

The University of Tulsa

Disparities in the Sources of Price Inflation during the Covid-19 Pandemic and Non-Pandemic
Periods – the Distinctive Nature of Pandemic Inflationary Pressures

Dominic Lupo

Dr. Hendricks

Econ 4973 – Economics Senior Project

2, May 2022

Index

I. Introduction and Research Abstract	4-14
i. Research Question and Gaps in the Existing Body of Literature.....	4-7
ii. Data Sources and Methods Used.....	8-10
iii. Analysis Issue Areas.....	10-12
iv. Findings and Policy Implications.....	12-13
v. Outlined Structure of the Paper.....	13-14
II. Data Sources and Values.....	14-17
III. Methods.....	17-32
i. Empirical Model Description and Formulation.....	17-24
ii. Empirical Model Assumptions and Estimator Properties.....	24-25
iii. Empirical Model Limitations and Potential Sources of Bias.....	25-29
iv. Model Applicability: Addressing the Research Question.....	29-32
IV. Results.....	32-41
i. Summary Regression Table 1.....	33-36
ii. Summary Regression Table 2.....	36-40
iii. Results Review and Analysis.....	40-41

V. Policy Implications and Proposal.....	41-43
VI. Conclusion.....	43-44
VII. Works Cited.....	44-46

Disparities in the Sources of Price Inflation during the Covid-19 Pandemic and Non-Pandemic Periods – the Distinctive Nature of Pandemic Inflationary Pressures

I. Introduction

i. Research Question: What We Seek to Uncover and The Gaps in the Existing Body of Literature We Hope to Address

My broad research question seeks to discover if the inflation of goods and services prices has grown at the same rate during the Covid-19 pandemic compared to that of normal periods. That is, has inflation changed due to Covid-19? Due to the recency of the Covid-19 pandemic, the literature exploring the effects of the pandemic on inflation is still in its infancy. However, subject area research into the effects of the pandemic is growing at an increasing rate. Though the extraordinary constraints the pandemic has placed on the supply chain of goods and services and the way we interact with such products and people remain obvious, the quantitative effects these pandemic restrictions have had on inflation remain disputed. Notably, the very nature of the inflation being experienced by consumers, being transitory or more permanent, was in conflict only until recently.

Recent research suggests the Consumer Price Index (CPI) and Personal Consumption Expenditure (PCE) Index are in fact not reflecting the true nature of inflation occurring during the pandemic via analysis of debit card transactions (Seiler, 2020). However, using debit card transactions themselves as a measure of the increase in the pricing of goods runs the risk of distorting findings as increased consumption by consumers via debit cards, may reflect added stimulus in the economy. The most standard form of independent inflation analysis by contributors entails the process of item tracking, where specific goods and services, similar to the CPI and PCE basket of goods, are tracked for price increases over time (Van Hoomissen, 1988).

Yet, this more common method also presents downsides, namely, the act of selecting certain goods and services themselves may reflect the biases of an author and may not be indicative of the most often consumed goods and services that present the biggest inflation impact to consumers. Further, such an approach may also be slow to adjust to consumer trends in consumption of different goods and services, therefore being unrepresentative of inflationary trends until such goods and services are phased out of measurement. Unrepresentative inflation indexes constitute an issue as inexact analysis may determine the appropriate level of a given policy, as well as the determined need of policy decision-making itself, which can lead to incorrect applications of policy and reduced consumer confidence in government economic policy (Claeys Jeanrenaud 2021). Such a disconnect between inflation reality and economic policy can lead to the visible effects of increases in individual inflation uncertainty and in inflation disagreement amongst consumers as well as potentially invisible or yet to be seen implications brought forward by this misalignment of government policy and actual policy need (Olivier et al, 2021). The debate on the appropriate methods to measure inflation, and the exactness of economic policymakers in identifying and addressing it, loom large amongst the body of research exploring the pandemic's inflation effects which is rife with both subject-area and methodological "gaps".

Thus, there exist two substantial gaps in the research exploring the effects of inflation during exogenous events, most specifically in the case of a pandemic. The first gap is such that the present body of research exploring how exogenous shocks have affected inflation in the past have explored events as varied as natural disasters to political turmoil. However, the current body of research into exogenous shock inflation analysis has not ventured into analyzing the

effect pathogens have had on price increases. This is partly due to the lack of such widespread and global pathogens since the Spanish Flu beginning in 1918.

Turning to the second major gap, this is to say nothing about the different methodologies used in the fledgling body of research into inflation during the Covid-19 pandemic. Baskets of goods, the main measurement figures for the Consumer Price Indexes of the United States, the European Union, and Canada, remain the same as during non-pandemic periods. Therefore, serious value can be added to the burgeoning area of research measuring the inflation experienced during the Covid-19 pandemic relative to past periods on both subject-area, and methodological fronts.

During non-pandemic periods of inflation, the most common method of analyzing price increases was the use of the Consumer Price Index basket of goods. Other entries in the literature exploring inflationary growth have typically examined a set of twelve to twenty goods' prices in their research. This paper seeks to incorporate the real transactional dynamics occurring during the pandemic period, many of which entail the consumption of products used in the United States Consumer Price Index, but also, to reflect various products that more accurately reflect the goods and services most commonly used by consumers during both non-pandemic and pandemic periods. In essence, offering a more inclusive view into the way consumers are being affected in either time period and then comparing the magnitude of the affect consumers are experiencing in one time period to another.

As a result, this paper seeks to provide two major contributions to the body of research exploring the inflation effects of the Covid-19 pandemic. First, to measure the impact of the Covid-19 pandemic on consumer price inflation. Second, to modify the means of price inflation

analysis, via alternative methods including real transactional data, to measure the impact of the pandemic on inflation to provide a more complete picture into this relationship.

Research into the specific nature of inflation during the Covid-19 pandemic period and that of more normal “regular” periods of inflation is growing but still in its early stages. This paper seeks to add to the developing niche of the effects of the pandemic on inflation and to determine if there was a noticeable difference in the magnitude or nature of the inflation experienced during this period, compared to the past.

However, the paper seeks to also address another “gap” in the existing literature so to speak. That is, the methods used to measure the inflationary periods. Various alternative inflationary measures such as the basket of goods measure known as the Personal Consumption Expenditures Index and a NACIS sorted record of real transactional increases in price offer a more inclusive wider set of products and services to be measured than that of the Consumer Price Index. Such alternative measures in conjunction with the Consumer Price Index are employed to offer a wider, more applicable examination into the effect inflation has had on prices of goods and services based on actual usage and pertinence to consumer behavior rather than the existing system of a benchmark set of goods which may be changed incrementally every four years. This approach offers greater fidelity into the actual effect experienced by the consumer during the Covid-19 pandemic relative to times of normalcy.

Our research question, the driver for this paper is thus; has there been a substantive, statistically significant difference in the rates of growth of wider more extensive measures of product and goods price growth prior to and during the Covid-19 pandemic period compared to that of the normal CPI market basket of goods during periods of both normal inflation and the Covid-19 pandemic period?

ii. Addressing the Question: Data Sources and Methods Used

The three primary data sources used for comparison purposes is that of the US Consumer Price Index, the Personal Consumption Expenditure Index, and NAICS delineated real transactional data. The first such source, the Consumer Price Index, serves as the primary means of inflationary analysis of goods and services by the market as a whole and is the key barometer into the price inflation occurring within the US. The second source, the Personal Consumption Expenditures Index, is the key metric the U.S. Federal Reserve, the monetary policy makers for the U.S. domestically, and by extension, the global markets, utilize in measuring inflation. The final source, the NAICS real transactional data, reflects many products and services entries included in the Consumer Price Index but is larger in its entries including products and services most commonly used by consumers in 2020, 2021, and the first three months of 2022. As such, the NAICS real transactional data source has the expectation of offering a more powerful indicative analysis of goods and products most used by consumers via the actual price increases for various goods and services during the period.

These three data sets are sorted by products and services, some similar some different, but compared based on their product or service niches. For instance, the Consumer Price Index contains a line item such as “food away from home” while the additive index that reflects real transactional data may reflect this category via “food services and drinking places” establishing that though there are different products being measured, as well as a larger constituency of such goods and services being measured in the case of the real transactional data, both categories serve similar purposes as they examine similar areas of the economy and can thus justly be compared. The correlative relationship between prices will be measured to determine the exactness of the Consumer Price Index, our control, to that of the Personal Consumption

Expenditures Index and the NAICS sorted real transactional data set. Moreover, the change in price of similar goods and services subgroups will be measured between the three data sets to identify the extent to which the control data accurately returns similar inflationary results to that of the PCE and NAICS real transactional data. Used CPI data is sourced from the Bureau of Labor Statistics while used PCE data is collected from the Bureau of Economic Analysis Finally, separate credit and debit card transactional data is drawn from the North American Industry Classification System. Collection of data occurred on each respective site with the years constrained to the non-pandemic period of January 2002 – February 2020 and the pandemic period of March 2020 – January 2021.

The method of analysis used to find the above target measurements was linear regression via Ordinary Least Squares analysis through the Anaconda package of the Python coding language. This form of analysis seeks to measure the unknown parameters of a linear function for a set of selected explanatory variables. “Least squares” refers to the process of minimizing the sum of the squared values of the differences between the observed dependent variables, or the values of the variables being observed in our data sets, and those predicted by the independent variable within the linear function. In other words, Ordinary Least Squares seeks to fit a regression line of best fit between all included data points situated in the exact spot where the regression line’s slope is closest to the sum of the squared values of the distance between every point. Such a model works best when operating under the Gauss-Markov theorem when errors are homoscedastic and serially uncorrelated. In addition to these conditions, errors being normally distributed is also assumed so that Ordinary Least Squares is the maximum likelihood estimator. These conditions are assumed for the CPI, PCE and NAICS real transactional data sets.

Broadly speaking, this paper attempts to measure the effectiveness of the Consumer Price Index in reflecting inflationary conditions experienced by consumers and then comparing these real inflationary effects to prior periods of normal inflation to ascertain if Covid-19 induced price inflation of goods and services is substantively different in magnitude or nature to past periods.

iii. Analysis Drawbacks: Issue Areas with our Approach and the Existing Body of Literature

It is necessary to address various issue areas and points of contention with the methods used, data sets, the sourcing and manipulation of said data, and the surrounding literature on the topic.

Concerning our first data set, the additive basket of goods for the Covid-19 period was very short, (a little more than two years) as such, it may not offer a large enough data set to accurately compare against a longer, normal period of inflation. Selection of what can be defined as a “normal” period of price inflation can be difficult and is subject to the potential for bias in data sourcing as well. The creation of a new basket of goods that seeks to be additive and more representative of the Consumer Price Index also presents an opportunity for selection risk and the inclusion of products and services deemed to be more pertinent to the inflation experience of consumers than others, whilst not in fact being more representative. Furthermore, the additive basket of goods may not be applicable to all population groups and demographics so there is a potentiality that various racial or socioeconomic biases may be reflected in product and service selection in our approach.

Like data selection and cleaning, price data also presents potential cases for bias as prices may differ regionally, being higher or lower than others based on the cost of living of a certain

geographic location. Averages of prices of goods amongst many regions will help to compensate for this issue but will not reflect any one specific consumer experience, but rather, a blend of consumer's experiences across the United States potentially causing a reduction in the representability of the model's findings to any one geographic region. Selected consumer prices and overall habits are only being considered for the United States and are not representative, or indicative, of international inflationary trends or product and service usages which are outside the scope of the paper. Sampling error, or the risk that the sample chosen might not accurately represent the entire population and non-sampling error, such as errors associated with price-data collection and errors associated with operational implementation may also be present, but steps have been taken for the mitigation of these and other potential sources of bias which will be explored further in the methods section.

As for our control data set, the US Consumer Price Index, there are various issue points with using such a tool. First, most CPI index series use 1982-84 as the basis for comparison. The U.S. Bureau of Labor Statistics (BLS) sets the index level during this 1982-84 period at 100. This long-ago time period reflects a significantly different economy where a different set of goods and services were used which may limit the applicability of the control effect of this tool relative to the Covid-19 pandemic economic period. Like the additive basket of goods, sampling error, or the risk of the right sample not being chosen may occur. Or in other words, the sample chosen might not accurately represent the entire population. Again, like the additive basket of goods, non-sampling error could be present such as errors associated with price-data collection and errors associated with operational implementation. Errors pertaining to data sourcing and cleaning as well as time period selection also carry over to the control data.

The CPI also does not include energy costs even though they are a major expenditure for most households. This is one of the issues the NAICS delineated data will seek to alleviate. The Consumer Price Index may also not be applicable to all population groups. For example, CPI-U (Urban) better represents the U.S. urban population, but it does not reflect the status of the populace in more rural areas. The CPI also does not offer official estimates for subgroups of a population. Another potential drawback with CPI data utilization is that the index is a conditional cost-of-living measure that does not measure every aspect that affects living standards. Moreover, CPI data cannot be divided regionally to compare two areas within the United States as a higher index value for one area relative to another does not always mean that prices are higher in that former area. Lastly, like the additive NAICS price measurement data, social and environmental factors are beyond the definitional scope of the index.

The PCE index exhibits many of the same issues to that of the CPI index listed above including: the limited applicability of the measurement effect of PCE due to its usage of a set basket of goods, sampling error, non-sampling error, errors pertaining to data sourcing and cleaning as well as time period selection, omission of energy costs, limited application to population and demographic differences, and geographic comparison difficulties.

iv. Findings: Their Significance and Potential Policy Implications

The findings of the below analysis are pertinent for both researchers seeking to better understand inflationary periods driven by exogenous affects, most namely a pandemic, but also policy makers, namely politicians and the Federal Reserve for how to best measure and understand price inflation situationally and then in what capacity and magnitude to act.

Our findings indicate that the disparity in the nature of inflation during the non-pandemic period relative to the pandemic period is pronounced, while food and energy were one of many key affecters of inflation in the non-pandemic period, they now nearly solely constitute the sum of key indicators for inflation during the pandemic period. Hence, we find that the significance of omission of such pivotal energy and food data may have consequently been underestimated.

Methods of alleviating the issues found above include the inclusion of energy and food inflationary effects in PCE and CPI data to ensure actual price increases are measured commensurately with CPI and PCE. In addition, added research into the nature of pandemic inflationary affects with further back testing into non-pandemic periods is needed to ensure the results arrived at in this paper remain robust.

v. Structure: The Outline of the Paper

The paper is organized as follows. Section II describes the data sources and reports summary statistics as well as cross-sectional correlations between price inflation of the various data sets and empirical models during the non-pandemic and pandemic time periods. In Section III, we formalize our estimating equations, explain how the parameters we estimate should be interpreted using the Ordinary Least Squares (OLS) regression model, the assumptions included within the various empirical models, and the limitations and potential sources of bias within the empirical models. Section IV presents results on price inflation and the correlative relationship between each of the three methods of inflation measurement during the non-pandemic and pandemic periods using the research designs described above. In Section V, we analyze several policy proposals and considerations based on our results findings. Finally, Section VI presents a concise conclusion on the extent and nature of inflation in the non-pandemic and pandemic periods for all three methods of measurement as well as a future outlook on the extent and nature

of inflation in the pandemic period going forward. Section VII lists our various data and literature sources.

II. Data

Data was sourced via The Bureau of Labor Statistics (BLS), The Bureau of Economic Analysis (BEA) and the North American Industry Classification System (NAICS) entailing both non-pandemic period (January 2002 – February 2020) and pandemic period (March 2020 – January 2022) data. Indexes sourced from the BLS, BEA, and NAICS include Consumer Price Index (CPI) data, Personal Consumption Expenditures (PCE) data, and real debit card NAICS transactional growth rate data respectively. Relevant Y's of such data include the aggregate CPI, PCE and NAICS real transactional growth rates for both periods. Relevant X's for utilized data include the dozens of categorized sections of CPI, PCE, and NAICS real transaction growth rate data specific to certain types of goods, services and, in the case of the NAICS data, store types and functions. Arranging and cleaning the data in such a way offers both a high-level snapshot of inflationary pressures in both non-pandemic and pandemic time periods, as well as more focused, categorized examinations of specific areas within the wider economy experiencing inflationary effects.

Sample restrictions imposed were the predetermined date periods sets for the non-pandemic (January 2002 – February 2020) and pandemic (March 2020 – January 2022) time periods. For the non-pandemic time period the selected two-hundred-eighteen-month period was chosen to reflect an as-long-as-possible period with as minimum exogenous economic affects experienced, with the exception of the inclusion of the Great Recession period, this two decade period is reflective of relatively varied but non-extreme economic conditions to provide a strong and relevant back-testing period. The end date for the time period, (February 2020) reflects the

last month prior to the official declaration of the presence of a pandemic in the United States and the enactment of stay-at-home measures, government stimulus, and economically favorable monetary policy. As for the pandemic time period, to best reflect the time period for which a pandemic was officially declared, the start month (March 2020) was selected as well as a firm stopping point so newer data would not be included irregularly when it became available during the research process.

Conditions for dropped data were based on a variety of factors including; such data existing outside the confines of the aforementioned time periods to facilitate easy comparison between all data sets when made uniform via the dropping of data that was null or incomplete, the adjustment of various categories of inflation via merging component categories or entirely similar categories within each data set to make each category more comparable and uniform between the data sets for regression purposes. After such restrictions were made, the final sample size of the non-pandemic data was twenty-one columns and two-hundred and nineteen rows (for each month during the period and section headers) of data for the non-pandemic CPI data sets. The pandemic PCE data set was ten columns and the same two-hundred and nineteen figure for the number of rows of data. As for the pandemic data sets, the CPI, PCE, and NAICS data sets contained twenty-one rows (for each month during the period and section headers) while total columns were twenty-one, ten, and twenty respectively.

Summary statistics entailing means, standard deviations, and correlations for the above data sets were collected to provide a high-level idea of what each data set's contents look like. Means for non-pandemic CPI, non-pandemic PCE, pandemic CPI, pandemic PCE, pandemic NAICS are as follows; 0.020748, 0.153670, 0.032130, 0.304545, and 0.859406 respectively. Turning to standard deviation statistics, standard deviations for non-pandemic CPI,

non-pandemic PCE, pandemic CPI, pandemic PCE, and pandemic NAICS are as follows; 0.012545, 0.209036, 0.024726, 0.276848, and 11.278619 respectively. Concerning correlative relationships between the data sets, the NAICS data set was split into two different regressions, one measuring pandemic NAICS against each individual category of CPI and PCE, and one measuring pandemic NAICS against Aggregate PCE and Aggregate CPI in their entirety explaining the addition of a sixth correlation, one for each regression, compared to the five previously listed means and standard deviations, one for each data set. Correlations for non-pandemic CPI, non-pandemic PCE, pandemic CPI, pandemic PCE, pandemic NAICS vs CPI and PCE categories, and Pandemic NAICS vs Aggregate CPI and Aggregate PCE are as follows; 0.332, 0.251, 0.627, 0.774, 0.848, and 0.445 respectively.

Adjusted R-squared, a modified version of R-squared that is adjusted for the number of predictors within each regression model for non-pandemic CPI, non-pandemic PCE, pandemic CPI, pandemic PCE, pandemic NAICS vs CPI and PCE categories, and pandemic NAICS vs Aggregate CPI and Aggregate PCE are as follows; 0.313, 0.204, 0.477, 0.407, -0.599, and 0.386 respectively. Adjusted R-squared increases when a new term addition improves the model more than what would be expected by chance while the measure decreases when a predictor improves the model by less than what would reasonably be expected. The adjusted R-squared values of the pandemic CPI and PCE regressions seem to exhibit a greater departure from their respective R-squared values than that of their CPI and PCE non-pandemic regressions indicating that the regressions conducted for the non-pandemic period are more improved based on their variables than that of the pandemic regressions. Notably, the pandemic NAICS vs CPI and PCE categories regression has a negative adjusted R-squared therefore indicating that the fit created by the addition of the many variables within the model is actually worse than just fitting a horizontal

line. This finding is consistent as the aforementioned NAICS vs CPI and PCE categories regression model contains the largest number of (X_N) values of any of the regressions, meaning each subsequent additional variable is reducing the strength of the model rather than improving it marginally.

III. Methods

i. Empirical Model Formulation and Description

Our research question pertains to examining whether there has been a substantive, statistically significant difference in the rates of growth of a wider more extensive luxury and cost-efficient measure of product and goods price growth during the Covid-19 pandemic period compared to that of the normal CPI market basket of goods during periods of normal inflation. The regression models constructed to examine this question are six-fold and are designed first and foremost to limit the potential of bias within the results of the model, while garnering as reasonably accurate, predictive results as possible.

Each of the six empirical models used follow the same basic formulation that consists of a multivariable regression such that:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots \beta_{11} X_{11} + \mu$$

Where the outcome/ dependent variable is denoted as (Y), exposure independent variables for which (Y) is regressed against are denoted as (X_N) The value of (Y) when all of the independent variables (X_1 through X_{11}) are equal to zero is denoted as (β_0), (β_N) is the estimated regression coefficient that quantifies the association between the potential confounder (X_N) and the outcome (Y), lastly, (μ) relates to the error term of the model containing all potential casual

factors and data points not expressly included as dependent (Y) or independent variables (X_N) within the model.

There are two non-pandemic period regression models and four pandemic period models. The first non-pandemic model pertains to non-pandemic Aggregate PCE vs the various components of CPI. The first regression model is as follows:

1. Empirical Model One: Non-Pandemic Aggregate PCE

$$\begin{aligned} \text{Non-pandemic Aggregate PCE} = & \beta 0 + \beta 1 \text{ food} + \beta 2 \text{ food at home} + \beta 3 \text{ food away from} \\ & \text{home} + \beta 4 \text{ energy} + \beta 5 \text{ gasoline all types} + \beta 6 \text{ electricity} + \beta 7 \text{ natural gas piped} + \beta 8 \\ & \text{apparel} + \beta 9 \text{ new vehicles} + \beta 10 \text{ medical care commodities} + \beta 11 \text{ shelter} + \beta 12 \text{ medical} \\ & \text{care services} + \beta 13 \text{ education and communication} + \mu \end{aligned}$$

The above regression is an ordinary least squares multivariable regression that compares non-pandemic Aggregate PCE to that of non-pandemic CPI components to determine its correlative relationship with that of the non-pandemic Aggregate CPI. Each of the (X_N) values pertains to a specific constituent variable used in the calculation of the aggregate PCE number reported by the Bureau of Economic Analysis, less any duplicate or unconfigured categories (as stated previously any dropped categories were combined with other similar categories for ease of comparison). Each category is labeled as such to indicate the metrics and parameters for which it is calculated, for instance, the “food at home” variable represents price changes during the non-pandemic data period in the price of foods traditionally purchased for consumption at home, such as those purchased at a grocery store, whilst the “food away from home” variable relates to foods purchased during the non-pandemic data period and consumed away from home including a variety of restaurants in the category. Such an arrangement allows us to compare aggregate

PCE to that of each specific constituent variable of CPI to determine commonalities as well as extreme differences amongst both indicators on a categorical basis during the non-pandemic data period. The significance of (β_0) is tantamount as it is the key parameter distinguishing the intercept of the regression line at the point in which the empirical model's various X variables are equal to zero. As such in this specific model, (β_0) represents the value returned of the regressed line when the sum of the non-pandemic variables; "food", "food at home", "food away from home", "energy", "gasoline all types", "electricity", "natural gas piped", "apparel", "new vehicles", "medical care commodities", "shelter", "medical care services", and "education and communication" equals zero.

The second non-pandemic model pertains to non-pandemic Aggregate CPI vs the various components of PCE. The regression model is as follows:

2. Empirical Model Two: Non-Pandemic Aggregate CPI

$$\text{Non-pandemic Aggregate CPI} = \beta_0 + \beta_1 \text{ regular goods} + \beta_2 \text{ durable goods} + \beta_3 \text{ nondurable goods} + \beta_4 \text{ services} + \beta_5 \text{ energy goods and services} + \mu$$

The above regression is an ordinary least squares multivariable regression that compares non-pandemic Aggregate CPI to that of non-pandemic PCE components to determine its correlative relationship with that of the non-pandemic Aggregate PCE. Each of the (X_N) values pertains to a specific constituent variable used in the calculation of the aggregate CPI number reported by the Bureau of Labor Statistics, less any duplicate or unconfigured categories. Each category is labeled as such to indicate the metrics and parameters for which it is calculated, for instance, the "services" variable represents price changes in a variety of services purchased by consumers during the non-pandemic data period with the exception of services pertaining to

energy, whilst the “energy goods and services” variable relates to a variety of energy goods and energy-related services purchased during the non-pandemic data period in the category. Such an arrangement allows us to compare aggregate CPI to that of each specific constituent variable of PCE to determine commonalities as well as extreme differences amongst both indicators on a categorical basis during the non-pandemic data period. The significance of (β_0) is key as it once again the parameter distinguishing the intercept of the regression. In this specific model, (β_0) represents the value returned of the regressed line when the sum of the non-pandemic variables; “regular goods”, “durable goods”, “nondurable goods”, “services”, and “energy goods and services 2” equals zero.

The first pandemic model pertains to pandemic Aggregate PCE vs the various components of CPI. The regression model is as follows:

3. Empirical Model Three: Pandemic Aggregate PCE

$$\begin{aligned} \text{Pandemic Aggregate PCE} = & \beta_0 + \beta_1 \text{ Food} + \beta_2 \text{ food at home} + \beta_3 \text{ food away from home} + \\ & \beta_4 \text{ energy} + \beta_5 \text{ gasoline all types} + \beta_6 \text{ electricity} + \beta_7 \text{ natural gas piped} + \beta_8 \text{ apparel} + \beta_9 \\ & \text{new vehicles} + \beta_{10} \text{ medical care commodities} + \beta_{11} \text{ medical care services} + \\ & \beta_{12} \text{ education and communication} + \mu \end{aligned}$$

The above regression is an ordinary least squares multivariable regression that compares pandemic Aggregate PCE to that of pandemic CPI components to determine its correlative relationship with that of the pandemic Aggregate CPI. Each of the (X_N) values pertains to a specific constituent variable used in the calculation of the aggregate PCE number reported by the Bureau of Economic Analysis, less any duplicate or unconfigured categories. Each category is labeled as such to indicate the metrics and parameters for which it is calculated. Such an

arrangement allows us to compare aggregate PCE to that of each specific constituent variable of CPI to determine commonalities as well as extreme differences amongst both indicators on a categorical basis during the pandemic data period. In empirical model three, (β_0) represents the value returned of the regressed line when the sum of the pandemic variables; “food”, “food at home”, “food away from home”, “energy”, “gasoline all types”, “electricity”, “natural gas piped”, “apparel”, “new vehicles”, “medical care commodities”, “shelter”, “medical care services”, and “education and communication” is equivalent to zero.

The second pandemic model pertains to pandemic Aggregate CPI vs the various components of PCE. The regression model is as follows:

4. Empirical Model Four: Pandemic Aggregate CPI

$$\text{Pandemic Aggregate CPI} = \beta_0 + \beta_1 \text{ regular goods} + \beta_2 \text{ durable goods} + \beta_3 \text{ nondurable goods} + \beta_4 \text{ services} + \beta_5 \text{ energy goods and services} + \mu$$

The above regression is an ordinary least squares multivariable regression that compares pandemic Aggregate CPI to that of PCE components to determine its correlative relationship with that of the pandemic Aggregate CPI. Each of the (X_N) values pertains to a specific constituent variable used in the calculation of the aggregate PCE number reported by the Bureau of Economic Analysis, less any duplicate or unconfigured categories. Each category is labeled as such to indicate the metrics and parameters for which it is calculated. Such an arrangement allows us to compare aggregate PCE to that of each specific constituent variable of CPI to determine commonalities as well as extreme differences amongst both indicators on a categorical basis during the pandemic data period. For empirical model four, (β_0) represents the value

returned of the regressed line when the sum of the pandemic variables; “regular goods”, “durable goods”, “nondurable goods”, “services”, and “energy goods and services 2” equates to zero.

The third pandemic model pertains to pandemic Aggregate NAICS real transaction data vs Aggregate Total PCE and Aggregate Total CPI. The regression model is as follows:

5. Empirical Model Five: Pandemic Aggregate NAICS vs. Pandemic Aggregate PCE and Pandemic Aggregate CPI

$$\text{Pandemic Aggregate NAICS} = \beta_0 + \beta_1 \text{Aggregate total PCE} + \beta_2 \text{Aggregate total CPI} + \mu$$

The above regression is an ordinary least squares multivariable regression that compares pandemic Aggregate NAICS to that of pandemic Aggregate PCE and pandemic Aggregate CPI to determine the correlative relationship of the pandemic Aggregate NAICS figure to that of the aforementioned pandemic Aggregate PCE and pandemic Aggregate CPI figures. Each of the (X_N) values pertains to a specific constituent variable used in the calculation of the aggregate PCE number reported by the Bureau of Economic Analysis, less any duplicate or unconfigured categories. Each category is labeled as such to indicate the metrics and parameters for which it is calculated, for instance, the “Aggregate total PCE” variable represents price changes in all types of goods and services purchased by consumers during the pandemic data period that are measured by PCE measuring the sum of all category-specific inputs into PCE as a whole, whilst the “Aggregate total PCE” variable represents price changes in all types of goods and services purchased by consumers during the pandemic data period that are measured by CPI measuring the sum of all category-specific inputs within CPI in its entirety. Such an arrangement allows us to compare aggregate NAICS real transactional data during the pandemic period to that of aggregate pandemic PCE and aggregate pandemic CPI to determine commonalities as well as

extreme differences amongst both indicators during the pandemic data period. In our fifth empirical model, (β_0) represents the value returned of the regressed line when the sum of the pandemic variables; “Aggregate total PCE” and “Aggregate total CPI” equals zero.

The fourth and final pandemic model pertains to pandemic Aggregate NAICS real transaction data vs complete pandemic PCE components and complete pandemic CPI components. The regression model is as follows:

6. Empirical Model Six: Pandemic Aggregate NAICS vs. All Pandemic PCE Constituent Variables and All Pandemic CPI Constituent Variables

$$\begin{aligned} \text{Pandemic Aggregate NAICS} = & \beta_0 + \beta_1 \text{ regular goods} + \beta_2 \text{ durable goods} + \beta_3 \text{ nondurable} \\ & \text{goods} + \beta_4 \text{ services} + \beta_5 \text{ energy goods and services} + \beta_6 \text{ Food} + \beta_7 \text{ food at home} + \beta_8 \\ & \text{food away from home} + \beta_9 \text{ energy} + \beta_{10} \text{ gasoline all types} + \beta_{11} \text{ electricity} + \beta_{12} \text{ natural} \\ & \text{gas piped} + \beta_{13} \text{ apparel} + \beta_{14} \text{ new vehicles} + \beta_{15} \text{ medical care commodities} + \beta_{16} \text{ shelter} \\ & + \beta_{17} \text{ medical care services} + \beta_{18} \text{ education and communication} + \mu \end{aligned}$$

The above regression is an ordinary least squares multivariable regression that compares pandemic Aggregate NAICS to that of pandemic PCE components and pandemic CPI components to determine its correlative relationship with that of the entire field of components of CPI and PCE pandemic data. Each of the (X_N) values pertains to a specific constituent variable used in the calculation of the aggregate PCE number reported by the Bureau of Economic Analysis or a specific constituent variable used in the calculation of the aggregate CPI number reported by the Bureau of Labor Statistics, less any duplicate or unconfigured categories. Each category is labeled as such to indicate the metrics and parameters for which it is calculated. Such

an arrangement allows us to compare aggregate NAICS real transactional data during the pandemic period to that of each specific pandemic constituent variable of PCE and each specific constituent pandemic variable of CPI to determine commonalities as well as extreme differences amongst both indicators during the pandemic data period. For the final model, empirical model six, (β_0) represents the value returned of the regressed line when the sum of the pandemic variables; “regular goods”, “durable goods”, “nondurable goods”, “services”, “energy goods and services 2”, “food”, “food at home”, “food away from home”, “energy”, “gasoline all types”, “electricity”, “natural gas piped”, “apparel”, “new vehicles”, “medical care commodities”, “shelter”, “medical care services”, and “education and communication” is zero.

ii. Empirical Model Assumptions and Estimator Properties

In my empirical models I am assuming the various X variables I am examining are not perfectly correlated. The complete list of X variables measured in one or more models are as follows: β_1 regular goods + β_2 durable goods + β_3 nondurable goods + β_4 services + β_5 energy goods and services 2 + β_6 Food + β_7 food at home + β_8 food away from home + β_9 energy + β_{10} gasoline all types + β_{11} electricity + β_{12} natural gas piped + β_{13} apparel + β_{14} new vehicles + β_{15} medical care commodities + β_{16} shelter + β_{17} medical care services + β_{18} education and communication. However, though efforts are taken to protect against correlation between the X variables, some level of relationship between the X variables in the model is virtually guaranteed. Further, I am also assuming that the data was sampled via random sampling from the U.S Bureau of Labor Statistics and The Bureau of Economic Analysis and therefore, is very unlikely to contain non-random sampling bias. Additionally, I am also assuming my model is fully in line with the essential MLR.1-MLR.4 rules for multivariable OLS estimators to be unbiased. Namely, that the Whole Common Rank assumption must hold true meaning that each

of the model's X variables must vary independently of each other and that the variance in the X variables is non-zero. Additionally, the model must be defined such that $Y = B_0 + B_1X_1 + B_2X_2 \dots + \text{Error Term}$ for the multiple regression case as stated previously when discussing the empirical models used. Furthermore, that a random sample from the population must be utilized, this was done via sampling the entire U.S. economy which includes a vast variety of different variables that can affect purchasing frequency and price inflation such as demographics, income-levels, education-level etc. Lastly, and most importantly, I assume that our model and the data sourced for it follow the Exogeneity condition that the error term must be uncorrelated with each of the X variables in the multiple regression model.

iii. **Empirical Model Limitations and Potential Sources of Bias**

The potential sources of bias in my model, though I believe they have been adequately controlled for and mitigated, is omitted variable bias, measurement error bias, and non-random sampling bias. Firstly, I may have incurred omitted variable bias as there may exist a non-irrelevant variable other than the various aggregate and category specific PCE, CPI and real transactional data NAICS values included in the various non-pandemic and pandemic data values that I did not include in my various empirical models and thus placed in the error term that is biasing my regression results. Such a placing of a non-irrelevant variable in the error term of any one of my empirical models would be an erroneous mistake dismissing my subsequent analysis due to the exclusion of a necessary variable thereby biasing my results. Since omitted variable bias can stem from any non-irrelevant variable that is excluded and placed in the error term when the omitted variable is correlated with one or more Xs in the model, unknowingly placing one or more non-irrelevant variables in the error term by not including it in the empirical model constitutes a risk for bias (endogeneity) to occur. The below omitted variable bias table helps to

predict which direction the model may be biased on the covariance of X and X_{omitted} being greater than or less than zero and whether β_{omitted} is greater than or less than zero:

1. Omitted Variable Bias Directional Table

	Positive $\text{Cov}(x, x_{\text{omit}}) > 0$	Negative $\text{Cov}(x, x_{\text{omit}}) < 0$
Positive $\beta_{\text{omitted}} > 0$	Upward Bias $E(B_1^{\text{OLS}}) > B_1$	Downward Bias $E(B_1^{\text{OLS}}) < B_1$
Negative $\beta_{\text{omitted}} < 0$	Downward Bias $E(B_1^{\text{OLS}}) < B_1$	Upward Bias $E(B_1^{\text{OLS}}) > B_1$

Of all of the necessary assumptions that guarantee unbiasedness the exogeneity condition is the most likely empirical model assumption to fail and can fail for a number of reasons including the presence of omitted variable bias in a model.

Turning to measurement error, my model may also have incurred some level of bias due to measurement error which stems from the difference between the observed value and the actual value of the dependent variable. Hence, I may have unknowingly incurred measurement error in my dependent variables which can cause bias if my dependent variables are related to one or more of my explanatory variables. Hence, endogeneity could arise if the dependent variable of any given empirical model, in the case of model one, (non-pandemic aggregate PCE) is overly related to one or more explanatory variables, again in the case of empirical model one (“food”, “food at home”, “food away from home”, “energy”, “gasoline all types”, “electricity”, “natural

gas piped”, “apparel”, “new vehicles”, “medical care commodities”, “shelter”, “medical care services”, and “education and communication”). Lastly, this bias, if present within the model, would likely have a direction towards zero as measurement error in X always has a direction of bias towards zero due to the concept of attenuation bias. However, it is important to note if the measurement error is independent of the explanatory variable, bias is not present. Of all of the necessary assumptions that guarantee unbiasedness, the exogeneity condition is the most likely empirical model assumption to fail and can fail for a number of reasons such as omitted variable bias, as previously mentioned, as well as the presence of measurement error in any one of the empirical models. The potential of the exogeneity condition failing is unlikely due to the broad categorical X variables within each data set which reduce the likelihood of instances of omitted variables coupled with the inclusion of a high number of price inflation categories that are of a distinctly different nature and focus compared to that of the aggregate inflation measures in question. As such, due to the variety and completely different areas of focus of each category, strong correlation to that of any such aggregate inflationary measure is low.

Lastly, it is possible I may have unknowingly biased any one of the six models used via non-random sampling as I deliberately utilized data from two predetermined time periods, January 2002 – February 2020 and March 2020 – January 2022 and excluded all other periods outside these two from my data set. This deliberate selection of time period and non-random removal of all years outside the confines of my pre-selected time period could constitute a non-random selection of data and may reflect my own bias. Furthermore, the categorization of these time periods, the first one (January 2002 – February 2020) being labeled as a non-pandemic period of time and the other (March 2020 – January 2022) being designated as a pandemic period of time, may also represent my biases and therefore the aforementioned

selected periods may not be indicative of the labels they have been designated, indicating the results derived from either period are inaccurate. Non-random sampling bias is brought on by the usage of non-random samples as opposed to stratified sampling or random sampling.

Endogeneity occurs when sampling selection is based on the dependent variable, typically when sample selection is based on whether the dependent variable is above or below a given value. For non-random sampling, the direction of bias depends on the relationship between your sample and the population, for a given set of data if you are oversampling a group of people for instance, there would likely be a large effect of X on Y prompting upward bias to occur and vice versa. It is thus vital samples are randomly obtained as pre-selection of a sample can reflect the biases of the person selectively picking data to analyze. Though the likelihood for non-random sampling bias being present in the model is low due to the robust data selection standards of the Bureau of Labor Statistics and the Bureau of Economic Analysis, my own purposeful parsing out of periods outside my measured timeline, may constitute a risk for non-random sampling bias within the various empirical regression models.

Lastly, two non-bias specific limitations of my data are multicollinearity within both PCE regressions and the second NAICS real transactional regression and the data interval I selected. Concerning multicollinearity, or a high degree of correlation between the X's of the model, three of my regressions, ("Empirical Model One: Non-pandemic PCE", Empirical Model Three: Pandemic PCE", and "Empirical Model Six: Pandemic Aggregate NAICS vs. All Pandemic PCE Constituent Variables and All Pandemic CPI Constituent Variables", experienced multicollinearity. While this does not invalidate the findings of the regressions, it does increase the size of the standard errors and variance of the models meaning the conclusions drawn from them, through unbiased, may not be as exact as possible. Secondly, since my interval is monthly,

(as the common date interval between the various Bureau of Labor Statistics and Bureau of Economic Analysis data sets were all offered in monthly intervals), the number of data points within my model are relatively small leaving a higher chance for the data to be less representative of the overall population (or all other years of price inflation weekly data) thereby reducing the efficiency of the models. This issue is further exasperated within the pandemic data values, two of which experienced multicollinearity issues, as the time frame for these models is only twenty-one months compared to the non-pandemic period entailing a two-hundred and eighteen data month period. The effect if any these instances of large condition numbers, and therefore multicollinearity, had on the three empirical models, the potential sources and circumstances behind it occurring, and possible means of alleviating the issue will be addressed with respect to each model further during the results analysis section.

iv. Model Applicability: Addressing the Research Question

Our research question sought to examine if there existed a significant difference in the rates of growth of more extensive measures of product and goods price growth prior to and during the Covid-19 pandemic period compared to that of the normal CPI market basket of goods. Each specific empirical model is designed to address this question from different angles.

Empirical models one and two examine the relationship between CPI and an alternative inflationary measure, PCE, during the non-pandemic data set period. Model one accomplishes this by regressing aggregate PCE data values for each month from January 2002 – February 2020 to that of the constituent goods and services components of CPI. In doing so, the model measures the relationship of PCE findings to that of different aspects of CPI to identify specific variables of CPI that exhibit a statistically significant relationship with PCE thus denoting aspects of CPI that coincide with this alternative inflationary measurement as well as others that fail to mirror it.

As PCE is the main inflationary measurement that informs Federal Reserve monetary policy this regression examines the effectiveness of different categories of CPI during non-pandemic periods to that of the leading inflation measurement that guides interest rates and expansionary or contractionary economic policy during a non-pandemic period.

Empirical model two also effectively measures for and addresses our research question through regressing aggregate CPI data values for each month from January 2002 – February 2020 to that of the constituent goods and services components of PCE. In doing so, the model measures the relationship of CPI findings to that of different aspects of PCE to identify specific variables of PCE that exhibit a statistically significant relationship with CPI thus indicating the applicability of aggregate CPI in measuring varying categories of PCE. As stated previously, PCE is the main inflationary measurement that informs Federal Reserve monetary policy. This regression examines the effectiveness of CPI as an aggregate measurement of total inflation compared to the various categories of PCE during non-pandemic periods to establish the ability of CPI in tracking category specific inflationary trends influencing Federal Reserve policy during instances of normal inflation.

Empirical models three and four also examine the relationship between CPI and PCE but do so during the pandemic data set period. Empirical model three regresses aggregate PCE March 2020 – January 2022 to that of the same constituent good and service components of CPI. The model therefore measures the relationship of PCE findings to that of different aspects of CPI to identify specific variables of CPI that exhibit a statistically significant relationship with PCE, denoting aspects of CPI that coincide with this alternative inflationary measurement as well as others that fail to mirror it. This regression importantly measures this relationship during the pandemic period accomplishing two separate goals that address our research question. Firstly, the

model measures the effectiveness of various CPI goods and services categories in matching the results of aggregated PCE. Secondly, empirical model three allows us to compare the ability of CPI constituent variables in correlating with PCE during non-pandemic periods to that of the pandemic period measured in empirical model one therefore demonstrating the effect the pandemic had on the success of CPI in its inflationary measurement abilities.

Just as empirical model three closely resembles the structure of model one, model four is largely the same as model two with the one difference between the two being model four's measurement of pandemic time period data as opposed to the normal inflationary period. Via regressing aggregate CPI data values for each month from March 2020 – January 2021 to that of the constituent goods and services components of PCE, the model measures the relationship of CPI findings to that of different aspects of PCE to identify if any specific variables of PCE register a statistically significant relationship with CPI thus indicating the applicability of aggregate CPI in measuring varying categories of PCE. Due to PCE serving as a measure that guides Federal Reserve policy, this model examines the ability of CPI to measure various areas of inflation examined via PCE during the Covid-19 pandemic period.

The final two empirical models, five and six, represent a slight departure from the model structuring of the previous four models. Empirical models five and six utilize NAICS delineated real transactional data of goods and services as a second alternative measurement to compare CPI, as well as PCE exclusively during the pandemic period. Empirical model five regresses aggregate NAICS data values for each month from March 2020 – January 2021 to that of aggregate CPI and aggregate PCE values during that same period. The model measures the relationship of CPI and PCE aggregate approximations of price inflation to that of real price growth rates tracked via debit card purchases at various business that are sorted based on their

service or good type to coincide with the categories of CPI and PCE. The regression seeks to identify the relationship of CPI and PCE in their entirety to the real transactional price increases occurring during the pandemic period. As such, the model therefore examines the applicability of CPI and PCE to that of measuring actual price growth during the pandemic inflationary period.

The final model, empirical model six, compares aggregate NAICS data values for each month from March 2020 – January 2021 to that of the constituent component measurements of CPI and PCE during that same period. The model measures the relationship of CPI and PCE component inflationary measurements to real price inflation growth derived from the NAICS real transactional data thereby identifying the relationship of CPI and PCE component parts that measure specific goods or services to the real transactional price increases occurring during the pandemic period of those same categories. As such, this final model therefore examines the applicability of CPI and PCE categories in measuring the actual price growth of the goods and services they claim to represent during the pandemic inflationary period.

IV. Results

Consolidated regression results containing information pertaining to intercepts, R-squared values, adjusted R-squared values, counts of total data points within each regressed data set, and categorized variable values within each regression can be found below. Importantly, values that are denoted with a single asterisk are statistically significant at the ninety percent confidence interval level. Two asterisks denote a data point that is statistically significant at the ninety five percent confidence level meeting the common-practice threshold of statistical significance with a p-value less than or equal to five percent. Finally, results with three asterisks represent results that are statistically significant at the ninety eight percent confidence interval. P-values for each category variable within each regression are bolded immediately below the

respective value readings they pertain to. All below regression models and their findings contain standard errors that were heteroskedasticity robust. However, empirical models one, three, and six contained exceedingly large condition numbers potentially indicating the model encountered strong multicollinearity issues. Multicollinearity issues for each of the regressions containing such issues are accordingly addressed on a regression specific basis below. Results are separated into two distinct regression tables, “Summary Regression Table 1” and “Summary Regression Table 2”.

“Summary Regression Table 1” contains key findings from regression models one, two and three. Empirical Model One: Non-Pandemic Aggregate PCE is denoted as “PCE I (Reg1)”, Empirical Model Two: Non-Pandemic Aggregate CPI is referenced as CPI I (Reg2)”, and Empirical Model Three: Pandemic Aggregate PCE is listed as “PCE II (Reg3)”:

i. Summary Regression Table 1 (Empirical Models 1 – 3 Regression Results)

Regression	PCE I (Reg1)	CPI I (Reg2)	PCE II (Reg3)
Intercept	0.03 0.562	0.01*** 0.000	-0.91 0.768
R-squared (R2)	0.25	0.33	0.77
R-squared Adj.	0.20	0.31	0.41
N	218	218	22
apparel	-1.44* 0.084		-1.87 0.769
durablegoods		0.00 0.827	
educationandcommunication	-1.58 0.240		5.01 0.869
electricity	-0.51 0.502		-20.26* 0.064
energy	3.27*** 0.017		1.16 0.906
energygoodsandservices2		0.00 0.150	
food	-1.27 0.959		139.85 0.440

food1		0.02***	
		0.000	
foodathome	-1.27		-69.85
	0.929		0.511
Regression	PCE I (Reg1)	CPI I (Reg2)	PCE II (Reg3)
foodawayfromhome	2.57		-38.14
	0.833		0.626
gasolinealltypes	-1.19*		0.75
	0.088		0.878
medicalcarecommodities	-0.67		-6.44
	0.461		0.796
medicalcareservices	3.11*		2.55
	0.071		0.905
naturalgaspiped	-0.98***		4.01
	0.000		0.396
newvehicles	0.06		-8.16
	0.969		0.681
nondurablegoods		-0.01	
		0.607	
regulargoods		0.00	
		0.836	
services		0.03***	
		0.003	
shelter	-0.11		-0.31
	0.957		0.995

Of the three model regressions displayed in the table above there were a total of nine statistically significant variables amongst the three regressions. Concerning “PCE I (Reg1)”, the regression in which non-pandemic aggregate PCE was regressed against the various components of non-pandemic CPI, five total statistically significant variables were identified. The variables “apparel”, “energy”, “gasolinealltypes”, “medicalcareservices”, and “naturalgaspiped”, all were statistically significant to varying degrees with p-values of 0.084, 0.084, 0.017, 0.088, and 0.000 respectively. Empirical model one’s findings concerning the aforementioned significant variables were as follows; for every one unit increase in “apparel”, the intercept (non-pandemic PCE) declined by -1.44, for each one unit increase in “energy” PCE rose by 3.27, As for “gasolinealltypes”, for each one unit increase in the variable PCE declined by -1.19, with a one

unit increase in “medicalcareservices”, PCE increased by 3.11, lastly, for each one unit increase in “naturalgaspiped”, PCE fell by -0.98.

This model experienced multicollinearity issues returning a large condition number, likely due to the sheer number of X variables contained within the model. These values already constitute a reduced count of the pre-adjusted various components of CPI as similar aggregable CPI components were combined when applicable in an effort to reduce the potentiality for large standard errors. To attempt to fix this issue, the above regression could be subdivided into two different regressions halving the number of independent variables within the present model. Although, such an adjustment would reduce the ability to accurately measure the entire combined relationship between the elements of CPI to PCE. Alternatively, the CPI variables could be further consolidated together, although this method runs the risk of denaturing each variable to such a point that it would be difficult to understand the true relationship between each area of CPI and PCE.

Turning to “CPI I (Reg2)”, the regression in which non-pandemic aggregate CPI was regressed against the various components of non-pandemic PCE, three total statistically significant variables were identified, (including the coefficient value). The “coefficient” variable, “food1”, and “services” were all found to be statistically significant with p-values of 0.000, 0.000, and 0.003 respectively. Empirical model two’s findings pertaining to the aforementioned significant variables were as follows; for each one unit increase in the “intercept” variable itself the intercept (non-pandemic CPI) rose by 0.01, With each one unit increase in “food1” CPI rose by 0.02, and lastly, with each one unit increase in “services”, CPI rose by 0.03.

Moving on to the final regression model within “Summary Regression Table 1”, “PCE II (Reg3)”, the regression in which pandemic aggregate PCE was regressed against the various

components of pandemic CPI, there was only a single variable identified to be statistically significant. The variable “electricity” was found to be significant with a p-value of 0.064. Further, when the “electricity” variable was increased by one unit, the intercept (pandemic PCE) was found to have decreased by -20.26.

Unfortunately, empirical model three also ran into multicollinearity induced problems. Similar to empirical model one, the sheer number of X variables contained within the model likely played a large role in the enlargement of the standard errors within the regression. To attempt to fix this issue as previously mentioned, the regression could be subdivided into two different regressions reducing the number of independent variables within the present model. Yet, due to the already reduced and consolidated nature of the X variables within CPI already, such an alteration may reduce the predictive and comprehension value of the model. Alternatively, another potential contributor to the multicollinearity issue may be the shortened nature of the pandemic period compared to that of the prior non-pandemic period being measured in the above regression. Shortened time periods reduce the total number of data points being evaluated, creating a situation where there exists a higher potentiality for a correlative relationship between the reduced number of X variables during the shortened pandemic period. This is unfortunately a problem that cannot be artificially mitigated, with the only solution being the addition of more data points as they become available over time.

“Summary Regression Table 2” contains key findings from regression models four, five and six. Empirical Model Four: Pandemic Aggregate CPI is denoted as “CPI II (Reg4)”, Empirical Model Five: Pandemic Aggregate NAICS vs. Pandemic Aggregate PCE and Pandemic Aggregate CPI is referred to as “NAICSTrans. (Reg5)”, and Empirical Model Six: Pandemic

Aggregate NAICS vs. All Pandemic PCE Constituent Variables and All Pandemic CPI

Constituent Variables is labeled as “NAICSTrans. II (Reg6)”:

ii. Summary Regression Table 2 (Empirical Models 4 – 6 Regression Results)

Regression	CPI II (Reg4)	NAICSTrans. (Reg5)	NAICSTrans. II (Reg6)
Intercept	0.00 0.611	-0.41 0.817	28.40 0.778
R-squared (R2)	0.63	0.44	0.85
R-squared Adj.	0.48	0.39	-0.60
N	22	22	22
allitemsadjustedcpi		-124.01*** 0.015	
apparel			-96.12 0.705
durablegoods	-0.03 0.496		-2.97 0.921
educationandcommunication			-897.81 0.469
electricity			-297.22 0.676
energy			218.27 0.698
energygoodsandservices2	-0.01* 0.063		-1.62 0.721
food			7526.70 0.403
food1	0.00 0.977		-0.27 0.977
foodathome			-4250.52 0.382
foodawayfromhome			-2841.45 0.294
gasolinealltypes			-134.01 0.658
medicalcarecommodities			482.25 0.500
medicalcareservices			-732.77 0.593
naturalgaspiped			19.54 0.928
newvehicles			-114.91 0.920
Regression	CPI II (Reg4)	NAICSTrans. (Reg5)	NAICSTrans. II (Reg6)

nondurablegoods	-0.01 0.870		-5.01 0.934
pce		15.39*** 0.000	
regulargoods	0.08 0.406		18.18 0.829
services	0.08*** 0.017		24.85 0.342
shelter			-61.15 0.984

Concerning the next three regression models displayed in the table above, there were four instances of statistically significant variables within the three regression models. First addressing “CPI II (Reg4)”, the regression in which pandemic aggregate CPI was regressed against the various components of pandemic PCE, there were two statistically significant variables discovered. The variables “energygoodsandservices2” and “services” were identified as statistically significant with p-values of 0.063 and 0.017 respectively. Empirical model four’s significant variables were found to have the following relationship with the intercept variable; for each one unit increase in “energygoodsandservices2” the intercept (pandemic CPI) saw a -0.01 decrease while a one unit increase in the “services” variable led to a 0.08 increase for CPI.

Moving to “NAICSTrans. (Reg5)”, this regression compared the pandemic NAICS real transactional data to that of pandemic aggregate CPI and aggregate PCE during the pandemic period. Of the two variables regressed against the coefficient variable both were found to be statistically significant variables. The variables “allitemsadjustedcpi” and “pce” had p-values of 0.015 and 0.000 respectively. Empirical model five’s significant variables were determined to have a varied relationship with that of the NAICS real transactional data. For each one unit increase in “allitemsadjustedcpi”, the intercept (pandemic NAICS) was found to decrease by

-124.01. Interestingly “pce” did not exhibit the same directional trend with one unit increases in the variable eliciting a 15.39 unit increase in NAICS real transactional data.

The final regression iteration, “NAICSTrans. II (Reg 6)”, compared pandemic NAICS real transactional data to that of the various components of pandemic PCE, and the various components of pandemic CPI. Though “NAICSTrans. II (Reg 6)” was the largest regression in terms of distinct variables contained within the model, empirical model six returned zero statistically significant variables. This is likely a symptom of the high number of variables present in the model enlarging the standard errors to a point that eroded accuracy.

It is important to note that multicollinearity was likely present in the model such that there was a high level of correlation between the various independent variables being used to estimate the dependent variable with the effect that the contribution of each independent variable and its variation in the dependent variable cannot be accurately determined. Again, due to the short nature of the pandemic period being measured in the above regression, and the resulting effect shortened time periods have on the number of data point being evaluated, multicollinearity issues also could have generated from the reduced number of data points being measured indicating a higher potentiality for a correlative relationship between the reduced number of X variable data points during the period. As before, this is unfortunately a problem that can only be remedied over time with the inclusion of newer more varied data points. It is important to reiterate that the presence of multicollinearity amongst the various X variables does not constitute the existence of bias within the model. Though ultimately unproductive, this model was still necessary to attempt to ascertain if any of the components of CPI and PCE have an extremely high relationship with that of the NAICS real transactional data. Reduction of the total number of X variables within empirical model six, perhaps by means of creating two additional

regressions that compared NAICS real transactional data to that of the components of PCE only, and a similar second regression that determines the correlation between NAICS real transactional data and the individual components of CPI, may help to reduce multicollinearity within each of the models potentially providing statistically significant variables without the presence of large standard errors.

iii. Results Review and Analysis

The volume of statistically significant variables within the first two empirical models, those that measured CPI and PCE during the non-pandemic period, was far higher than that of the regressions measuring CPI, PCE and NAICS real transactional data during the pandemic period. In empirical model one and two, a total of eight statistically significant variables were identified compared to that of just five in the remaining four regression models. Despite having nearly double the number of regressions analyses, the pandemic period failed to identify nearly as many significant variables to that of the non-pandemic “normal” time period.

In the non-pandemic regressions, the eight identified significant variables were once again: “apparel”, “energy”, “gasolinealltypes”, “medicalcareservices”, “naturalgaspiped”, “services”, “food1”, and the “intercept” value. Turning to the pandemic period, empirical model regressions, aside from the two aggregated PCE and CPI variables, (“pce” and allitemsadjustedcpi”) which were found to be significant, the only categorical variables found to have a statistically high correlative relationship with that of the index being measured were “electricity”, “services”, and “energygoodsandservices2”. Importantly, there is a clear trend of the diversity of major effectors of inflation being reduced during the pandemic period. Notably, the number and variety of inflation driver variables falls during the pandemic period compared to the non-pandemic period. Whilst the non-pandemic inflationary period’s largest effectors of

inflation were differentiated including apparel, energy, medical care, services, and food, pandemic inflationary drivers are characterized as being concentrated around service and energy specific categories.

Overall, the implications of these results point to a clear decrease in the role of all good and service categorical factors in affecting inflation during the pandemic era save for energy and services which have grown to become the only significant drivers behind price increases. This finding departs with the previous twenty years of inflationary price increases where inflation itself was driven by several differentiated factors.

Moreover, though there were many instances of significant variable components of CPI affecting aggregate PCE and vice versa in models one through four, aside from aggregate PCE and CPI themselves, none of the many categorical variables of PCE or CPI were identified as having a statistically significant relationship with real transactional NAICS price increase data. As such, the Consumer Price Index and the Personal Consumption Expenditures Index exhibit a highly correlative relationship between the variables and aggregate values of one another, both CPI and PCE have little in common with the NAICS real transactional price increase data in empirical models five and six. Meaning, such indexes are not accurately correlating with the real consumer price increases being experienced during the pandemic as expressed in the NAICS real transactional data. Hence, the CPI and PCE indexes have seemingly declined since Covid-19 in their ability to accurately predict inflation on a categorical basis during the pandemic era.

V. Policy Implications and Discussion

Analytical findings are pertinent for both researchers seeking to better understand inflationary periods driven by exogenous affects, namely a pandemic, and policy makers,

including politicians and the Federal Reserve, for how to best measure and understand price inflation situationally and then in what capacity and magnitude to act.

At present, CPI, the common barometer of inflation used by the wider market and countless financial intermediaries and institutions, as well as PCE, the key inflation measure used in the Federal Reserve's monetary policy making, are not as effective as they once were before the pandemic. As such, updates and changes to the indices, as well as the addition of other data sets to stress test and ensure either index's accuracy, would help to boost the predicative ability, and predictive relevance of CPI and PCE thereby bolstering market and consumer confidence in inflation measurements and the actions taken to mitigate inflationary pressures.

The disparity in the nature of inflation during the non-pandemic period relative to the pandemic period is pronounced, while food, apparel, medical care, services, and energy were some of the many key affecters of inflation in the non-pandemic period, services and energy now nearly solely constitute the sum of key indicators for inflation during the pandemic period. The significance of omission of such pivotal energy and food data may have consequently been underestimated.

Issue alleviation could include the inclusion of more robust and explanatory energy and food inflationary effects in the CPI and PCE indexes to ensure actual price increases are measured commensurately with the newly adjusted CPI and PCE models. Secondly, the introduction of a standardized transactional price data set, such as the NAICS sorted transactions data set to benchmark and stress test CPI and PCE indexes, would be a welcome addition to the toolkits of the Federal Reserve and financial institutions leading to more accurate and far quicker comprehension of the present state of inflation by policymakers and key stakeholders, and therefore, stronger policy and action by such organizations.

Added research into the nature of pandemic inflationary affects with further back testing into non-pandemic periods would be helpful in developing a greater sample size for stronger, more bias-resistant results. With further examinations into the nature of inflation in the pandemic period compared to the past, particularly as the Covid-19 pandemic progresses, our collective understanding of the extent Covid-19, an exogenous shock, affected price increases, how long the shock may last, and how to mitigate such effects, will only continue to grow. The measurement of other indexes with longer time periods, both those of a market basket of goods and or real transactional data, would allow researchers to further refine our concept of how exogenous shocks affect inflation and reinforce our inflation-centric analysis of the present exogenous shock we find ourselves in. Conclusively, more frequent research examining price inflation in the pandemic utilizing larger timer periods and a combination of traditional methods of inflation analysis as well as alternatives is essential.

VI. Conclusion

This research paper sought to answer the question whether there existed a significant difference in the rates of growth of more extensive measures of product and goods price growth prior to and during the Covid-19 pandemic period compared to that of the normal CPI market basket of goods prior to and during Covid-19. As such, we used a combination of six empirical models that examined the correlative relationship between aggregate and categorical data of the Consumer Price Index (CPI) from the Bureau of Labor Statistics (BLS), the Personal Consumption Expenditures (PCE) Index from the Bureau of Economic Analysis (BEA), and real transactional data delineated by NAICS categorizations sourced through the North American Industry Classification System. Via a variety of robustness checks and assumptions we

painstakingly worked to reduce the potentiality of bias while preserving the accuracy of our empirical model findings.

In our findings we discovered a clear concentration of variables driving pandemic era inflation. Namely, these variables entailed energy and services price increases exclusively. This was a significant departure from the more numerous and differentiated set of significant variables that were found to be guiding inflation during the twenty-year non-pandemic period (apparel, food, services, medical care, and energy). CPI and PCE, the main indexes used in the U.S. to measure inflation by the market and the Federal Reserve respectively, were also found to have reduced predictive ability during the pandemic period compared to the results obtained from the non-pandemic period. As illustrated via the real transactional NAICS delineated data, both aforementioned indexes failed to accurately measure the real price increases occurring for consumers during Covid-19.

The implications of these findings are two-fold. Firstly, institutions should utilize real transactional data sets to qualify the accuracy of their inflation indexes, and potentially retool them in light of the pandemic era decreases in effectiveness of these inflation measurement tools. Secondly, further research with varied back testing of longer time periods and other alternative forms of inflation measurement are necessary to refine and codify the results found in this paper. As for the future, the nature of inflation will remain incredibly difficult to measure without data sets and measurements that adapt to the new post-exogenous shock environment we find ourselves in. The accuracy of such readings will continue to be in question until such recommended changes and further research occur, or the economy returns to a state similar to that of the non-pandemic years from January 2002 – February 2020 where the Consumer Price Index and the Personal Consumption Expenditure Index were more effective than at present.

VII. Works Cited

- Armantier, Olivier, et al. "How economic crises affect inflation beliefs: Evidence from the Covid-19 pandemic." *Journal of Economic Behavior & Organization* 189 (2021): 443-469.
- Claeys, GréGory, and Lionel Guetta-Jeanrenaud. "How has COVID-19 affected inflation measurement in the euro area?." *Bruegel-Blogs* (2021): NA-NA.
- Seiler, Pascal. "Weighting bias and inflation in the time of COVID-19: evidence from Swiss transaction data." *Swiss Journal of Economics and Statistics* 156.1 (2020): 1-11.
- United States. Bureau of Economic Analysis. (Pandemic Mar 2020 – Jan 2022) "Monthly Event Study of Estimates Using Card Transactions ." U.S. Dept. of Commerce, 1 Feb. 2022, <https://www.bea.gov/recovery/estimates-from-payment-card-transactions>. Accessed 1 February 2022.
- United States. Bureau of Economic Analysis. (Non-pandemic Jan 2002 – Feb 2020) "Table 2.8.7. Percent Change From Preceding Period in Prices for Personal Consumption Expenditures by Major Type of Product, Monthly." *Databases, Tables & Calculators by Subject*, U.S. Dept. of Commerce, 1 Feb. 2022, <https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=3&isuri=1&1921=survey&1903=84>. Accessed 1 February 2022.
- United States. Bureau of Economic Analysis. (Pandemic Mar 2020 – Jan 2022) "Table 2.8.7. Percent Change From Preceding Period in Prices for Personal Consumption Expenditures by Major Type of Product, Monthly." *Databases, Tables & Calculators by Subject*, U.S. Dept. of Commerce, 1 Feb. 2022,

<https://apps.bea.gov/iTable/iTable.cfm?reqid=19&step=3&isuri=1&1921=survey&1903=84>.

Accessed 1 February 2022.

United States. Bureau of Labor Statistics. (Non-pandemic Jan 2002 – Feb 2020) - “12-month percentage change, Consumer Price Index, selected categories (past 20 year).” Databases, Tables & Calculators by Subject, U.S. Dept. of Labor, 1 Feb. 2022,

<https://www.bls.gov/charts/consumer-price-index/consumer-price-index-by-category-line-chart.htm>. Accessed 1 February 2022.

United States. Bureau of Labor Statistics. (Pandemic Mar 2020 – Jan 2022) - “12-month percentage change, Consumer Price Index, selected categories (past 20 year).” Databases, Tables & Calculators by Subject, U.S. Dept. of Labor, 1 Feb. 2022,

<https://www.bls.gov/charts/consumer-price-index/consumer-price-index-by-category-line-chart.htm>. Accessed 1 February 2022.

Van Hoomissen, Theresa. "Price dispersion and inflation: Evidence from Israel." *Journal of Political Economy* 96.6 (1988): 1303-1314.