
```

```
_____KEEP=DATE_CODE_VALUE) _OUT=DOCS.FF;
```

```
_____BY _DATE;
```

```
RUN;
```

```
DATA_DOCS.M2(KEEP=DATE_VALUE);
```

```
_____SET_DOCS.LOGSERIES(WHERE=(COMPRESS(CODE) _IN _(" M2" ) ) );
```

```
RUN;
```

```
DATA_DOCS.INDEX(KEEP=DATE_VALUE);
```

```
_____SET_DOCS.LOGSERIES(WHERE=(COMPRESS(CODE) _IN _(" Index " ) ) );
```

```
RUN;
```

```
PROC_SORT_DATA=DOCS.INDEX;
```

```
.....BY DATE;
```

```
RUN;
```

```
PROC_SORT_DATA=DOCS.MACROS;
```

```
.....BY DATE;
```

```
RUN;
```

```
PROC_SORT_DATA=DOCS.LOGSERIES;
```

```
.....BY DATE;
```

```
RUN;
```

```
DATA_PERM.AUTO;
```

```
MERGE_DOCS.M2(RENAME=(VALUE=M2) IN=A)
```

```
.....DOCS.LOGSERIES(WHERE=(COMPRESS(CODE)
```

```
.....NOT_IN_("M2","FEDFUNDS","Index")) IN=B)
```

```
.....DOCS.FF(RENAME=(VALUE=FEDFUNDS) DROP=CODE IN=C)
```

```
.....DOCS.INDEX(RENAME=(VALUE=PDEX) IN=D);
```

```
BY DATE;
```

```
IF A AND B AND C AND D;
```

```
RUN;
```

```
%MACRO_REDUCE;
```

```
PROC_SQL;
```

```
CREATE_TABLE_DOCS.SUBSET_AS
```

```
.....SELECT DISTINCT CODE FROM PERM.AUTO;
```

```

DATA _DOCS.SUBSET;

  _ALL_ SET _DOCS.SUBSET(OBS=50);

RUN;

PROC _SORT _DATA=PERM.AUTO; _BY _CODE; _RUN;

PROC _SORT _DATA=DOCS.SUBSET; _BY _CODE; _RUN;

DATA _PERM.AUTO;

  _ALL_ MERGE _PERM.AUTO

  _DOCS.SUBSET(IN=A);

  _BY _CODE;

  _IF _A;

RUN;

PROC _SORT _DATA=PERM.AUTO; _BY _CODE _DATE; _RUN;

%MEND _REDUCE;

*%REDUCE;

PROC _SORT _DATA=PERM.AUTO; _BY _CODE; _RUN;

TITLE _"ARIMA _with _No _Difference _in _Macro-variables";

ODS _PDF _FILE="/scratch/jschule4/arima_report.pdf"

  _STYLE=STATISTICAL;

PROC _ARIMA _DATA=PERM.AUTO;

```

```

.....BY_CODE;

.....IDENTIFY_VAR=VALUE(1) _CROSSCORR=(FEDFUNDS(0 1)
.....M2(0 1) _PDEX(0 1))
.....OUTCOV=SDOCS.OCOVAR_NLAG=12
.....STATIONARITY=(PP=12);

.....ESTIMATE_P=12_Q=0_INPUT=(
.....0 _$_FEDFUNDS
.....1 _$_FEDFUNDS
.....2 _$_FEDFUNDS
.....3 _$_FEDFUNDS
.....4 _$_FEDFUNDS
.....5 _$_FEDFUNDS
.....6 _$_FEDFUNDS
.....7 _$_FEDFUNDS
.....8 _$_FEDFUNDS
.....9 _$_FEDFUNDS
.....10 _$_FEDFUNDS
.....11 _$_FEDFUNDS
.....12 _$_FEDFUNDS

.....0 _$_M2
.....1 _$_M2
.....2 _$_M2
.....3 _$_M2
.....4 _$_M2
.....5 _$_M2
.....6 _$_M2

```

```

.....7...$_M2
.....8...$_M2
.....9...$_M2
.....10...$_M2
.....11...$_M2
.....12...$_M2

.....0...$_PDEX
.....1...$_PDEX
.....2...$_PDEX
.....3...$_PDEX
.....4...$_PDEX
.....5...$_PDEX
.....6...$_PDEX
.....7...$_PDEX
.....8...$_PDEX
.....9...$_PDEX
.....10...$_PDEX
.....11...$_PDEX
.....12...$_PDEX_)

```

```

outest=DOCS.ESTS_ _OUTMODEL=SDOCS.MODS_OUTSTAT=SDOCS.STATS;
.....FORECAST_LEAD=0_INTERVAL=MONTH_ID=DATE_OUT=SDOCS.RESULT;
RUN;
QUIT;
ODS_PDF_CLOSE;
PROC_SORT_DATA=SDOCS.RESULT; _BY _DATE; _RUN;

```

```

PROC TRANSPOSE DATA=SDOCS.RESULT(KEEP=CODE DATE RESIDUAL)
    OUT=MDOCS.RESTRANS(DROP=_NAME_ _LABEL_ );
    BY DATE;
    ID CODE;
    VAR RESIDUAL;
RUN;

```

```

* Now, we want to merge the
macro variables in with the residuals;
PROC SORT DATA=DOCS.LAGFED; BY DATE; RUN;
PROC SORT DATA=DOCS.LAGM2; BY DATE; RUN;
PROC SORT DATA=DOCS.LAGPDEX; BY DATE; RUN;
PROC SORT DATA=MDOCS.RESTRANS; BY DATE; RUN;

```

```

DATA PERM.RESLAGALL;
    MERGE DOCS.LAGFED
    DOCS.LAGM2
    DOCS.LAGPDEX
    DOCS.FF(RENAME=(VALUE=FEDFUNDS) DROP=CODE)
    DOCS.M2(RENAME=(VALUE=M2) )
    DOCS.INDEX(RENAME=(VALUE=PDEX) )
    MDOCS.RESTRANS;
    BY DATE;
RUN;

```

```

ODS CSV FILE="/scratch/jschule4/DATA/RAW/resids1.csv";

```

```

PROC_PRINT DATA=PERM.RESLAGALL(DROP=DATE) _NOOBS;

RUN;

ODS_CSV_CLOSE;

X_” Rscript_dcorr_analyst004.R_resids1.csv_output2.csv_0_”

___XSYNC;

*_Now_prepare_the_BETAS;

DATA_SDOCS.MODS;

___SET_SDOCS.MODS(WHERE=(COMPRESS( _PARM_ )

_____IN_(”MU”,”AR”,”NUM”)));

_____LAG=MAX( _LAG_ , _SHIFT_ );

_____LENGTH_BYGRP_$ 50.;

RUN;

PROC_SORT DATA=SDOCS.MODS; _BY_ _NAME_ _LAG; _RUN;

PROC_TRANSPOSE DATA=SDOCS.MODS_OUT=SDOCS.BETA;

___BY_ _NAME_ _LAG;

___ID_CODE;

___VAR_ _VALUE_;

RUN;

ODS_CSV_FILE=”/scratch/jschule4/DATA/RAW/betas.csv”;

PROC_PRINT DATA=SDOCS.BETA

_NOOBS;

```

```
RUN;
```

```
ODS_CSV_CLOSE;
```

```
PROC_SORT_DATA=PERM.AUTO; _BY_CODE; _RUN;
```

```
PROC_EXPAND_DATA=PERM.AUTO(WHERE=(COMPRESS(CODE)
```

```
.....NOT_IN_("M2","FEDFUNDS","Index")))
```

```
.....OUT=PERM.AUTOLAG;
```

```
...._BY_CODE;
```

```
....CONVERT_VALUE=VALUE_/ _TRANSFORMOUT=(LOG_DIF1 );
```

```
....CONVERT_VALUE=L1...../ _TRANSFORMOUT=(LOG_DIF1 _LAG_1 );
```

```
....CONVERT_VALUE=L2...../ _TRANSFORMOUT=(LOG_DIF1 _LAG_2 );
```

```
....CONVERT_VALUE=L3...../ _TRANSFORMOUT=(LOG_DIF1 _LAG_3 );
```

```
....CONVERT_VALUE=L4...../ _TRANSFORMOUT=(LOG_DIF1 _LAG_4 );
```

```
....CONVERT_VALUE=L5...../ _TRANSFORMOUT=(LOG_DIF1 _LAG_5 );
```

```
....CONVERT_VALUE=L6...../ _TRANSFORMOUT=(LOG_DIF1 _LAG_6 );
```

```
....CONVERT_VALUE=L7...../ _TRANSFORMOUT=(LOG_DIF1 _LAG_7 );
```

```
....CONVERT_VALUE=L8...../ _TRANSFORMOUT=(LOG_DIF1 _LAG_8 );
```

```
....CONVERT_VALUE=L9...../ _TRANSFORMOUT=(LOG_DIF1 _LAG_9 );
```

```
....CONVERT_VALUE=L10..../ _TRANSFORMOUT=(LOG_DIF1 _LAG_10 );
```

```
....CONVERT_VALUE=L11..../ _TRANSFORMOUT=(LOG_DIF1 _LAG_11 );
```

```
....CONVERT_VALUE=L12..../ _TRANSFORMOUT=(LOG_DIF1 _LAG_12 );
```

```
RUN;
```

```
PROC_SORT_DATA=PERM.AUTOLAG; _BY_DATE; _RUN;
```

```
PROC_SORT_DATA=DOCS.LAGMACROS; _BY_DATE; _RUN;
```

```

PROC_SORT _DATA=DOCS.M2(RENAME=(VALUE=M2))

        _OUT=DOCS.FM2; _BY _DATE; _RUN;

PROC_SORT _DATA=DOCS.FF(RENAME=(VALUE=FEDFUNDS) _DROP=CODE)

        _OUT=DOCS.FFF; _BY _DATE; _RUN;

PROC_SORT _DATA=DOCS.INDEX(RENAME=(VALUE=PDEX))

        _OUT=DOCS.FPDEX; _BY _DATE; _RUN;


DATA _PERM.LAGFIN(DROP=TIME);

        _MERGE _PERM.AUTOLAG

        _DOCS.LAGMACROS

        _DOCS.FM2

        _DOCS.FFF

        _DOCS.FPDEX;

        _BY _DATE;

        IF _MISSING(FEDFUNDS) _THEN _DELETE;

RUN;


PROC_SORT _DATA=PERM.LAGFIN(DROP=DESC);

        _BY _DESCENDING _DATE _CODE; _RUN;


*_Now, _reorder _the _data;

DATA _PERM.LAGFIN;

        _ARRAY _VARBS{*} _FEDFUNDS

        _FEDFUNDS _L _ :

        _M2

        _M2 _L _ :

```

```

.....PDEX
.....PDEX_L_:
.....VALUE
.....L;;
SET _PERM.LAGFIN;
RUN;

*_Now, _make_sure_that_all_dates_and_codes
are_represented_in_each_date_,_even_if_blank;
PROC_SQL;
CREATE_TABLE _PERM.CODE_DT_AS
.....SELECT _A.CODE, _B.DATE FROM
.....(SELECT DISTINCT _CODE FROM _PERM.LAGFIN) _A,
.....(SELECT DISTINCT _DATE FROM _PERM.LAGFIN) _B
.....;
QUIT;

PROC_SORT _DATA=PERM.CODE_DT; _BY _DESCENDING _DATE _CODE; _RUN;

DATA _PERM.LAGFIN;
.....MERGE _PERM.LAGFIN
....._PERM.CODE_DT;
BY _DESCENDING _DATE _CODE;
RUN;

PROC_DATASETS _lib=PERM;
MODIFY _LAGFIN;
FORMAT _ _all _;
INFORMAT _ _all _;

```

```

RUN;

QUIT;

ODS_CSV_FILE="/scratch/jschule4/DATA/RAW/X.csv";

PROC_PRINT_DATA=PERM.LAGFIN_NOOBS;

RUN;

ODS_CSV_CLOSE;


*Now, get the dependency structure of the error term;
X" Rscript lasso_dcorr_analyst001.R\
lassores.csv_output3.csv_0" _XSYNC;


*Now, we can begin the simulation;


*First, study the difference in Federal funds rate;


PROC_EXPAND_DATA=DOCS.STACK(WHERE=(CODE=_ "FEDFUNDS" ))
      _OUT=DOCS.FDELTA(KEEP=DATE_D_FEDFUNDS_L_0);
      _BY_CODE;
      _CONVERT_VALUE=D_FEDFUNDS_L_0_/ _TRANSFORMOUT=(DIF_1);


RUN;


ODS_PDF_FILE="/scratch/jschule4/fedfunds.pdf"
      _STYLE=STATISTICAL;

PROC_SGPLOT_DATA=DOCS.FDELTA;
      _HISTOGRAM_D_FEDFUNDS_L_0;

RUN;

```

```

ODS_PDF_CLOSE;

*Now, we will start with the lowest
observed federal funds rate for T=0;
PROC_MEANS_DATA=DOCS.FF_NOPRINT;
_____OUTPUT_OUT=DOCS.FF_SMRY
_____MIN(VALUE)=FED_MIN
_____MAX(VALUE)=FED_MAX;
RUN;

DATA_PERM.SIMU;
_____SET_DOCS.FF_SMRY(KEEP=FED_MIN);
_____T=0;
RUN;

PROC_RANK_DATA=DOCS.FDELTA_OUT=DOCS.F_RANKS;
_____VAR_D_FEDFUNDS_L_0;
_____RANKS_D_RANK;
RUN;

DATA_DOCS.QUANT(KEEP=cnt);
_____SET_DOCS.F_RANKS_END=EOF;
_____IF_EOF_THEN_DO;
_____CNT=_N_;
_____END;
_____IF_CNT=_N_ THEN_OUTPUT;
RUN;

```

```
PROC SQL; CREATE TABLE DOCS.PCT AS SELECT A.*,B.CNT
        FROM DOCS.F_RANKS A,DOCS.QUANT B; QUIT; RUN;
```

```
DATA DOCS.PCT;
    SET DOCS.PCT;
    PCT=D_RANK/_CNT;
RUN;
```

```
PROC MEANS DATA=DOCS.PCT;
    VAR PCT;
RUN;
```

```
*Now, we want a smooth path for the federal funds rate from
the lowest to the highest observed
over a period of 100 ticks;

*We need a sigmoid function for which
we can use the normal CDF.

We will arrange this such that the min
and max observed federal
funds rate are 4 sds from the mean;

DATA DOCS.FF_SMRY;
SET DOCS.FF_SMRY;
DELTA=FED_MAX-FED_MIN;
SD=DELTA/16;
MN=FED_MIN+DELTA/2;
DO T=0 TO 119;
X=FED_MIN+_DELTA*
    _CDF("NORMAL",FED_MIN+(T/119)*DELTA,MN,SD);
```

```
OUTPUT;
```

```
END;
```

```
RUN;
```

```
ODS_PDF_FILE="/scratch/jschule4/ffpath.pdf"
```

```
=====STYLE=STATISTICAL;
```

```
PROC_SGPLOT_DATA=DOCS.FF_SMRY;
```

```
=====SERIES_X=T_Y=X;
```

```
RUN;
```

```
ODS_PDF_CLOSE;
```

```
DATA_DOCS.FDELTA;
```

```
=====SET_DOCS.FDELTA;
```

```
=====IF _MISSING(D_FEDFUNDS_L_0) _THEN_DELETE;
```

```
RUN;
```

```
*Now, we can assemble an approximation of this sigmoid
```

```
function out of actual observed differences
```

```
in the FF rate month to month;
```

```
PROC_SQL; _CREATE_TABLE_DOCS.SIGMOID
```

```
=====AS_SELECT_A.FED_MIN, _A.X, _A.T,
```

```
=====B.D_FEDFUNDS_L_0 FROM DOCS.FF_SMRY_A,
```

```
=====DOCS.FDELTA_B
```

```
=====ORDER_BY_T; _QUIT; _RUN;
```

```
DATA_DOCS.SIGMOID;
```

```
=====SET_DOCS.SIGMOID;
```

```

      _BY_ T;
      IF _T=0 THEN _CURR_VAL=FED_MIN;
      RETAIN _CURR_VAL _CLOSE_DIF;
      IF _FIRST . T THEN _CLOSE_DIF=D_FEDFUNDS_L_0;

      LR=ABS(X-(CURR_VAL+D_FEDFUNDS_L_0));
      GR=ABS(X-(CURR_VAL+_CLOSE_DIF));

      IF _ABS(X-(CURR_VAL+D_FEDFUNDS_L_0)) <
          _ABS(X-(CURR_VAL+_CLOSE_DIF))
          THEN _CLOSE_DIF=D_FEDFUNDS_L_0;

      IF _LAST . T THEN DO;
          _CURR_VAL=CURR_VAL+_CLOSE_DIF;
      END;
RUN;

```

```

DATA _DOCS . SIGMOID;
      SET _DOCS . SIGMOID;
      _BY_ T;
      IF _LAST . T THEN OUTPUT;
RUN;

```

```

ODS_PDF_FILE="/scratch/jschule4/ffpath_emp.pdf"
      STYLE=STATISTICAL;
PROC_SGPLOT_DATA=DOCS . SIGMOID;

```

```

      SERIES X=T Y=X;

      SERIES X=T Y=CURR_VAL;

RUN;

ODS_PDF_CLOSE;


*Now that we have a plausible FF path,
we can simulate the other macro variables;

*First, we need to model the stochastic
processes for the univariate macro variables;

PROC_SORT DATA=DOCS.LAGMACROS; BY DATE; RUN;


DATA_PERM.COPMACROS;

      MERGE DOCS.FF (RENAME=(VALUE=FEDFUNDS) _DROP=CODE)

      DOCS.M2(RENAME=(VALUE=M2))

      DOCS.INDEX (RENAME=(VALUE=PDEX))

      DOCS.LAGMACROS;

      BY DATE;

RUN;


ODS_CSV

FILE="/scratch/jschule4/DATA/RAW/macro_copula_varbs.csv";

PROC_PRINT DATA=PERM.COPMACROS_NOOBS;

RUN;

ODS_CSV_CLOSE;


*Now, we can predict the M2 and the Price
index from the Federal funds rate;

*We need to prepare the macroeconomics data set by

```

```

adding the simulated period;
DATA DOCS.SIMFED;
    SET DOCS.SIGMOID(RENAME=(X=FEDFUNDS) KEEP=X);
RUN;

PROC MEANS DATA=PERM.COPMACROS NOPRINT;
    OUTPUT OUT=DOCS.FINALDATE MAX(DATE)=DATE;
RUN;

PROC SQL; CREATE TABLE PERM.SIMMACROS AS SELECT A.*,
    B.DATE AS FIRDT
    FROM DOCS.SIMFED A, DOCS.FINALDATE B; QUIT; RUN;

DATA PERM.SIMMACROS(DROP=FIRDT);
    SET PERM.SIMMACROS;
    DATE=INTNX("MONTH",FIRDT,( _N_ -1),"SAME");
    SIM=1;
RUN;

DATA PERM.COPMACROS;
    SET PERM.COPMACROS(WHERE=(DATE<=21975))
    PERM.SIMMACROS;
RUN;

PROC ARIMA DATA=PERM.COPMACROS;
    TITLE "M2 1st Difference";

```

```

IDENTIFY VAR=M2(1) CROSSCORR=(FEDFUNDS(0 1))
OUTCOV=DOCS.M2_COVAR_1 NLAG=12 STATIONARITY=(PP=12);
ESTIMATE P=12 Q=0 INPUT=(
0 $ FEDFUNDS
1 $ FEDFUNDS
2 $ FEDFUNDS
3 $ FEDFUNDS
4 $ FEDFUNDS
5 $ FEDFUNDS
6 $ FEDFUNDS
7 $ FEDFUNDS
8 $ FEDFUNDS
9 $ FEDFUNDS
10 $ FEDFUNDS
11 $ FEDFUNDS
12 $ FEDFUNDS

)

OUTEST=DOCS.M2_ESTS_1 OUTMODEL=DOCS.M2_MODS_1
OUTSTAT=DOCS.M2_STATS_1;
FORECAST LEAD=120
INTERVAL=MONTH ID=DATE OUT=DOCS.M2_RESULTS_1;
RUN;
QUIT;

*Now we have an estimate for M2,
we can use this estimate to forecast PDEX;

```

```

DATA _DOCS.M2_CAST(DROP=FORECAST);

_____SET _DOCS.M2_RESULTS_1(KEEP=DATE_M2_FORECAST);

_____M2=COALESCE(M2,FORECAST);

RUN;


DATA _PERM.COPMACROS;

_____MERGE _PERM.COPMACROS(DROP=M2)

_____ _DOCS.M2_CAST;

_____BY _DATE;

RUN;


PROC _ARIMA _DATA=PERM.COPMACROS;

_____TITLE_"PDEX_1st_Difference - no target";

_____IDENTIFY _VAR=PDEX(1) _CROSSCORR=(M2(1) _FEDFUNDS(1) _)

_____OUTCOV=_DOCS.PDEX_COVAR_1 _NLAG=24 _STATIONARITY=(PP=12);

_____ESTIMATE _P=12 _Q=0 _INPUT=(

_____0 ___$ _M2

_____1 ___$ _M2

_____2 ___$ _M2

_____3 ___$ _M2

_____4 ___$ _M2

_____5 ___$ _M2

_____6 ___$ _M2

_____7 ___$ _M2

_____8 ___$ _M2

_____9 ___$ _M2

_____10 ___$ _M2

_____11 ___$ _M2

```

```

.....12__$ _M2

.....0__$ _FEDFUNDS
.....1__$ _FEDFUNDS
.....2__$ _FEDFUNDS
.....3__$ _FEDFUNDS
.....4__$ _FEDFUNDS
.....5__$ _FEDFUNDS
.....6__$ _FEDFUNDS
.....7__$ _FEDFUNDS
.....8__$ _FEDFUNDS
.....9__$ _FEDFUNDS
.....10__$ _FEDFUNDS
.....11__$ _FEDFUNDS
.....12__$ _FEDFUNDS

.....)

outest=DOCS.PDEX_ESTS_1__$OUTMODEL=DOCS.PDEX_MODS_1
OUTSTAT=DOCS.PDEX_STATS_1;
.....FORECAST_LEAD=120_INTERVAL=MONTH_ID=DATE
.....OUT=DOCS.PDEX_RESULTS_1;
RUN;
QUIT;

DATA DOCS.PDEX_CAST (RENAME=(FORECAST=PDEX) );
.....SET DOCS.PDEX_RESULTS_1 (KEEP=DATE_FORECAST);
RUN;

```

```

* Now, we have predictions of all the macro variables;
DATA _PERM.COPMACROS(KEEP=DATE_FEDFUNDS_M2_PDEX);
    _MERGE _PERM.COPMACROS(WHERE=(SIM=1) _IN=A)
    _DOCS.PDEX_CAST;
    _BY DATE;
    _IF _A;
RUN;

PROC _EXPAND _DATA=_PERM.COPMACROS _OUT=_PERM.DIFFCOP;
    _CONVERT _FEDFUNDS=_FEDFUNDS_/_TRANSFORMOUT=(DIF1);
    _CONVERT _M2=_M2_/_TRANSFORMOUT=(DIF1);
    _CONVERT _PDEX=_PDEX_/_TRANSFORMOUT=(DIF1);
    _RUN;

DATA _PERM.DIFFCOP;
    _SET _PERM.DIFFCOP;
    _IF _N_=1 _THEN _DELETE;
RUN;

ODS_CSV_FILE="/scratch/jschule4/DATA/RAW/finsimmacro.csv";
PROC _PRINT _DATA=_PERM.DIFFCOP (DROP=TIME_DATE) _NOOBS;
RUN;

ODS_CSV_CLOSE;

```

```

ODS_CSV_FILE="/scratch/jschule4/DATA/RAW/prices.csv";
PROC_PRINT_DATA=PERM.AUTO_NOOBS;

****VAR_DATE_CODE_VALUE;

RUN;

ODS_CSV_CLOSE;


*Now that we have the simulated data,
we can begin prepping the simulation code;

*We need to run the error term program once per round;

FILENAME DELTAS
"/scratch/jschule4/DATA/RAW/predictions.csv";
FILENAME START"/scratch/jschule4/DATA/RAW/beginnings.csv";


DATA PERM.DELTA;

  INFILE DELTAS dlm=' ' dsd lrecl=4096 MISSOVER
  *****firstobs=1 termstr=LF;

  LENGTH IDX 8.
  *****VALUE 8.;
  ****INPUT IDX VALUE;
  ****K=0;


  ****TIME=FLOOR( _N_/381);

RUN;


DATA PERM.XINIT (KEEP=IDX_VALUE_K_TIME);
INFILE START dlm=' ' dsd lrecl=4096 MISSOVER
*****firstobs=1 termstr=LF;

```

```

LENGTH_IDX_8.
      VALUE_8.
      L1_8.
      L2_8.
      L3_8.
      L4_8.
      L5_8.
      L6_8.
      L7_8.
      L8_8.
      L9_8.
      L10_8.
      L11_8.
      L12_8.;

```

```

      K=_N_;
      TIME=0;
RUN;

```

```

DATA_PERM.XSTACK;
      SET_PERM.XINIT
      PERM.DELTA;
      RUN;

```

```

PROC_EXPAND_DATA=PERM.XSTACK_OUT=PERM.FDELTA;
      BY_TIME;
      CONVERT_VALUE=CUMSUM_/_TRANSFORMOUT=(CUSUM);

```

```

RUN;

PROC MEANS DATA=PERM.FDELTA NOPRINT;
  CLASS TIME;
  OUTPUT OUT=PERM.COV MEAN(CUMSUM)=MN STD(CUMSUM)=STDV;
RUN;

DATA PERM.COV;
  SET PERM.COV;
  COV=STDV/ABS(MN);
RUN;

TITLE "Coefficient of Variation";
ODS PDF FILE="/scratch/jschule4/plot.pdf"
  STYLE=STATISTICAL;
ODS HTML FILE="/scratch/jschule4/plot.html"
  STYLE=STATISTICAL;
PROC SGPLOT DATA=PERM.COV;
  LABEL TIME="Time Step"
  COV="Coefficient";
  SERIES X=TIME Y=COV;
RUN;

ODS HTML CLOSE;
ODS PDF CLOSE;
TITLE "";

ODS CSV FILE="/scratch/jschule4/DATA/RAW/cov.csv";
PROC PRINT DATA=PERM.COV NOOBS; RUN;

```

```
ODS_CSV_CLOSE;
```

```
PROC_TRANPOSE_DATA=PERM.FDELTA_OUT=PERM.TDELT;
```

```
____BY_TIME;
```

```
____VAR_CUMSUM;
```

```
____ID_IDX;
```

```
RUN;
```

```
ODS_CSV_FILE="/scratch/jschule4/DATA/RAW/simsum.csv";
```

```
PROC_PRINT_DATA=PERM.TDELT_NOOBS; _RUN;
```

```
ODS_CSV_CLOSE;
```

Appendix 7: Python Code for Full Model

```
#           Code to Model the Error Term  
#           February 2020  
#           John S. Schuler.
```

```
# Import librarys
```

```
import networkx as nx
```

```
import numpy as np
```

```
import scipy as sp
```

```
import skpp
```

```
import pandas as pd
```

```
import random as R
```

```
import multiprocessing
```

```

import pickle as PK
import sys
import copy
import datetime

reps=1
R.seed(78)

cmdargs = sys.argv

# Now convert this into a number. We start the simulation
# at the 13th place and the simulated macro data is 120 long
start=0
end=13


varbs=382


R.seed(525)

# Multithreading note: The main process is single-threaded.
# The projection pursuit estimation may be farmed
# out to another thread.


# now import residuals

resdat=\
pd.read_csv("/scratch/jschule4/DATA/RAW/resids1.csv",\
na_values=['.','])

# and import the distance correlation matrix
dcor=\

```

```

pd.read_csv("/scratch/jschule4/DATA/RAW/dependencies.csv")

# and import the dependency information

# of the macroeconomic| variables

mvars=\

pd.read_csv("/scratch/jschule4/DATA/RAW/macrobool.csv")


# We also need the simulated macro data

simmacro=\

pd.read_csv("/scratch/jschule4/DATA/RAW/finsimmacro.csv")


prices=pd.read_csv("/scratch/jschule4/DATA/RAW/prices.csv")


# Now, test import


resdat=resdat[\
[ 'WPU01110104',\
'WPU01130113',\
'WPU01130211',\
'WPU01130212',\
'WPU01130213',\
'WPU01130214',\
'WPU01130215',\
'WPU01130216',\
'WPU01130217',\
'WPU01130218',\
'WPU01210101',\

```

'WPU01210102' ,\
'WPU01210103' ,\
'WPU01210104' ,\
'WPU01220101' ,\
'WPU01220205' ,\
'WPU01510101' ,\
'WPU01610102' ,\
'WPU01710703' ,\
'WPU01810101' ,\
'WPU01830131' ,\
'WPU01PLUS02' ,\
'WPU02230101' ,\
'WPU02450202' ,\
'WPU02550201' ,\
'WPU02610101' ,\
'WPU02610103' ,\
'WPU02630103' ,\
'WPU02640101' ,\
'WPU02780109' ,\
'WPU02940301' ,\
'WPU03810623' ,\
'WPU03810645' ,\
'WPU03810646' ,\
'WPU03T15M05' ,\
'WPU03THRU15' ,\
'WPU05320104' ,\
'WPU05320105' ,\
'WPU05720201' ,\

'WPU05720301 ' ,\
'WPU05730201 ' ,\
'WPU06130213 ' ,\
'WPU06220209 ' ,\
'WPU06220407 ' ,\
'WPU06371801 ' ,\
'WPU06380202 ' ,\
'WPU0652013A ' ,\
'WPU06710402 ' ,\
'WPU07120105 ' ,\
'WPU09130291 ' ,\
'WPU09140551 ' ,\
'WPU09150337 ' ,\
'WPU09150441 ' ,\
'WPU10720122 ' ,\
'WPU10750103 ' ,\
'WPU10810231 ' ,\
'WPU10890424 ' ,\
'WPU10890507 ' ,\
'WPU10890701 ' ,\
'WPU11382611 ' ,\
'WPU11382631 ' ,\
'WPU11382651 ' ,\
'WPU11430406 ' ,\
'WPU11720501 ' ,\
'WPU11790551 ' ,\
'WPU11930700 ' ,\
'WPU11941301 ' ,\

'WPU12220325 ' ,\
'WPU12310101 ' ,\
'WPU12410220 ' ,\
'WPU12680101 ' ,\
'WPU13210121 ' ,\
'WPU13330101 ' ,\
'WPU13440131 ' ,\
'WPU13450101 ' ,\
'WPU14120514 ' ,\
'WPU30160107 ' ,\
'WPU33110402 ' ,\
'WPUFD413111 ' ,\
'WPUFD413112 ' ,\
'WPUFD413121 ' ,\
'WPUFD413122 ' ,\
'WPUSI004011 ' ,\
'WPUSI006111 ' ,\
'WPUSI012011 ' ,\
'WPUSI013011 ' ,\
'WPUSI017011 ' ,\
'WPUSI018011 ' ,\
'WPUSI019011 ' ,\
'WPUSI020011 ' ,\
'WPUSI021011 ' ,\
'WPUSI021012 ' ,\
'WPUSI023011 ' ,\
'WPUSI023012 ' ,\
'WPUSI024011 ' ,\

'WPUSI024012' ,\
'WPUSI070011' ,\
'WPUSI089011' ,\
'WPUSI092011' ,\
'WPUSI093011' ,\
'WPUSI094011' ,\
'WPUSI095011' ,\
'WPUSI134011' ,\
'WPUSISTEEL1' ,\
'WPUSISTEEL2' ,\
'WPUSISTEEL3' ,\
'WPU10150211' ,\
'WPU02420305' ,\
'WPU02620609' ,\
'WPU06790302' ,\
'WPU10150501' ,\
'WPU11420221' ,\
'WPU13312101' ,\
'WPU13313101' ,\
'WPU03380301' ,\
'WPU10790354' ,\
'WPU13920201' ,\
'WPU13314101' ,\
'WPU02610105' ,\
'WPU04410132' ,\
'WPU09470102' ,\
'WPU09470202' ,\
'WPU09470203' ,\

'WPU09470302' ,\
'WPU09470402' ,\
'WPU09470506' ,\
'WPU09470603' ,\
'WPU09470609' ,\
'WPU09471101' ,\
'WPU10170710' ,\
'WPU10810206' ,\
'WPU10880101' ,\
'WPU10880701' ,\
'WPU11680111' ,\
'WPU11782890' ,\
'WPU11784901' ,\
'WPU13990101' ,\
'WPU13990111' ,\
'WPU13990121' ,\
'WPU02310301' ,\
'WPU14120513' ,\
'WPU02310302' ,\
'WPU02310303' ,\
'WPU02310304' ,\
'WPU02310401' ,\
'WPU02310501' ,\
'WPU02840102' ,\
'WPU02840104' ,\
'WPU02850109' ,\
'WPU02850111' ,\
'WPU02850113' ,\

'WPU09141104 ' ,\
'WPU09141105 ' ,\
'WPU09220125 ' ,\
'WPU10260301 ' ,\
'WPU10890425 ' ,\
'WPU11140611 ' ,\
'WPU11140711 ' ,\
'WPU11490201 ' ,\
'WPU11490202 ' ,\
'WPU11490203 ' ,\
'WPU11490204 ' ,\
'WPU11490205 ' ,\
'WPU11490209 ' ,\
'WPU07210606 ' ,\
'WPU02112103 ' ,\
'WPU02112104 ' ,\
'WPU02120301 ' ,\
'WPU02340201 ' ,\
'WPU02540104 ' ,\
'WPU02540105 ' ,\
'WPU02550301 ' ,\
'WPU02550302 ' ,\
'WPU06210201 ' ,\
'WPU06210301 ' ,\
'WPU06220206 ' ,\
'WPU06220299 ' ,\
'WPU06710401 ' ,\
'WPU06710403 ' ,\

'WPU06720103' ,\
'WPU07130604' ,\
'WPU07130606' ,\
'WPU07130608' ,\
'WPU08210112' ,\
'WPU08210132' ,\
'WPU08210142' ,\
'WPU08210152' ,\
'WPU08210162' ,\
'WPU08210183' ,\
'WPU10280207' ,\
'WPU10540211' ,\
'WPU10710515' ,\
'WPU11350243' ,\
'WPU11350501' ,\
'WPU113A0201' ,\
'WPU11490208' ,\
'WPU11490534' ,\
'WPU11490535' ,\
'WPU11490537' ,\
'WPU11490538' ,\
'WPU13130116' ,\
'WPU13130117' ,\
'WPU11730901' ,\
'WPU02350201' ,\
'WPU03810407' ,\
'WPU06790402' ,\
'WPU06790501' ,\

'WPU06790502' ,\
'WPU06790918' ,\
'WPU09150216' ,\
'WPU09150218' ,\
'WPU09150322' ,\
'WPU10270111' ,\
'WPU10740803' ,\
'WPU10890521' ,\
'WPU10890522' ,\
'WPU11411911' ,\
'WPU11450108' ,\
'WPU11470143' ,\
'WPU11470144' ,\
'WPU11670502' ,\
'WPU12220407' ,\
'WPU12220409' ,\
'WPU12670141' ,\
'WPU15410701' ,\
'WPU15420601' ,\
'WPU113A0101' ,\
'WPU02350301' ,\
'WPU02350303' ,\
'WPU02130201' ,\
'WPU03390101' ,\
'WPU06790919' ,\
'WPU08120311' ,\
'WPU08120401' ,\
'WPU08490901' ,\

'WPU11410701 ' ,\
'WPU11440212 ' ,\
'WPU11440216 ' ,\
'WPU11440601 ' ,\
'WPU11440602 ' ,\
'WPU11950501 ' ,\
'WPU12110102 ' ,\
'WPU12110104 ' ,\
'WPU12150112 ' ,\
'WPU12690101 ' ,\
'WPU12690102 ' ,\
'WPU13620121 ' ,\
'WPU14150101 ' ,\
'WPU14160101 ' ,\
'WPU14160201 ' ,\
'WPU12680102 ' ,\
'WPU01710701 ' ,\
'WPU01710702 ' ,\
'WPU01710704 ' ,\
'WPU01710705 ' ,\
'WPU04410112 ' ,\
'WPU06130271 ' ,\
'WPU08610103 ' ,\
'WPU08610104 ' ,\
'WPU09150636 ' ,\
'WPU09150999 ' ,\
'WPU10890811 ' ,\
'WPU11410801 ' ,\

'WPU11440485' ,\
'WPU11450201' ,\
'WPU11790103' ,\
'WPU11790104' ,\
'WPU12230102' ,\
'WPU13950111' ,\
'WPU15120182' ,\
'WPU08610102' ,\
'WPU09150901' ,\
'WPU09220131' ,\
'WPU02140907' ,\
'WPU02540107' ,\
'WPU03460102' ,\
'WPU03460103' ,\
'WPU03470105' ,\
'WPU03830321' ,\
'WPU03940101' ,\
'WPU04450101' ,\
'WPU04450111' ,\
'WPU05760303' ,\
'WPU05760401' ,\
'WPU06790904' ,\
'WPU06790961' ,\
'WPU06790999' ,\
'WPU08420101' ,\
'WPU08710101' ,\
'WPU08710102' ,\
'WPU09480104' ,\

'WPU10410311 ',\n
'WPU10830522 ',\n
'WPU10830524 ',\n
'WPU10890564 ',\n
'WPU10890589 ',\n
'WPU11790512 ',\n
'WPU12210112 ',\n
'WPU12210113 ',\n
'WPU12210114 ',\n
'WPU12210116 ',\n
'WPU13990209 ',\n
'WPU13990211 ',\n
'WPU13990299 ',\n
'WPU14911105 ',\n
'WPU15110152 ',\n
'WPU15110154 ',\n
'WPU15110155 ',\n
'WPU15930113 ',\n
'WPU15930117 ',\n
'WPU60110501 ',\n
'WPU02120401 ',\n
'WPU05810319 ',\n
'WPU02890151 ',\n
'WPU02890161 ',\n
'WPU02890172 ',\n
'WPU02890175 ',\n
'WPU02940202 ',\n
'WPU03470108 ',\n

'WPU03810441 ' ,\
'WPU03830351 ' ,\
'WPU09480201 ' ,\
'WPU11760121 ' ,\
'WPU11760141 ' ,\
'WPU11760301 ' ,\
'WPU11760302 ' ,\
'WPU11760303 ' ,\
'WPU11792902 ' ,\
'WPU11840102 ' ,\
'WPU11840103 ' ,\
'WPU13450199 ' ,\
'WPU15120103 ' ,\
'WPU15120127 ' ,\
'WPU15120193 ' ,\
'WPU15950210 ' ,\
'WPU15950307 ' ,\
'WPU15970501 ' ,\
'WPU15970502 ' ,\
'WPU15970503 ' ,\
'WPU159A0402 ' ,\
'WPU159A0404 ' ,\
'WPU33310101 ' ,\
'WPU15320100 ' ,\
'WPU02630313 ' ,\
'WPU02940203 ' ,\
'WPU08210122 ' ,\
'WPU11621201 ' ,\

'WPU11627701 ' ,\
'WPUIP230000 ' ,\
'WPUIP231000 ' ,\
'WPUIP231100 ' ,\
'WPUIP232000 ' ,\
'WPUIP232100 ' ,\
'WPUIP232200 ' ,\
'WPU02890102 ' ,\
'WPU02930102 ' ,\
'WPU02930118 ' ,\
'WPU10230101 ' ,\
'WPU10230102 ' ,\
'WPU10230104 ' ,\
'WPU10230201 ' ,\
'WPU10230205 ' ,\
'WPU10230206 ' ,\
'WPU10260314 ' ,\
'WPU11440378 ' ,\
'WPU08110503 ' ,\
'WPU0613020T ' ,\
'WPU06140399 ' ,\
'WPU06220298 ' ,\
'WPU09150301 ' ,\
'WPU61110201 ' ,\
'WPU02440102 ' ,\
'WPU02440127 ' ,\
'WPU02440139 ' ,\
'WPU09130321 ' ,\

'WPU09130322 ', \
 'WPU02630104 ', \
 'WPU02210579 ', \
 'WPU10260333 ', \
 'WPU12140114 ', \
 'WPU33410101 ', \
 'WPU08610101 ', \
 'FEDFUNDS ', \
 'FEDFUNDS_L_1 ', \
 'FEDFUNDS_L_2 ', \
 'FEDFUNDS_L_3 ', \
 'FEDFUNDS_L_4 ', \
 'FEDFUNDS_L_5 ', \
 'FEDFUNDS_L_6 ', \
 'FEDFUNDS_L_7 ', \
 'FEDFUNDS_L_8 ', \
 'FEDFUNDS_L_9 ', \
 'FEDFUNDS_L_10 ', \
 'FEDFUNDS_L_11 ', \
 'FEDFUNDS_L_12 ', \
 'M2 ', \
 'M2_L_1 ', \
 'M2_L_2 ', \
 'M2_L_3 ', \
 'M2_L_4 ', \
 'M2_L_5 ', \
 'M2_L_6 ', \
 'M2_L_7 ', \

```

'M2_L_8 ',\
'M2_L_9 ',\
'M2_L_10 ',\
'M2_L_11 ',\
'M2_L_12 ',\
'PDEX ',\
'PDEX_L_1 ',\
'PDEX_L_2 ',\
'PDEX_L_3 ',\
'PDEX_L_4 ',\
'PDEX_L_5 ',\
'PDEX_L_6 ',\
'PDEX_L_7 ',\
'PDEX_L_8 ',\
'PDEX_L_9 ',\
'PDEX_L_10 ',\
'PDEX_L_11 ',\
'PDEX_L_12 ' ] ]

```

```

fulldat=resdat

```

```

mvars.rename(columns={"Unnamed: 0": "VARB"}, inplace=True)

```

```

macrodat=resdat.iloc[:, -1: -39]

```

```

resdat=resdat.iloc[:, 0:382]

```

```

print (" Simulated_Data")

```

```

print (simmacro)

```

```

# We will use objects containing the column name as nodes

class colnode:
    '''
    An object to represent each column.
    '''

    node_obj_list = []

    def __init__(self, name, macrodeps):
        self.name = name
        self.deg = None
        self.macrodeps = macrodeps

        # We set a lock on each node so the
        # agents cannot run into each other.
        # We also lock the node once we provide it a value.
        self.lock = False
        self.value = None
        self.models = []
        colnode.node_obj_list.append(self)

Depgraph = nx.Graph()

namedict = {}
revnamedict = {}

for col in resdat.columns:
    #for col in resdat.columns[0:139]:
        deps = mvars[mvars['VARB'] == col]
        namedict[col] = colnode(col, deps)

```

```

    revnamedict[namedict[col]]=col

for node in colnode.node_obj_list:
    Depgraph.add_node(node)

# Now, we can add nodes
for i in range(dcor.shape[0]):
    for j in range(i,dcor.shape[0]):
        if i!=j:
            if dcor.iloc[i,j] !=0:
                Depgraph.add_edge(\
namedict[resdat.columns[i]],namedict[resdat.columns[j]],\
dcor=dcor.iloc[i,j])

# Now that the graph is complete,
# we can add to each node some information
for node in nx.nodes(Depgraph):
    node.deg=Depgraph.degree(node) /\
    Depgraph.number_of_nodes()

# Now, we need to make some decisions regarding the agents
# Do they try to add values to the
# highest degree nodes first?

```

```

# or to the ones with the highest distance correlation with
# the current node?

# we can give each agent a utility alpha where it goes for
# the node that maximizes the weighted average
# of these quantities

```

```

# Now, let's play with edges

```

```

class agent:

```

```

    '''

```

```

    These agents walk around the network deciding what
    models to run and what predictions to make

```

```

    '''

```

```

    agt_list=[]

```

```

    retired_list=[]

```

```

    def __init__(self,alpha,node):

```

```

        self.alpha=alpha

```

```

        self.node=node

```

```

        self.hist=[]

```

```

        self.backs=[]

```

```

        agent.agt_list.append(self)

```

```

    def util(self):

```

```

        '''

```

```

        Have the agent determine how much

```

```

        it wants to jump to each node.

```

```

        '''

```

```

    # What are the neighbors of the current node?

```

```

    edges=Depgraph.edges(self.node)

```

```

adjdict={}
for e in edges:
    # The agent does not consider locked nodes
    if not e[1].lock:
        adjdict[e[1]]=\
self.alpha*Depgraph.edges[e]['dcor']+\\
(1-self.alpha)*e[1].deg
    return(adjdict)
def jump(self):
    '''
    Now, have the agent jump to the unlocked node
    it most prefers. If there is no unlocked node,
    it teleports to an unlocked node
    '''
    mx=0
    choice=None
    options=self.util()
    if len(options.keys())==0:
        # find a list of unlocked nodes
        unlocked=[]
        for n in colnode.node_obj_list:
            if n.lock==False:
                unlocked.append(n)

        # First, if we failed to set a value,
        # then unlock the node
        if self.node.value==None:
            self.node.lock=False

```

```

        if len(unlocked) > 0:
            select=R.choice(list(range(len(unlocked))))
            self.hist.append(select)
            self.node=unlocked[select]
            self.node.lock=True
        else:
            agent.retired_list.append(self)
            agent.agt_list.remove(self)

    else:
        if self.node.value==None:
            self.node.lock=False
        for k in options.keys():
            if options[k] > mx:
                mx=options[k]
                choice=k
            self.hist.append(choice)
            self.node=choice
            choice.lock=True

def eval(self):
    '''
    Now, the agent checks to see if there are valued
    nodes adjacent to its current node.
    '''
    neighbors=Depgraph.neighbors(self.node)
    vlist=[]
    for n in neighbors:
        if n.value != None:

```

```

        vlist.append(n)

    if len(vlist) > 0:
        return True
    else:
        return False

# Now, we want a function that estimates a given node in
# terms of the values of other nodes. This is the part
# that should eventually be parallelized
# We also need a dictionary of dictionaries to store
# these models for re-use
model_dict={}

# Now we need a model object
class model:
    '''
    This object encapsulates the model along with its data.
    '''
    mod_list=[]
    def __init__(self,node,r,degree):
        self.node=node
        self.r=r
        self.degree=degree
        self.mod_neighbors=[]
        # Now we can prepare the data.
        # What nodes are we connected to?
        model.mod_list.append(self)

```

```

neighbors=Depgraph.neighbors( self.node)

# Now, which have values?

self.valued=[]

for n in neighbors:

    if n.value != None:

        self.valued.append(n)

# Now, turn these into columns

cols=[]

for n in self.valued:

    cols.append(revnamedict[n])

ycol=revnamedict[ self.node]

# Now, we have to deal with the

# possibility of missing data

# Now, finally , we need the macro columns

mcols=[]

if list( self.node.macrodeps["FEDFUNDS"])[0]:

    mcols=mcols+[ 'FEDFUNDS', 'FEDFUNDS_L_1 ',\

                    'FEDFUNDS_L_2 ', 'FEDFUNDS_L_3 ',\

                    'FEDFUNDS_L_4 ', 'FEDFUNDS_L_5 ',\

                    'FEDFUNDS_L_6 ', 'FEDFUNDS_L_7 ',\

                    'FEDFUNDS_L_8 ', 'FEDFUNDS_L_9 ',\

                    'FEDFUNDS_L_10 ', 'FEDFUNDS_L_11 ',\

                    'FEDFUNDS_L_12 ' ]

if list( self.node.macrodeps["M2"])[0]:

    mcols=mcols+[ 'M2', 'M2_L_1 ', 'M2_L_2 ', 'M2_L_3 ',\

                    'M2_L_4 ', 'M2_L_5 ', 'M2_L_6 ',\

                    'M2_L_7 ', 'M2_L_8 ', 'M2_L_9 ',\

                    'M2_L_10 ', 'M2_L_11 ', 'M2_L_12 ' ]

```

```

if list(self.node.macrodets["PDEX"])[0]:
    mcols=mcols+['PDEX', 'PDEX_L_1', 'PDEX_L_2',\
        'PDEX_L_3', 'PDEX_L_4', 'PDEX_L_5',\
        'PDEX_L_6',\
        'PDEX_L_7', 'PDEX_L_8',\
        'PDEX_L_9',| 'PDEX_L_10',\
        'PDEX_L_11', 'PDEX_L_12']

while True:
    working_dat=fulldat [[ ycol]+cols+mcols]
    popdat=working_dat.dropna(0, 'any', inplace=False)

    if popdat.shape[0] >= 50:
        break

    else:
        for col in cols:

            reduce_dt=working_dat[col]
            popcol=reduce_dt.dropna(0, 'any',\
                inplace=False)

            if popcol.shape[0] < 50:
                cols.remove(col)

            break

# Now, we can generate the list of nodes
# on which this model depends
for c in cols:
    self.mod_neighbors.append(namedict[c])

# Now, we can split it
ydat=popdat.iloc[:,0:1]
xdat=popdat.iloc[:,1:popdat.shape[1]]

```

```

try:

    self.mod=skpp.ProjectionPursuitRegressor(\
        r=self.r, fit_type='spline',\
        show_plots=False, out_dim_weights="uniform",\
        degree=self.degree, stage_maxiter=100,\
        backfit_maxiter=25)

    self.fit=self.mod.fit(xdat,ydat)

    self.preds=self.fit.predict(xdat)

    ydat=np.array(ydat).reshape(self.preds.shape)

    model.res=ydat-self.preds

    self.retval=True

except:

    self.retval=False


def predict(self,X):

    '''

    Predict a Y

    '''

    #print("X")

    #print(X)

    y_hat=self.fit.predict(X)

    Rdex=R.choice(list(range(X.shape[0])))

    res=self.res[Rdex]

    outp=y_hat

    return (outp[0],y_hat[0],res)

def ObjPredict(self,start,end):

    '''

```

```

This function wraps around the predict function
# to take in the neighbor objects in this model
and turn them into a numpy array
'''

nvals=[]

for n in self.mod_neighbors:
    nvals.append(n.value)

# We also need to add the macrovariables
deps=np.array(self.node.macrodets.iloc[0,1:4])
# Now, turn deps into a list of numbers
# We can pull a parameter from the shell instructing
# the program what row to use in the
# macro-variable data

macros=simmacro.iloc[start:end,deps]

flat=list(np.array(macros).flatten('F'))

return(self.predict(np.array([nvals+flat])))

# Now, we need a few tuning parameters
# What percentage of the data do we seed?

seedpct=.05

def SeedRand():
    '''

    Select random columns and seed nodes with the values.

    '''

    cnt=int(seedpct*len(colnode.node_obj_list))

```

```

sampnodes=R.sample(colnode.node_obj_list , cnt )

# Now generate a list of columns

cols=[]

for node in sampnodes:

    cols.append(revnamedict [node])


currdt=resdat [ cols ]

# Now remove missing data

popdat=currdt.dropna(0, 'any', inplace=False)

idx=R.choice( list (range(popdat.shape [0])) )


for j in range(len(popdat.columns)):

    namedict [popdat.columns [j]].value=popdat.iloc [idx ,j]

    namedict [popdat.columns [j]].lock=True


def AgtRand( agt_cnt ):

    '''

    This function randomly generates agents and assigns

agents to nodes. Agents only go to

nodes adjacent to valued nodes.

    '''

    # We first need a list of the valued nodes

    availist=[]

```

```

for node in colnode.node_obj_list:
    if not node.value==None:
        availist.append(node)
# and then a list of lists of the unvalued neighbors
# of the valued nodes
    neighborlist=[]
for node in availist:
    lst=[]
    for n in Depgraph.neighbors(node):
        if n.value == None:
            lst.append(n)
        neighborlist.append(lst)
# Now, turn this list of lists into a deduped list
    base=set()
for lst in neighborlist:
        base=base.union(lst)

    base=list(base)

targnodes=R.sample(availist, agt_cnt)
for i in range(agt_cnt):
    agent(R.uniform(0,1), base[i])
    base[i].lock=True

# Now we begin the main process

```

```

def Process():
    '''
    This function generates a realization of the error term.
    It requires information on where in the input
    macro process we are.
    '''
    unvalhist=[]
    # Step 1: initialize
    SeedRand()
    AgtRand(10)
    tstlist=[]
    for n in colnode.node_obj_list:
        if n.value != None:
            tstlist.append(n)
    t=0
    while True:
        t=t+1
        # Find any nodes that are unvalued but locked
        unvalued=[]
        for node in colnode.node_obj_list:
            if node.value==None:
                unvalued.append(node)
        unvalhist.append(len(unvalued))
        if t >=3:
            if unvalhist[-1]==unvalhist[-2]:
                # find the columns we can't value and
                # pull a random error from the data
                for node in unvalued:

```

```

        # Find the non-null residuals
        node.value=resdat[\
            resdat[revnamedict[node]].notna()][\
            revnamedict[node].sample(1).iloc[0]

        break

    if len(unvalued)==0:
        break

    for agt in agent.agt_list:
        # each agents checks whether it wants to jump
        if not agt.eval():
            # If we could not fit a model, jump
            agt.jump()
        else:
            # else, if we did fit a model, we set
            # the node value to its prediction
            # jump, and leave the node locked.
            M=model(agt.node,10,1)
            if M.retval:
                out=M.ObjPredict(start,end)
                agt.node.value=out[0]

            agt.jump()

    # Now, we gather up the estimates and output them.
    nodelist=[]

    for node in colnode.node_obj_list:
        nodelist.append(node.value)

```

```

    return np.array(modelist)

# Now, we can run the simulation

# First, load X

timeln=50

R.seed(25252)

Beta=pd.read_csv("/scratch/jschule4/DATA/RAW/betas.csv")

X=pd.read_csv(\

"/scratch/jschule4/DATA/RAW/X.csv",na_values=['.']);

# Now, we rearrange them

X=X[["FEDFUNDS",\

    "FEDFUNDS_L_1",\

    "FEDFUNDS_L_2",\

    "FEDFUNDS_L_3",\

    "FEDFUNDS_L_4",\

    "FEDFUNDS_L_5",\

    "FEDFUNDS_L_6",\

    "FEDFUNDS_L_7",\

    "FEDFUNDS_L_8",\

    "FEDFUNDS_L_9",\

    "FEDFUNDS_L_10",\

    "FEDFUNDS_L_11",\

    "FEDFUNDS_L_12",\

    "M2",\

    "M2_L_1",\

    "M2_L_2",\

    "M2_L_3",\

```

"M2_L_4" ,\
 "M2_L_5" ,\
 "M2_L_6" ,\
 "M2_L_7" ,\
 "M2_L_8" ,\
 "M2_L_9" ,\
 "M2_L_10" ,\
 "M2_L_11" ,\
 "M2_L_12" ,\
 "PDEX" ,\
 "PDEX_L_1" ,\
 "PDEX_L_2" ,\
 "PDEX_L_3" ,\
 "PDEX_L_4" ,\
 "PDEX_L_5" ,\
 "PDEX_L_6" ,\
 "PDEX_L_7" ,\
 "PDEX_L_8" ,\
 "PDEX_L_9" ,\
 "PDEX_L_10" ,\
 "PDEX_L_11" ,\
 "PDEX_L_12" ,\
 "VALUE" ,\
 "L1" ,\
 "L2" ,\
 "L3" ,\
 "L4" ,\
 "L5" ,\

```

    "L6" , \
    "L7" , \
    "L8" , \
    "L9" , \
    "L10" , \
    "L11" , \
    "L12" , \
    "DATE" , \
    "CODE" ]]

X=X[X[ 'CODE' ] !=np.nan]

Beta=Beta.iloc[:,3:]

X=X[[ "VALUE" ,"L1" ,"L2" ,"L3" ,"L4" ,"L5" ,"L6" ,"L7" ,"L8" ,"L9" , \
    "L10" ,"L11" ,"L12" ,"DATE" ,"CODE" ]]

# Now, we need to select a beginning point.
# Let's just start at the first X and take
# the first 13 differences
sasbase=datetime.date(1960, 1, 1)

def intnx(dt,delta):
    '''
    This function takes a Python datetime object and
emulates advances it k months in a manner
similar to the SAS
INTNX function.
    '''
    # First, what month are we currently at?

```

```

newdt=dt
if delta >= 0:
    for j in range(delta):
        if newdt.month==12:
            newdt=\
            newdt.replace(month=1,year=newdt.year+1)
        else:
            newdt=newdt.replace(month=newdt.month+1)
    else:
        delta=12+delta
        newdt=newdt.replace(year=newdt.year-1)
        for j in range(delta):
            if newdt.month==12:
                newdt=\
                newdt.replace(month=1,year=newdt.year+1)
            else:
                newdt=newdt.replace(month=newdt.month+1)
    return newdt

# Now, let's test this
tstdate=datetime.date(2020,1,1)

tst2=intnx(tstdate,4)

def SASDate(dt):
    '''
    This function converts a SAS date into a python date.
    '''

```

```

    print("SASDATE")

    print(dt)

    return datetime.date(1960, 1, 1)+\
        datetime.timedelta(days=int(dt))
def invSASDate(sdt):
    '''
    This function is the inverse of the above.
    '''

    return (sdt-datetime.date(1960,1,1)).days

```

```
tst2=SASDate(-2010)
```

```
tst3=SASDate(21946)
```

```
# Now, we need a unique list of dates
```

```
dates=pd.unique(X["DATE"])
```

```
begdt=R.choice(dates.astype(int))
```

```
# Now, we need the previous 12 dates also;
```

```
# Now that we have a date range, we can subset the data
```

```
X=X[X.DATE.isin([begdt])]
```

```
X=X[X["CODE"].notna()]
```

```

# Now, we need to join in the synthetic macro variables

# Now, we reverse it
simmacro=simmacro.iloc[::-1]

# Now, we need the last 13 rows
last=simmacro.iloc[-13:,:]

last=last[["FEDFUNDS","M2","PDEX"]]

flat=np.array(last).flatten('F')

# Now, we need the number of codes so
# we can combine them into a dataframe
codes=pd.unique(X["CODE"])

maclist=[]
for i in range(len(codes)):
    maclist.append(flat)
macarray=np.array(maclist)

# Now we need a numpy array of our starting data
X=X[["VALUE","L1","L2","L3","L4","L5","L6","L7","L8","L9",\
"L10","L11","L12"]]

X.to_csv("/scratch/jschule4/DATA/RAW/beginnings.csv",\
mode='a',header=False)

Xsum=np.sum(np.array(X),1)

```

```

# Now, let's joint the numpy arrays
Z=np.concatenate((macarray,np.array(X)),1)

NBeta=np.array(Beta)

def NAmult(x,b):
    '''
    This implements matrix multiplication with a the
    special rule that 0 * NA=0.
    # '''
    hld=[]
    for i in range(x.shape[0]):
        #print("I")
        #print(i)
        if b[i]==0 and x[i]==np.nan:
            hld.append(0)
        elif b[i] != 0 and x[i]==np.nan:
            hld.append(np.nan)
        else:
            hld.append(x[i]*b[i])
    return(np.sum(np.array(hld)))

# Now, for each code, find a dot product
res=[]

prices=prices[prices.DATE.isin(\
[invSASDate(intnx(SASDate(begdt),-1))])]

```

```

sval=np.array( prices["VALUE"] )

init=Xsum+sval

predlist=[np.array(X["VALUE"])]

for t in range(120):


    # second, get the range of simulated macro variables
    last=simmacro.iloc[-(t+14):- (t+1),:]

    flat=np.array(last).flatten('F')


    print( flat )
    print( flat.shape )
    maclist=[]
    for i in range(len(codes)):
        maclist.append( flat )
    macarray=np.array( maclist )


    Z=np.concatenate(( macarray , np.array(X) ),1)


    res=[]
    for i in range(Z.shape[0]):

```

```

        #print(i)
        res.append(NAmult(Z[i,:],NBeta[:,i]))
res=np.array(res)

# Now generate the error term

err=Process()

#print("Prediction")
pred=res+err
# Now write this out to a file
pd.DataFrame(pred).to_csv(
"/scratch/jschule4/DATA/RAW/predictions.csv",\
mode='a',header=False)

# Now, what is too big?

predlist.append(pred)

# Now that we have the prediction,
# let's get the new difference

new_delta=predlist[-1]-predlist[-2]
#print(new_delta)
init=pred

```

```

# Now, add the prediction to X
X[ 'L12' ]=X[ 'L11' ]
X[ 'L11' ]=X[ 'L10' ]
X[ 'L10' ]=X[ 'L9' ]
X[ 'L9' ]=X[ 'L8' ]
X[ 'L8' ]=X[ 'L7' ]
X[ 'L7' ]=X[ 'L6' ]
X[ 'L6' ]=X[ 'L5' ]
X[ 'L5' ]=X[ 'L4' ]
X[ 'L4' ]=X[ 'L3' ]
X[ 'L3' ]=X[ 'L2' ]
X[ 'L2' ]=X[ 'L1' ]
X[ 'L1' ]=X[ 'VALUE' ]
X[ 'VALUE' ]=pd.Series(new_delta ,index=X.index)

#print("Index")
#print(X.index)
#print(pd.Series(pred).index)
#print(pd.Series(pred))
#print("New X")
#print(X)

# clear the nodes and agents
# Now, we clear the nodes

for node in colnode.node_obj_list :
    node.value=None
    node.lock=False

## and clear the agents

```

Appendix 8: R Code for Copula Estimator

```

agent.agt[1,1]

# Let's begin with a simplified model
# again, we begin with unifdat
# But we will add the marginals
for (i in 1:nrow(unifdat))
{

}

# continuing to use the same RV's
translist <- list()
for (a in seq(.1,3,.25))
{
  for (b in seq(.1,3,.25))
  {
    for (j in 1:(ncol(unifdat)-1))
    {
      for (k in 1:(ncol(unifdat)-1))
      {
        if (j==k){translist[[length(translist)+1]]
          <- pkumar(unifdat[,j],a,b)} else
          {translist[[length(translist)+1]]
            <- pkumar(unifdat[,j]*unifdat[,k],a,b)}
      }
    }
  }
}

```

```

    }
  }

  tdat <- log(data.frame(translist))

  logy <- log(unifdat$y)
  beta <- Variable(ncol(tdat))

  obj <- sum((logy - tdat %*% beta)**2)
  prob <- Problem(Minimize(obj), list(sum(beta**2) <= .5))
  result <- solve(prob)
  beta_hat <- result[[1]]
  y_hat <- data.matrix(tdat) %*% beta_hat
  data.frame(exp(y_hat), unifdat$y) -> wrk
  plot(wrk$exp.y_hat., wrk$unifdat.y)
  names(wrk) <- c('mod_y', 'emp_y')
  ggplot(wrk, aes(x=mod_y, y=emp_y)) + geom_point() +
    xlab("Model_Psi") + ylab("Empirical_Psi")

# Now, let 's actually incorporate the weight function
  subfunc <- function(x)
  {
    vec <- c()
    for (i in 1:length(x))
    {
      idx <- rep(TRUE, length(x))

```

```

      idx[i] <- FALSE
      vec <- c(vec,prod(x[idx]))
    }
    return(max(vec)**4)
  }

```

```

WeightFunc <- function(dat)
{
  return(apply(dat,1,subfunc))
}

```

```

unifdat$y -> y
unifdat[,1:5] -> X

```

```

y_trans <- log(y)
X_trans <- X

```

```

logy <- y_trans

```

```

translist <- list()
for (a in seq(.1,3,.25))
{
  for (b in seq(.1,3,.25))
  {
    for (j in 1:(ncol(unifdat)-1))
    {
      for (k in 1:(ncol(unifdat)-1))

```

```

    {
      if (j==k){translist[[length(translist)+1]]
        <- pkumar(unifdat[,j],a,b)} else
        {translist[[length(translist)+1]]
          <- pkumar(unifdat[,j]*unifdat[,k],a,b)}
    }
  }
}
plot(y,y_trans)

```

```

set.seed(5)

tdat <- log(data.frame(translist))
# Now, let's remove the validation set
nfit <- round(.8*nrow(tdat),0)
nsave <- nrow(tdat)-nfit
sample(1:nrow(tdat),nfit) -> fits
saves <- setdiff(1:nrow(tdat), fits)

```

```

fdat <- tdat[fits,]
sdat <- tdat[saves,]

```

```

yfit <- logy[fits]
y_true_save <- unifdat$y[saves]
y_true_fit <- unifdat$y[fits]
beta <- Variable(ncol(tdat))

```

```

obj <- sum((yfit - fdat %*% beta)**2)
prob <- Problem(Minimize(obj), list(sum(beta**2) <= 5))
result <- solve(prob)
beta_hat <- result[[1]]
#y_hat <- exp(data.matrix(tdat) %*% beta_hat -
WeightFunc(X)*log(apply(X,1,prod)))
y_hat_in <- exp(data.matrix(fdat) %*% beta_hat )
y_hat_out <- exp(data.matrix(sdat) %*% beta_hat )
data.frame(y_hat_in, y_true_fit) -> wrkin
data.frame(y_hat_out, y_true_save) -> wrkout
names(wrkin) <- c('mod_y', 'emp_y')
names(wrkout) <- c('mod_y', 'emp_y')
ggplot(wrkin, aes(x=mod_y, y=emp_y)) + geom_point() +
xlab(TeX('Model_\\hat{C}$')) +
ylab(TeX('Empirical_\\hat{C}$')) +
  geom_segment(aes(x = 0, y = 0, xend = 1, yend = 1 ),
  colour = "red") + ggtitle("In_Sample_Fit")

ggplot(wrkout, aes(x=mod_y, y=emp_y)) + geom_point() +
xlab(TeX('Model_\\hat{C}$')) +
ylab(TeX('Empirical_\\hat{C}$')) +
  geom_segment(aes(x = 0, y = 0, xend = 1, yend = 1 ),
  colour = "red") + ggtitle("Out_of_Sample_Fit")

```

Bibliography

Bibliography

- [1] A. Mandel, D. Taghawi-Nejad, and V. P. Veetil, “The price effects of monetary shocks in a network economy,” *Journal of Economic Behavior & Organization*, vol. 164, pp. 300–316, 2019.
- [2] N. S. Balke and M. A. Wynne, “The relative price effects of monetary shocks,” *Journal of Macroeconomics*, vol. 29, no. 1, pp. 19–36, 2007.
- [3] W. D. Lastrapes, “Estimating and identifying vector autoregressions under diagonality and block exogeneity restrictions,” *Economics letters*, vol. 87, no. 1, pp. 75–81, 2005.
- [4] —, “Inflation and the distribution of relative prices: the role of productivity and money supply shocks,” *Journal of Money, Credit and Banking*, pp. 2159–2198, 2006.
- [5] S. Athey, “Machine learning and causal inference for policy evaluation,” in *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2015, pp. 5–6.
- [6] S. Athey and G. W. Imbens, “Machine learning methods for estimating heterogeneous causal effects,” *stat*, vol. 1050, no. 5, pp. 1–26, 2015.
- [7] S. Athey, “The impact of machine learning on economics,” in *The economics of artificial intelligence: An agenda*. University of Chicago Press, 2018.
- [8] L. von Mises, *Human Action*. Liberty Fund, 2007, vol. 2.
- [9] G. J. Székely and M. L. Rizzo, “Energy statistics: A class of statistics based on distances,” *Journal of statistical planning and inference*, vol. 143, no. 8, pp. 1249–1272, 2013.
- [10] —, “The distance correlation t-test of independence in high dimension,” *Journal of Multivariate Analysis*, vol. 117, pp. 193–213, 2013.
- [11] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013, vol. 112.
- [12] B. B. Mandelbrot, “Heavy tails in finance for independent or multifractal price increments,” in *Handbook of heavy tailed distributions in finance*. Elsevier, 2003, pp. 1–34.
- [13] L. Breiman *et al.*, “Statistical modeling: The two cultures (with comments and a rejoinder by the author),” *Statistical science*, vol. 16, no. 3, pp. 199–231, 2001.

- [14] Y. Fan and A. J. Patton, “Copulas in econometrics,” *Annu. Rev. Econ.*, vol. 6, no. 1, pp. 179–200, 2014.
- [15] C. Czado, “Analyzing dependent data with vine copulas,” *Lecture Notes in Statistics*, Springer, 2019.
- [16] P. Jaworski, F. Durante, W. K. Hardle, and T. Rychlik, *Copula theory and its applications*. Springer, 2010, vol. 198.
- [17] H. Abelson, G. J. Sussman, and J. Sussman, *Structure and interpretation of computer programs*. Justin Kelly, 1996.
- [18] T. L. Heath *et al.*, *The thirteen books of Euclid’s Elements*. Courier Corporation, 1956.
- [19] J. Klein, *Greek Mathematical Thought and the Origin of Algebra*. Dover, 1992.
- [20] —, *Lectures and Essays*. St. John’s College Press, 1985.
- [21] F. v. Hayek, “Scientism and the study of society. part ii,” *Economica*, vol. 10, no. 37, pp. 34–63, 1943.
- [22] E. Schechter, *Handbook of Analysis and its Foundations*. Academic Press, 1996.
- [23] L. J. Savage, *The foundations of statistics*. Courier Corporation, 1972.
- [24] G. Boole, *An investigation of the laws of thought: on which are founded the mathematical theories of logic and probabilities*. Dover Publications, 1854.
- [25] E. T. Jaynes, *Probability theory: The logic of science*. Cambridge university press, 2003.
- [26] W. Briggs, *Uncertainty: the soul of modeling, probability & statistics*. Springer, 2016.
- [27] R. Frisch, *Problems and Methods of Econometrics: The Poincaré Lectures of Ragnar Frisch 1933*. Routledge, 2009.
- [28] F. Louçã, *The years of high econometrics: A short history of the generation that reinvented economics*. Routledge, 2007.
- [29] T. Haavelmo, “The probability approach in econometrics,” *Econometrica: Journal of the Econometric Society*, pp. iii–115, 1944.
- [30] —, “The statistical implications of a system of simultaneous equations,” *Econometrica, Journal of the Econometric Society*, pp. 1–12, 1943.
- [31] J. Pearl, “Trygve haavelmo and the emergence of causal calculus,” *Econometric Theory*, vol. 31, no. 1, pp. 152–179, 2015.
- [32] F. Hayashi, *Econometrics*. Princeton University Press Princeton, 2000.
- [33] L. von Mises, *Human Action*. Liberty Fund, 2007, vol. 1.
- [34] R. Axtell, “The complexity of exchange,” *The Economic Journal*, vol. 115, no. 504, 2005.

Curriculum Vitae

John Schuler graduated from St. John's College Annapolis in 2009 with a BA in liberal arts. He spent the next academic year doing post-baccalaureate work at the University of Maryland- College Park. He then embarked on a five year career at Fannie Mae beginning in 2010. While at Fannie Mae, he began a MS degree in statistics at the American University. In 2015, he left Fannie Mae and entered the PhD program in Economics at George Mason University; finishing the MS degree in 2017. The same year, he completed the MA degree in economics at George Mason University. In the Spring of 2017, he served as a visiting assistant professor at Washington College in Maryland. Since, he has served as an adjunct professor of statistics and data analysis at Southern New Hampshire University's College of Online and Continuing Education.