2023

Applications in Macroeconometrics

Corey J.M. Williams
West Virginia University, cjw0074@mix.wvu.edu

Applications in Macroeconometrics

Corey J.M. Williams

Dissertation submitted to the
John Chambers College of Business & Economics
at West Virginia University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Economics

Feng Yao, Ph.D., Committee Chairperson
Bebonchu Atems, Ph.D.
Arabinda Basistha, Ph.D.
Scott Schuh, Ph.D.

Department of Economics

Morgantown, WV
2023

Keywords: Applied Macroeconomics, Applied Econometrics, Productivity, Technical Progress, Inflation, Pass-Through, Too-Big-to-Fail, Returns-to-Scale

Abstract

Applications in Macroeconometrics

by Corey J.M. Williams

The title of this work, "Applications in Macroeconometrics," reflects the unifying theme of my essays: new, and original research in applied macroeconomics, and applied econometrics. This dissertation consists of three chapters specifically looking at producer price inflation pass-through, technical progress, and moral hazard in United States critical banking markets.

My first chapter seeks to quantify the pass-through (PT) and causal direction of producer to consumer prices. We further disaggregate PPI down to commodity-specific indices to quantify disaggregated PPI PT coefficients. By estimating an augmented Phillips Curve, we find short-run PT of aggregate PPI to be around 7 percent. Using a VARX framework, we conduct Toda-Yamamoto tests that show evidence of bidirectional causality between CPI and PPI inflation. When disaggregating down to commodity-specific indices, we find several PPI series that exhibit unidirectional causality and express stronger pass-through coefficients than aggregate PPI.

My second chapter (coauthored with Scott Schuh and Brad Humphreys) expands a niche literature that argues how technical progress and productivity growth–which are ordinarily arduous to measure–can be expressed in athletic outcomes. In particular, we formalize the link between race outcomes, and total factor productivity (TFP) using deterministic and stochastic trend econometric models. A bivariate error-correction model reveals evidence that race outcomes and TFP share a common trend, and that race outcomes adjust to TFP but not vice versa. These results suggest aggregate technological progress partially diffuses to firms and industries, and motivates further investigation of the underlying structural relationships.

My third chapter contributes new evidence to the literatures on banking performance, and too-big-to-fail (TBTF) in US banking. We estimate a restricted translog semiparametric smooth coefficient seemingly unrelated regressions model (SPSC SUR), wherein model elasticities are functions of nonperforming assets, a proxy for moral hazard, to derive nonperformance-adjusted returns-to-scale estimates for critical market banks from 2001 through 2021. Over our full sample, the median critical market bank tends to operate under increasing returns-to-scale while half of all critical market banks exhibit slight decreasing returns-to-scale. Results taken over the past two decades suggest that most TBTF banks have exhausted their economies of scale concurrent with the shrinking competitive landscape.

Acknowledgments

I would like to first thank my parents, Merideth L. Williams (née Mount) and James C. Williams. There are no parents more supportive, loving, and empathetic than you two. I could write a whole dissertation on the qualities you both possess that make you both truly exceptional, and inspiring parents and role models. I would be proud to possess only a fraction of your character traits, and personal qualities. I love you both so immensely, and cannot imagine my life, or my world over the past twenty-eight years without either of you. I thank God every day that you are both in my life. In the same vein, I would like to thank my sisters, Kara N. Williams and Kristin M. Williams, as well as my nephew, Hunter. I am proud to call you my family. I love you all, and look forward to what the future brings for all of us!

Secondly, I would like to thank my advisor, Dr. Feng Yao, for your guidance, support, and willingness to take me on as a pupil. Without a doubt, I am a stronger, and more confident economist as a direct result of your guidance, patience, and expertise. I would also like to thank the other members of my dissertation committee, Dr. Scott Schuh, Dr. Arabinda Basistha, and Dr. Bebonchu Atems for their continued guidance, and support throughout this process. In particular, I owe Dr. Bebonchu Atems an unpayable debt for sparking my interest in economics, and bringing to my attention academia as a potential vocational path. If not for your Principles of Macroeconomics class in Spring of 2015, and your Introduction to Econometrics class in Spring of 2016, I would not have found my love for economics or teaching as I know and feel today. Finally, I would like to thank Dr. Adam Nowak, Dr. Kole Reddig, Dr. Brad Humphreys, and Dr. Bryan McCannon for the research opportunities I have had with all of you. In particular, I want to highlight Dr. Adam Nowak for his charisma, work ethic, and the genuine attention he gave to me and my research endeavors–coauthored or otherwise–throughout my years at West Virginia University.

Thirdly, I would like to acknowledge my closest friends–and in some cases, brothers fraternally–who have supported my career ambitions every step of the way, and who have always been in my foxhole since day one. In no particular order, thank you David Kerr, Nick Rogers, Kenny Galbraith, Steve Roll, Connor Bell, Charli Lomino III, Greg McMullin, and Joy Kanengiser. Know that I will always be in your respective corners as you all achieve your own career goals and personal ambitions. On a related note, thank you to my cohort members Justin Heflin, Bryan Khoo, Shaun Gilyard, Eli Kochersperger, and Dinushka Paranavitana for taking this journey with me.

Finally, thank you to my significant other, Dr. Allyssa Wadsworth, who I can always count on to make me laugh, raise my spirits, and show me the proverbial light at the end of the tunnel when I cannot seem to find one. I love you, and cannot express how pivotal you have been for my success and confidence throughout this endeavor.

Dedication

In memory of, and dedicated to Zoey–the most loving, and gentle companion I could ever ask for. For over nineteen years, you were there for me in my best moments, and during my darkest days. You are as human to me as my closest and most immediate family, and have been an integral part of their lives as well. You shaped my sense of empathy, and humanity in ways no one else could. I love you and miss you.

Zoey (**B**: December 16, 2002 **D**: August 5, 2022)

Quote

The biggest things that I cannot remember
The little things that I cannot forget about when I'm alone in bed
A picture might make my memory better
No matter how much I might regret the fact, it gets into my head

So take my picture, I'm decomposing
Make it a solemn silhouette
'Cause this time is no different
It might help you not forget my feeble voice

So take my picture as I burn
And print it out for all to see
Because words are wind
But image impacts our collective memory

"*From the Sky*" by Luke Hoskin, Mike Ieradi, Tim MacMillar, Cameron McLellan, and Rody Walker

Performed by *Protest the Hero*

Contents

List of Figures

List of Tables

Chapter 1 A Critical Exploration of United States Producer Price Pass-Through

Corey J.M. Williams

1.1 Introduction

United States inflation has reached its highest level in decades. By the end of February, 2022, a forty-year inflation high was recorded at 7.91 percent. Popular media and academics have pointed towards distortions in global supply chains, and shortages in key commodities as primary sources of this recent bout. This rhetoric is grounded in the longstanding theory that producer prices both cause and lead consumer prices.[1] As such, PPI can be, and is thought of as a leading indicator for future consumer prices. Stemming from this train of thought, if producer prices are increasing, one would expect some amount of pass-through to consumer prices. This assumption has established a precedent for utilizing PPI as a predictor for CPI. Despite this longstanding assumption on the relationship between CPI and PPI, and the vested interest forecasters, policymakers, firms and consumers have in understanding the direction that prices are moving, inflation remains difficult to forecast.[2]

A growing literature has found itself interested in investigating the degree to which non-consumer prices pass-through to consumer prices. Efforts in this literature aim to quantify the degree to which a given non-consumer inflation rate precedes and manifests itself in consumer price inflation. Studies in this literature have investigated oil prices, import prices, exchange rates, and producer prices predominantly.

In this study, we turn our attention to producer prices, which express what firms pay for a given good or commodity, rather than what consumers pay for the same product. Traditionally, the literature pays attention to the PPI index for all commodities, which can be thought of as a weighted average of the individual price indices for all commodity groups.[3] However, direct quantification of the aggregate pass-through of producer prices is not well established. Furthermore, PPI has a rather unique feature compared to import prices, oil prices, and exchange rates in that it can be disaggregated down to commodity-specific

[1] This is known as the *production view* theory of pass-through.

[2] As discussed in Stock & Watson (2007), multivariate inflation forecasting models seldom outperform a time-varying univariate model of inflation.

[3] This series is provided by the Bureau of Labor Statistics and possesses the pneumonic *PPIACO* or "PPI by Index: All Commodities."

indices. No piece in the literature has leveraged these commodity-specific producer price indices to investigate PT. Investigating disaggregated producer price indices are of significant benefit to the field if there is potential for a specific commodity or group of commodities to outperform aggregate PPI in predicting consumer price inflation.

Aggregate analysis, particularly of PPI, obscures commodity and industry-specific heterogeneity. If the existing rhetoric points toward disruptions in global supply chains and shortages in key commodities, then inflation's recent movements in the US might be tied only to a small handful of commodities whose effects are otherwise muted by the aggregation process from which the traditional PPI series is constructed. If this is the case, then aggregate PPI PT estimates and inferences may suffer from some degree of bias.

Distilling and differentiating disaggregated PPI PT coefficients from aggregate PT is an exercise that has not been conducted formally, despite the availability of thousands of industry-and-commodity specific indices. From the standpoint of a policymaker that is concerned with PPI PT, the tightening of monetary policy to reign in inflation, particularly cost-push inflation, may be ineffective if aggregate PPI inflation underestimates the true extent of PPI inflation among a select few commodities or industries. Simply put, it is important to ask ourselves if the aggregate PPI suffices for forecasting inflation, and if it does not, it is imperative to understand why. Similarly, the heterogeneity of previously unexplored commodity-specific PPI's may prove to forecast CPI inflation with more precision than the aggregate PPI.

1.2 Literature

While the core literature we consider looks explicitly at the pass-through of producer prices to consumer prices, there are some tangential literatures that contribute to our analysis. These literatures focus on the pass-through of wholesale prices, exchange rates, and oil prices to consumer prices. T. E. Clark et al. (1995) provides the groundwork for investigating PPI PT by attempting to fundamentally address the degree to which producer prices lead consumer prices and aggregates quarterly one-step-ahead rolling forecasts of US CPI inflation and averages them to generate the equivalent yearly forecasts. With forecasts generated from a series of vector autoregression (VAR) models, T. E. Clark et al. (1995) uses the mean absolute error (MAE) criterion to evaluate forecast quality. The theory is that if pass-through from PPI to CPI exists, then a VAR with PPI should forecast CPI strictly better than a VAR without PPI. His results illustrate that multivariate forecast models including PPI inflation tend to generate lower forecast errors over long samples regardless of whether or not one is forecasting core inflation or core goods inflation, but produce mixed evidence over small sample periods in both cases.

While T. E. Clark et al. (1995) investigates the predictive content of a VAR with PPI inflation, a more contemporary approach would be that of Akcay (2011) who compares causality and the direction of causality associated with CPI and PPI inflation for a handful of European nations. The author utilizes monthly data from August, 1995 through December, 2007 to construct vector autoregression models (VAR) for each country. The author then employs Toda & Yamamoto (1995) causality tests to assess the direction of causality associated with the price indices of interest in each country. Causal direction results are mixed, but generally supportive of the production view theory of pass-through.

In a similar vein, Ghazali et al. (2008) motivate their investigation of inflation pass-through by utilizing data from a less developed nation where the traditional view of the relationship between the production chain and consumer goods markets might differ when compared to developed countries. Using monthly Malaysian data from January, 1986 through April, 2007, the authors construct bivariate vector error correction models (VEC) leveraging an existing long-run stochastic trend shared between PPI and CPI. Following similar approaches in the literature, the authors utilize both Granger (1969) causality tests and Toda & Yamamoto (1995) causality tests to address the presence and direction of causality between CPI and PPI in Malaysia. Their findings tend to fall in-line with the production view theory wherein Malaysian PPI inflation leads CPI inflation.

Topuz et al. (2018) study the direction of causality associated with PPI and CPI indices for the United Kingdom and Turkey. Their underlying goal is not only to examine direction of causality alone, but the dominance of one causal direction versus the other. The authors utilize a bivariate VAR model to generate impulse-response functions (IRF), perform a variance decomposition exercise, and conduct Granger (1969) causality tests. Overall, the results of Topuz et al. (2018) support a strong link between CPI inflation and PPI inflation with the production view theory dominating the causal direction associated with most of the results.

The literature that examines PPI inflation alone tends to differentiate itself by nation or country of interest, rather than by disaggregated producer price indices or by commodity and/or industry. The level of analysis in this literature fails to directly quantify pass-through, and prefers to test for causal direction. However, the pass-through literature associated with oil prices and exchange rates provide some empirical methods that can be leveraged and applied to the PPI pass-through literature. Methods borrowed from these tangential literatures allow us to estimate PPI pass-through in the form of a regression coefficient or elasticity.

Looking first at oil price pass-through, consider De Gregorio et al. (2007) who motivate their analysis of oil price pass-through to consumer price inflation from the historical importance of oil price shocks in the US economy. De Gregorio et al. (2007) estimates an augmented Phillips Curve with oil price growth, and construct the pass-through coefficient using the model's reduced form parameters. Additional evidence on the size and significance of the estimated level of oil price pass-through is provided using structural break tests as well as impulse-response functions on a rolling basis. The authors conclude that, across most nations, the significance and magnitude of oil price pass-through has diminished greatly in recent decades. This attributed, in part, to the possibility that inflation expectations are closer to the inflation target as well as incentives to change prices in response to an oil supply shock being much smaller in comparison to earlier decades.[4]

A complementary study would be that of Chen (2009) who estimates oil price pass-through to consumer price inflation using an augmented Phillips curve similar to that of De Gregorio et al. (2007). Short-run pass-through is constructed using reduced form parameters from the estimated augmented Phillips Curve and its corresponding long-run model. The authors conduct a series of structural break tests to assess the stability of the long-run pass-through for several nations by including dummy variable to identify structural breaks. The author concludes that across all nations tested, oil price pass-through to inflation has been much weaker after the turn of the century compared to previous periods. The author posits that such reasons for the weakening of pass-through would be domestic currency appreciation, trade openness and the tightening of monetary policy efforts aimed at controlling inflation.

Looking beyond oil price pass-through, we turn toward contributions made from the exchange rate pass-through literature, which provides additional insights while maintaining some methodological similarities with the oil price pass-through and PPI pass-through literature. An overview of this literature and natural starting point would be Menon (1995), which provides a substantive review of this literature. From a thorough synthesis of the literature, the author identifies five gaps or areas where improvement efforts should be focused. Firstly, the dynamics of pass-through estimates are mixed and pervasive with many papers reporting incomplete or insufficient estimates of exchange rate pass-through. In large part, this is attributable to the mixed lags and lag dynamics of each model the author analyzes. Secondly, the author points out that pass-through varies widely across different nations, particularly when comparing nations in multi-country studies. Thirdly, the author highlights how heterogeneous pass-through studies that focus on a single nation. There are seventy-eight papers that the author takes account of for this literature review, of which, twenty-seven focus on the US specifically (more than any other country). Fourthly, the author critiques how exchange rate pass-through is estimated across different product categories.

Menon (1995) also notes that while the literature has employed a disaggregated approach to measure commodity-specific pass-through, variation across export markets can be extreme. Furthermore, most disaggregated models fail to explain inter-industry variation in pass-through. Finally, the author examines the stability of the pass-through coefficient estimates over time, which also present mixed results in the litera-

[4] This finding is now somewhat of a stylized fact in the pass-through literature, broadly speaking.

ture. More often than not, variation is tested via a one-time structural break, however, some studies yield no change in pass-through, or present evidence of weaker pass-through after the post-break period.

More recent papers are Campa & Goldberg (2005) and Gopinath et al. (2010). Campa & Goldberg (2005) evaluate exchange rate pass-through for 23 OECD nations. A key contribution of their empirical findings are that for nations with low exchange rate volatility, the pass-through elasticity estimates are very low and in some cases, negligible. However, nations with high exchange rate volatility express higher pass-through magnitudes. From an industry level, the authors find that long-run pass–through is at its highest within the manufacturing sector. The authors find that in the short-run for most OECD nations, import prices encompass 46 percent of exchange rate volatility in the short-run and 65 percent in the long-run. The authors describe partial pass-through as the most general explanation for OECD nations, as there are no short or long-run instances of where complete pass-through exists in the confidence intervals for each pass-through coefficient. Gopinath et al. (2010) consider a model of an open economy with nominal rigidities and endogenous currency choice. The findings when applied to the US data show that exchange rate pass-through is vastly different in magnitude for goods priced in non-US dollars (95 percent pass-through) versus goods priced in US dollars (25 percent pass-through).

A complementary piece to Campa & Goldberg (2005) would be Gagnon & Ihrig (2004) who estimate a more theoretically driven six-equation model via OLS in order to evaluate the effect of monetary policy on exchange rate pass-through. The authors apply their model to 20 developed nations and find that the stabilization policies pursued by most central banks have a dampening effect on exchange rate pass-through. When central banks are credible, and agents fully understand the actions their central banks are pursuing, then an aggressive attempt to target or stabilize domestic inflation correspondingly minimizes the pass-through elasticities associated with import prices and exchange rates.

1.3 Methodology & Data

A reasonable starting point in-line with the majority of pass-through studies is to consider CPI inflation and PPI inflation in the aggregate over a long sample. The advantage of starting with this analysis is that it serves as a reasonable benchmark for evaluating pass-through over certain subsamples and for comparing the pass-through of disaggregated producer price indices to aggregate pass-through. Unlike the PPI pass-through literature, however, which mostly leverages bivariate vector autoregression (VAR) or vector error correction (VEC) models, we begin our analysis with an augmentation of the Phillips Curve, which is a common practice in evaluating oil price and exchange rate pass-through to inflation, but not PPI. Broadly speaking, the short-run Phillips Curve takes on the following form described by equation (1.1).

$$\pi_t = \pi^e - \beta(u_t - u_t^n) \tag{1.1}$$

In theory, equation (1.1) expresses relationship between inflation (π_t) and cyclical unemployment ($u_t = u_t^n$), plus a disturbance term. Conveniently, due to Okun's Law, $y_t - y_t^* = \Theta(u_t - u_t^n)$, the short-run Phillips Curve can also be equivalently expressed as it is in equation (1.2).

$$\pi_t = \pi^e + \Gamma(y_t - y_t^*) \tag{1.2}$$

The underlying relationships expressed by both equations (1.1) and (1.2) provide the econometric framework from which Chen (2009) and De Gregorio et al. (2007) conduct their pass-through studies. Following the seminal work of Engle & Granger (1987), we leverage the use of a two-step error-correction model that expresses an augmented Phillips Curve. We opt for this approach over a VAR or VEC model for a few reasons. Firstly, the Phillips Curve expresses a well-known economic relationship whose structural parameters are easily mapped to the reduced form coefficients when estimated. Secondly, there are some critiques with regards to the literature's reliance on bivariate VAR and VEC models. Principally, bivariate systems tend to omit key covariates that may be important in explaining variation in both producer price indices and consumer price indices. Thirdly, as highlighted in pieces like Caporale et al. (2002), the role of monetary policy in mitigating inflation has changed significantly over time, thus bivariate systems cannot adequately capture monetary transmission mechanisms or solidify a structural link between PPI inflation and CPI inflation.

We believe that beginning our analysis with an augmented Phillips Curve overcomes the first critique quite readily by controlling for the output gap. With regards to the second critique, while we do not explicitly include a channel for monetary transmission, we can allow for time-varying estimation in PPI inflation pass-through in our aggregate modeling procedure, which indirectly accounts for changes in the degree to which monetary policy and monetary transmission channels can influence CPI inflation.

Beyond econometric critiques, the Phillips Curve has also been subject to the famous Lucas (1976) critique. However, despite this, many papers would still advocate for the specification of the Phillips Curve or the similar triangle model in describing movements in consumer inflation. Specifically, works like Gordon (2013) and Fuhrer (1995) advocate that the Phillips Curve is still alive and well and in many ways overcomes the Lucas (1976) critique, but can suffer from misidentification. One way to overcome this is to include longer lags of CPI inflation in the estimation procedure.

Additionally, while De Gregorio et al. (2007) and Chen (2009) present an augmented Phillips Curve as given, it should be noted that papers like Huang & Liu (2005) investigate the degree to which the central bank should target CPI inflation, PPI inflation, or both using a true structural model, specifically a dynamic stochastic general equilibrium (DSGE) model. Huang & Liu (2005) present a Phillips Curve with an endogenous cost-push term that arises from producers' marginal costs within the production chain. A key benefit of this augmentation is that it accounts for nominal rigidities present in both the production and consumption sectors. Huang & Liu (2005) provide evidence that the omission of producer prices as a targeting variable typically results in significant welfare loss.

While the Phillips Curve is excellent in its ability to describe short-run price dynamics, in order to leverage an error-correction framework, as is our goal, we need a corresponding long-run model to derive our error correction term. Such a long-run model is a technical necessity for a two-step error-correction model, particularly if our goal is to estimate an augmented Phillips Curve that fundamentally expresses short-run dynamics. We formulate our long-run relationship returning to the production view theory of producer price pass-through, chiefly, the idea that producer prices lead consumer prices. Equation (1.3) describes this view.

$$P_t^C = P_{t-1}^P \tag{1.3}$$

Where P_t^C are consumer prices and P_t^P are producer prices.

If producer prices lead consumers prices, then under the assumption that P_t^P and P_t^C are cointegrated, in steady-state with the absence of temporal variation and other rigidities, there is some stationary linear combination of P_t^C and P_t^P. This relationship is described in equation (1.4).

$$P_t^C - \Phi P_t^P = \epsilon_t \tag{1.4}$$

Where Φ describes the long-run pass-through of producer prices to current period consumer prices and $\epsilon_t \sim$ I(0) is our error-correction term (ECT). Overall, the Phillips Curve as a functional form of inflation along with the corresponding long-run production view theory taken together are well-motivated and reliable for our analysis of pass-through. For the purpose of our study, consider the following aggregate data series described in Table 1.

Table 1: Data Description: Aggregate Model

	CPI	Real GDP	PPI	CPI Inflation	Output Gap	PPI Inflation
Notation	$\log(CPI_t)$	$\log(GDP_t)$	$\log(PPI_t)$	π_t^{CPI}	$y_t - \bar{y}_t$	π_t^{PPI}
Mean	4.42	8.89	4.35	0.03	0.00	0.03
Std Dev.	0.85	0.68	0.74	0.02	0.02	0.05
Minimum	3.08	7.61	3.21	-0.03	-0.09	-0.14
Maximum	5.63	9.89	5.49	0.13	0.04	0.20
Length	1947 : $Q1$–2021 : $Q4$	1947 : $Q1$–2021 : $Q4$	1947 : $Q1$–2021 : $Q4$	1948 : $Q1$–2021 : $Q4$	1947 : $Q1$–2021 : $Q4$	1948 : $Q1$–2021 : $Q4$

Note: Data on prices indices are retrieved from the BLS while data on GDP is from the BEA.

The log levels of the data while relevant for our econometric analysis are less insightful than their respective growth rates or detrended states. We observe that on average, CPI inflation matches PPI inflation at three percent. We see, however, that the consumer price level on average is slightly higher than the producer price level. This could reflect markups or other imperfections at the point of consumption for the consumer that the producer might not experience in their prices for the same good. The levels of our variables of interest are described by Figure 1.

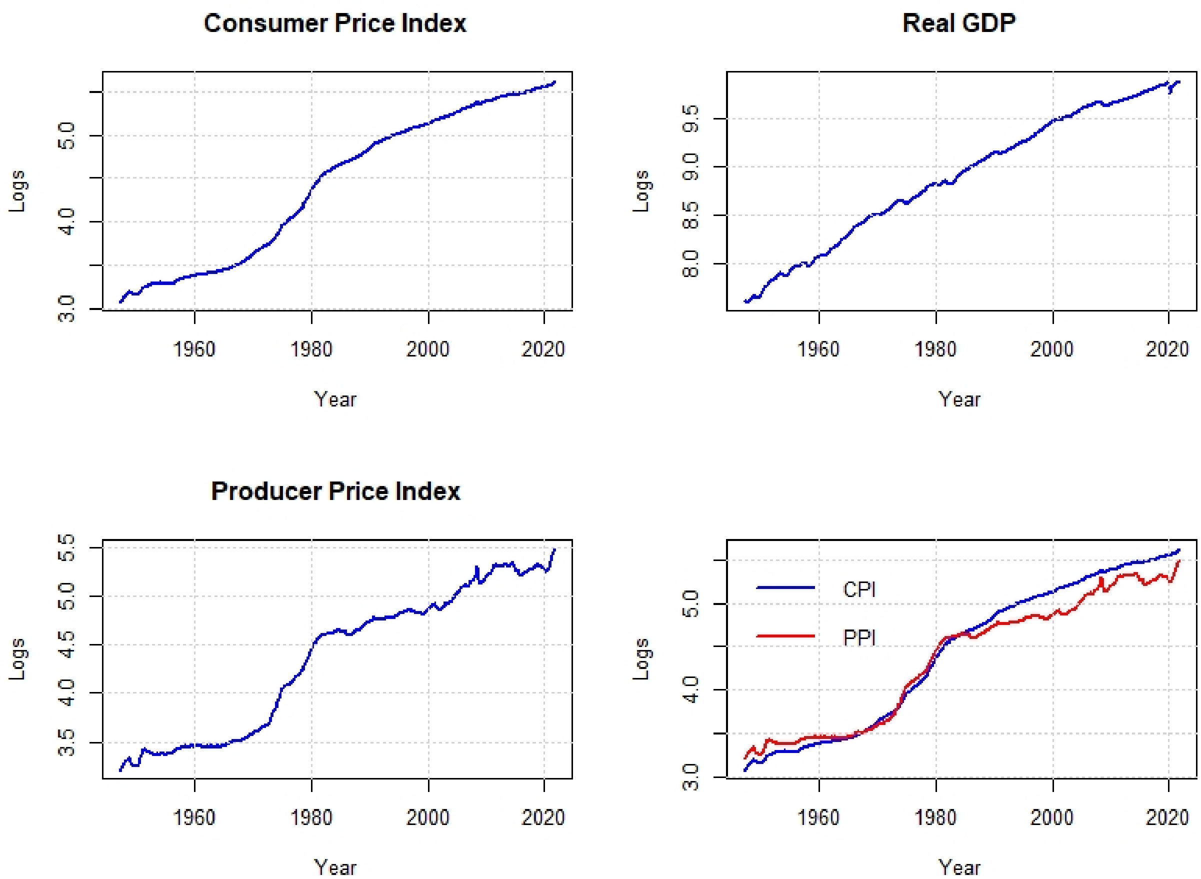


Figure 1: Aggregate Data: Levels

When looking at the extreme moments, we see a stark difference emerge between consumer and producer price inflation. The highest level of consumer inflation in our data is around 13 percent while the maximum producer price inflation is in our sample is more than 50 percent higher. Similarly, the highest deflationary levels (negative inflation) sit at around -3 percent for consumer prices and at -14 percent for producer prices. The growth rates and detrended data (which are ultimately stationary series) possess the following visual characteristics described by Figure 2.

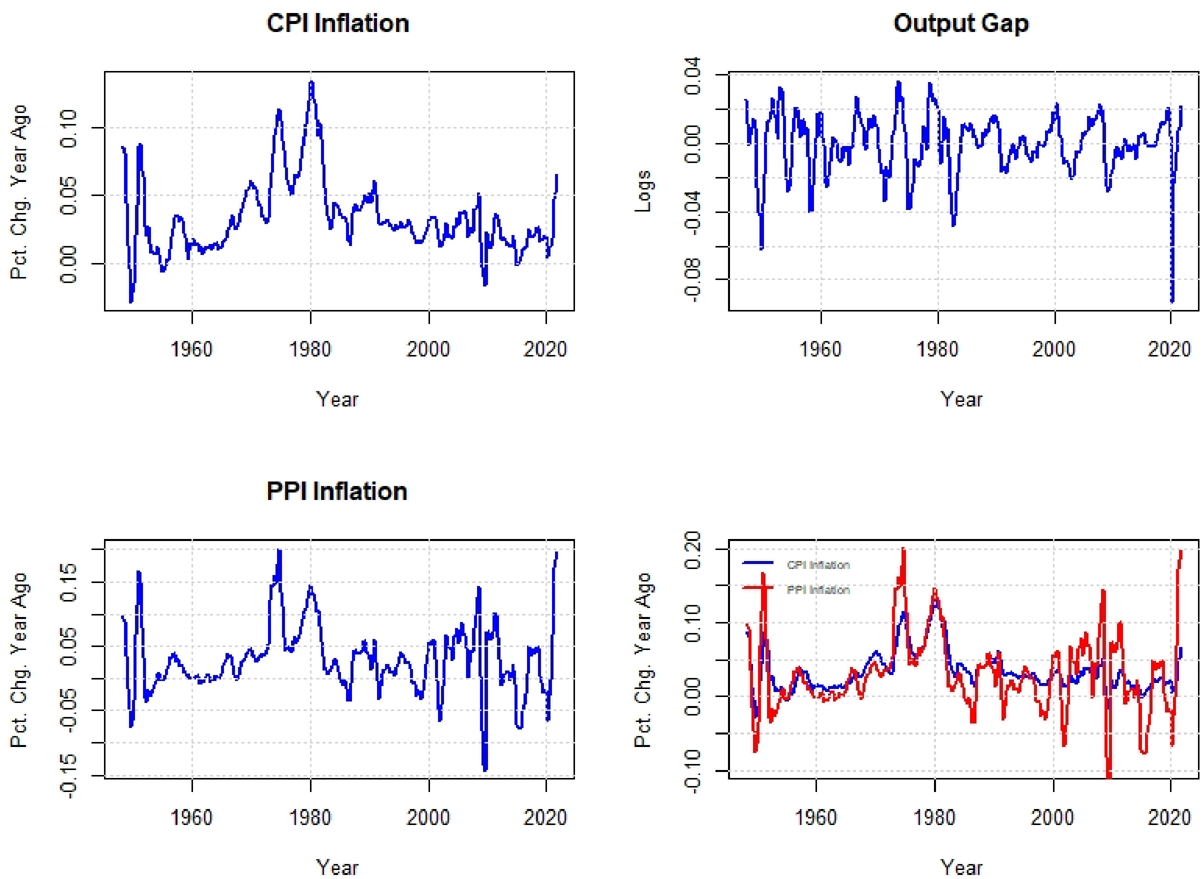


Figure 2: Aggregate Data: Stationary

These distinctions in the extremes underscore the importance of understanding pass-through. If the theory that producer prices lead consumer prices holds true, movements in producer prices should foretell future consumer prices, but not on a one-for-one basis. We also note that the volatility of producer price inflation tends to be higher on average than consumer prices. This volatility is visibly reflected in Figure 2.

1.4 Aggregate Models

In particular, we can see that from the early 1990's onward, the relative volatility of producer price inflation seems to increase despite the volatility of consumer price inflation remaining relatively stable. This visual discrepancy hints at the possibility that the magnitude of pass-through from producer-to-consumer prices has diminished in later years in our sample.

Full Sample Estimation

In order assess PPI pass-through using the data described in Table 1, we leverage a two-step error correction model using the Engle & Granger (1987) method (as utilized in De Gregorio et al. (2007) and Chen (2009)). Prior to estimating a short-run Phillips Curve, we must first estimate a long-run model describing the price level's relationship to GDP in its levels as well as the producer price level (a long-run augmentation to match the short-run augmentation). To reiterate our long-run formulation described by equation 1.4: in a steady-state, if there is cointegration between consumer prices and producer prices, the long-run relationship should describe the share of the producer price level reflected in consumer prices while controlling for output.

An important assumption of the augmented Phillips Curve and its corresponding long-run model is that PPI inflation leads CPI inflation and not the other way around. The literature tends to support this production view theory of PPI inflation in the US with causal direction results affirming this derived in studies such as Caporale et al. (2002). Our long-run model is described by equation (1.5), which is effectively a log transformation of equation (1.4).[5]

$$\log(CPI_t) = \alpha + \beta_1 \log(GDP_t) + \beta_2 \log(PPI_t) + \varepsilon_t \tag{1.5}$$

In equation (1.5), we pay particular attention to the β_2 coefficient, which captures the long-run relationship between consumer and producer prices. We also take note of the residual term ε_t, which is our error-correction term (ECT_t). Both these coefficients are essential for capturing short-run pass-through of producer prices to consumer prices. We estimate equation (1.5) and present the results in Table 2.

Table 2: Long-Run Model: Aggregate PPI

Dependent Variable: $\log(CPI_t)$		
Variable	Estimate	Standard Error
α	-2.25^{***}	(0.12)
$\log(GDP_t)$	0.33^{***}	(0.02)
$\log(PPI_t)$	0.86^{***}	(0.02)
Adjusted R^2	0.99	
Observations	300	

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Before proceeding to estimate a short-run model, we must verify that our data are I(1) or integrated at an order of one. We further must check the stationarity of our error-correction term, ε_t. To test for the order of integration and stationarity of our residual term, we use a series of Augmented Dickey-Fuller (ADF) tests that check for the presence of a unit root. Table 3 presents these results.

[5] It should be noted that this assumption can hold true even if causality is found to be bidirectional, which is also a common finding in the literature.

Table 3: ADF Test Results

Variable	Levels		Transformed	
	T. Stat	5% Critical Value	T. Stat	5% Critical Value
$\log(CPI_t)$	4.70	−1.95	−2.79*	−1.95
$\log(GDP_t)$	8.12	−1.95	−6.72*	−1.95
$\log(PPI_t)$	3.73	−1.95	−5.95*	−1.95
ECT_t (ε_t)	−2.58*	−1.95		

Note: * indicates rejection of the null hypothesis at a level of $\leqslant$ 5%

It is evident that given a null hypothesis that the data are possess a unit root, we fail to reject the null when we leave the data in their levels; however, after transforming the data, we find ourselves rejecting the null hypothesis of the presence of a unit root in favor of an alternative hypothesis that the data do not possess a unit root at a level $\leqslant$ 5 percent. To induce stationarity or mitigate the presence of a unit root for the CPI level, we convert it a log ratio of its current period value, t, relative to its lagged value four quarters, $t-4$, ago: $\pi_t^{CPI} = \log(CPI_t/CPI_{t-4})$. In effect, this transformation provides us the rate of inflation relative to a year ago. We employ the exact same procedure to the PPI level to derive PPI inflation: $\pi_t^{PPI} = \log(PPI_t/PPI_{t-4})$. For the level of GDP, we utilize a Hodrick-Prescott (HP) filter described in Hodrick & Prescott (1997) to extract the trend associated with the level of output, $\bar{y}_t$, and then difference output, y_t, from that derived trend: $y_t - \bar{y}_t$.

In effect, we difference our price levels to eliminate the presence of a unit root and detrend output to do the same.[6] Utilizing two different transformations for our data is important, as it allows our estimated short-run model to match the underlying theory described by equation (1.2).

Given that our variables of interest are non-stationary in their levels (via the presence of a unit root), but stationary as growth rates and relative deviations from a trend, we can confirm that the order of integration associated with each of variables is I(1). Given that our error-correction term does not possess a unit root in its levels, which is to say it is I(0), we can simultaneously confirm the presence of a cointegrating relationship as well. Visually, our error-correction term (ECT_t) is described by Figure 3.

[6]Pieces like Hamilton (2018) would strongly advise against the use of the HP filter, and with good reason, however, we believe that its use is sensible in this empirical application. Furthermore, in our appendix, we show an alternative specification of our model where we use the CBO's estimates for potential output and construct the output gap leveraging that series. The differences between both these specifications are negligible.

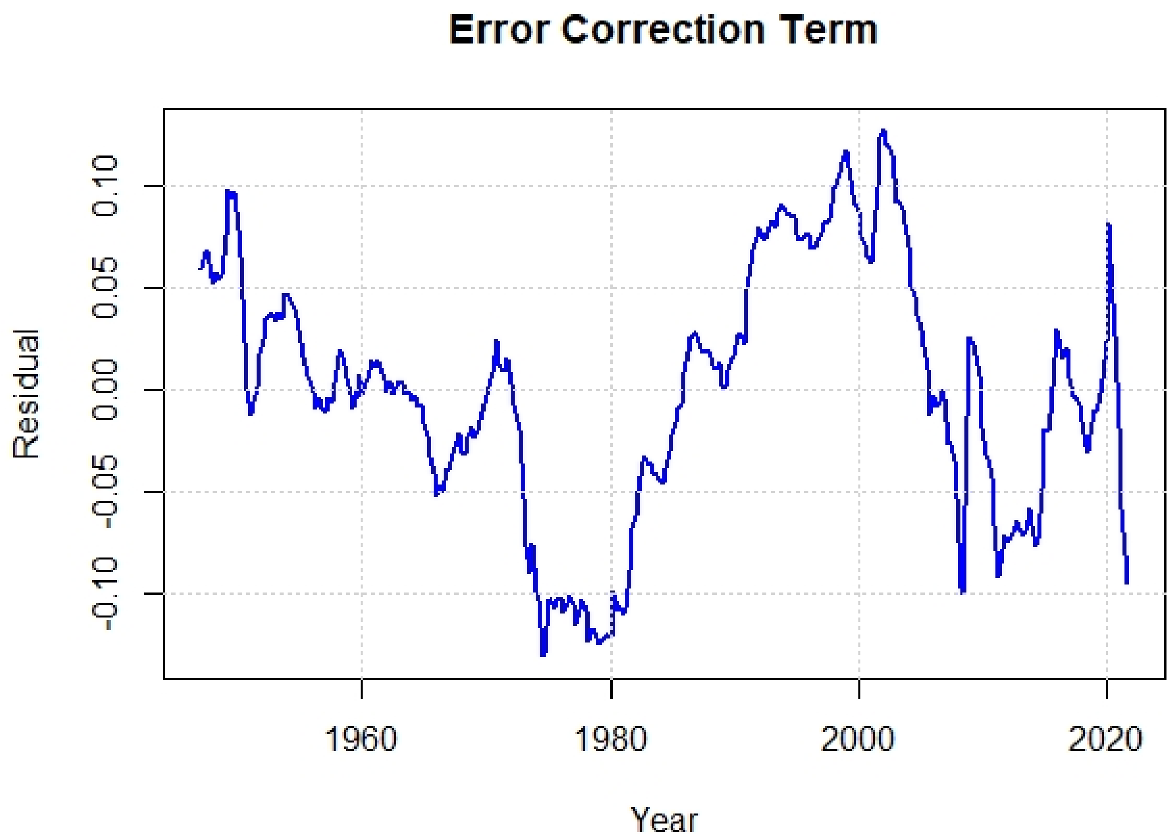


Figure 3: ECT: Aggregate Model

Beyond a unit root test on the residual vector, we can also leverage the Johansen test for cointegration described in Johansen (1995). The Johansen (1995) procedure considers a VAR(1) model (in our case): $\Delta Z_t = \alpha + \Gamma_0 Z_{t-1} + \Gamma_1 \Delta Z_{t-1} + \varepsilon_t$ where $Z_t = [\log(CPI_t), \log(GDP_t), \log(PPI_t)]^T$ and performs an Eigenvalue decomposition of the Γ_0 matrix. The Johansen procedure sequentially tests the rank, r, of this matrix starting at $r = 0$ through the maximum rank of $r = 2$ in this case. We generate test results associated with the trace statistic of each rank tested. The null hypothesis is that there are no cointegrating relationships. Our cointegration test results are described in Table 4.

Table 4: Johansen Test Results: Aggregate Model

Rank	T. Statistic	10% Crit.	5% Crit.	1% Crit.
$r = 0$	75.34	32.00	34.91	41.07
$r \leqslant 1$	20.34	17.85	19.96	24.60
$r \leqslant 2$	6.03	7.52	9.24	12.97

We observe that we can strongly reject the null hypothesis of zero ($r = 0$) cointegrating equations, and find evidence at a level between 5% and 1% of the possibility that there is at most one ($r \leqslant 1$) cointegrating equation, and fail to the reject the null hypothesis that there is a maximum of two cointegrating equations ($r \leqslant 2$). Between our ADF test results associated with the ECT_t and our results from the Johansen (1995) procedure, there is enough evidence of possible cointegration that we can estimate an error-correction model (ECM) with π_t^{CPI} as our dependent variable. We choose to estimate this ECM with four lags in both π_t^{CPI} and π_t^{PPI} to account for a year's (four quarters) worth of persistence in both inflation rates. This lag selection choice mirrors the models utilized in both Chen (2009) and De Gregorio et al. (2007). Our estimated short-run ECM is described by equation (1.6).

$$\pi_t^{CPI} = \alpha + \sum_{i=1}^{4} \beta_i \pi_{t-i}^{CPI} + \gamma(y_{t-1} - \bar{y}_{t-1}) + \sum_{i=1}^{4} \theta_i \pi_{t-i}^{PPI} + \psi ECT_{t-1} + \varepsilon_t \tag{1.6}$$

From equation (1.6), we can estimate short-run pass-through (SRPT) from the coefficient associated with the first lag of PPI inflation (this is known as partial short-run pass-through), θ_1, in conjunction with β_2 from our long-run model and our error correction term coefficient, ψ, which describes the adjustment speed of short-run inflation to the long-run price level. With this in mind, the results of our estimated equation (1.6) are reported in Table 5.[7]

Table 5: Short-Run Model: Aggregate

Dependent Variable: π_t^{CPI}		
Variable	Estimate	Standard Error
α	0.003**	(0.001)
π_{t-1}^{CPI}	1.070***	(0.099)
π_{t-2}^{CPI}	−0.039	(0.130)
π_{t-3}^{CPI}	0.110	(0.143)
π_{t-4}^{CPI}	−0.215**	(0.081)
$(y_{t-1} - \bar{y}_{t-1})$	0.035	(0.038)
π_{t-1}^{PPI}	0.088*	(0.035)
π_{t-2}^{PPI}	−0.125**	(0.041)
π_{t-3}^{PPI}	−0.009	(0.050)
π_{t-4}^{PPI}	0.020	(0.025)
ECT_{t-1}	−0.021*	(0.009)
SRPT	$\theta_1 + (\psi \times \beta_2) = 0.07$ (0.035)	
Adjusted R^2	0.93	
Observations	292	

Note: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

We define short-run pass-through as: $SRPT = \theta_1 + (\psi \times \beta_2) = 0.07$. There is a simple interpretation of this coefficient: suppose producer prices increase by some arbitrary amount such as 20 percent, this

[7] Most pass-through studies building SRPT this way omit standard errors or confidence intervals given SRPT itself is built using point estimates from two distinct models. This makes constructing statistically sound standard errors arduous; however, in an attempt to provide some measure of a crude confidence interval, we use the standard errors from the long-run and short-run models' coefficients to build naive standard errors for SRPT the same way we build the SRPT point estimates themselves. One should view these naive standard errors with some caution.

coefficient tells us that we would expect $0.20 \times 0.07 = 0.014$ of that producer price increase to pass-through to consumer prices in the aggregate. This contribution is first and foremost a direct quantification of aggregate PPI inflation pass-through. Secondly, it serves as a baseline for further Phillips Curve augmentations with disaggregate producer price indices.

Time-Varying Estimation

With aggregate pass-through estimated, we acknowledge that its interpretation has to be thought of as static over the full sample; however, as the literature tells us, pass-through has declined overall in recent decades. Thus, we propose a modification to equation (1.6) to allow for simple time variation. Following Chen (2009), we estimate a single breakpoint associated with CPI inflation using the methodology prescribed in Bai & Perron (1998). We report 1993:Q1 as the estimated break. We create a dummy variable, D_t, which partitions our data from 1947:Q1 through 1992:Q4 taking on a value of $D_t = 0$ and $D_t = 1$ from 1993:Q1 until the end of the sample period, 2021:Q4. An estimated model with time variation takes on the form described by equation (1.7).

$$\pi_t^{CPI} = \alpha + \sum_{i=1}^{4} \beta_i \pi_{t-i}^{CPI} + \gamma(y_{t-1} - \bar{y}_{t-1}) + \sum_{i=1}^{4} \theta_i \pi_{t-i}^{PPI} + \sum_{i=1}^{4} \delta_i (D_{t-i} \times \pi_{t-i}^{PPI}) + \psi ECT_{t-1} + \varepsilon_t \tag{1.7}$$

We define short-run inflation pass through as $\omega_1 = \theta_1 + (\psi \times \beta_2)$ for the pre-break period, and $\omega_2 = \theta_1 + (\psi \times \beta_2) + \delta_1$ for the post-break period. With this in mind, we summarize the results in Table 6. Standard errors are heteroscedastic and autocorrelation consistent as described in Newey & West (1986).

Table 6: Time-Varying Model: Aggregate

Dependent Variable: π_t^{CPI}		
Variable	Estimate	Standard Error
α	0.004***	(0.001)
π_{t-1}^{CPI}	0.807***	(0.142)
π_{t-2}^{CPI}	0.046	(0.166)
π_{t-3}^{CPI}	0.216	(0.144)
π_{t-4}^{CPI}	−0.192*	(0.087)
$(y_{t-1} - \bar{y}_{t-1})$	0.049	(0.031)
π_{t-1}^{PPI}	0.310***	(0.056)
π_{t-2}^{PPI}	−0.228***	(0.062)
π_{t-3}^{PPI}	−0.070	(0.080)
π_{t-4}^{PPI}	0.029	(0.055)
$(D_{t-1} \times \pi_{t-1}^{PPI})$	−0.230***	(0.046)
$(D_{t-2} \times \pi_{t-2}^{PPI})$	0.136	(0.085)
$(D_{t-3} \times \pi_{t-3}^{PPI})$	0.032	(0.094)
$(D_{t-4} \times \pi_{t-4}^{PPI})$	−0.006	(0.056)
ECT_{t-1}	−0.013	(0.009)
ω_1	$\theta_1 + (\psi \times \beta_2) = 0.299$	(0.102)
ω_2	$\theta_1 + (\psi \times \beta_2) + \delta_1 = 0.069$	(0.056)
Adjusted R^2	0.94	
Observations	292	

Note: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

We observe that, as expected, the pass-through between PPI inflation CPI inflation was much stronger in the pre-break period, which corresponds to the era of the Great Inflation in US history among other crises that are historically linked to high aggregate inflation rates. We can calculate the percent change in the pass through between both periods as: $\frac{\omega_2 - \omega_1}{\omega_1} \approx -0.77$, which indicates that the magnitude of the pass-through coefficient declined by roughly 77 percent from pre-break to post-break. The pass-through coefficients indicate that the spillover from PPI inflation to the CPI price level discernibly higher in the pre-break period.

It is well known that during the pre-break era, particularly during the late 1960's through the middle of the 1980's that the demise of the Bretton Woods system in combination with lax monetary policy among other events such as the energy crisis contributed to heightened pass-through rates. Additionally, prior to the Great Inflation, policymakers believed they could exploit the Phillips Curve and as such, assumed inflation would result from derived demand. However, as evident from papers such as P. N. Ireland (1999), the exploitation of the Phillips Curve by policymakers resulted in a level of higher inflation with no changes in output. With this in mind, papers such as Gordon (1975), which focused on food prices primarily, illustrated a more convincing case of cost-push inflation (production view theory) wherein supply chain disruptions manifested themselves in retailer prices. Gordon (1975) further examines external shocks to the price level from the perspective that they can be traced back to input material or commodity shortages. Our own result of pre-break pass-through around 0.30 reinforces this finding.

Overall, our aggregate models directly quantify the pass-through of PPI inflation over a long period of CPI inflation, which has not been directly achieved in the existing literature. We further motivate our findings by electing to use a reduced form model that is linked closely to a well known structural relationship that captures inflation. Secondly, our findings, both over the entire sample and with time-varying consideration, reaffirm the evidence in the literature that the production chain theory of pass-through holds true for the United States.

1.5 Disaggregate Models

We consider the possibility that aggregate PPI obscures possible pass-through coefficients from underlying, and potentially more volatile disaggregated PPI indices. If PPI is to be thought of as a leading indicator for inflation, then an implicit assumption would be that the average for all commodities accurately expresses pass-through despite some underlying commodities being much more volatile than others. This has been the standard assumption of the literature.

However, the patterns of aggregate producer price movements might be misleading if the underlying behavior of relevant commodity-specific producer price indices differ in proportion and direction relative to the aggregate. We acknowledge the possibility that disaggregate PPIs might not provide as much inference as the aggregate PPI, however, as this exploration has not been conducted before, we consider the alternative hypothesis that the pass-through coefficients of at least one disaggregated PPI series provides stronger evidence of pass-through than the aggregate PPI. The null hypothesis being that none of the disaggregate series offer any inference with regards to pass-through.

Disaggregated Data

For our disaggregation exercise, we leverage select quarterly data series from the Bureau of Labor Statistics. The choice of series were based primarily on two pieces of criteria: relevance of index, and length of series. For a fair comparison of long-run and short-run models between both aggregate PPI and a given disaggregated PPI, it is essential to match the data length even at the cost of other potential series. Additionally, while there are several hundreds individual indices to pick from, only a small subset are a sensible first-level disaggregation of PPI. The formal disaggregation from aggregate PPI that use are described by Figure 4 below.

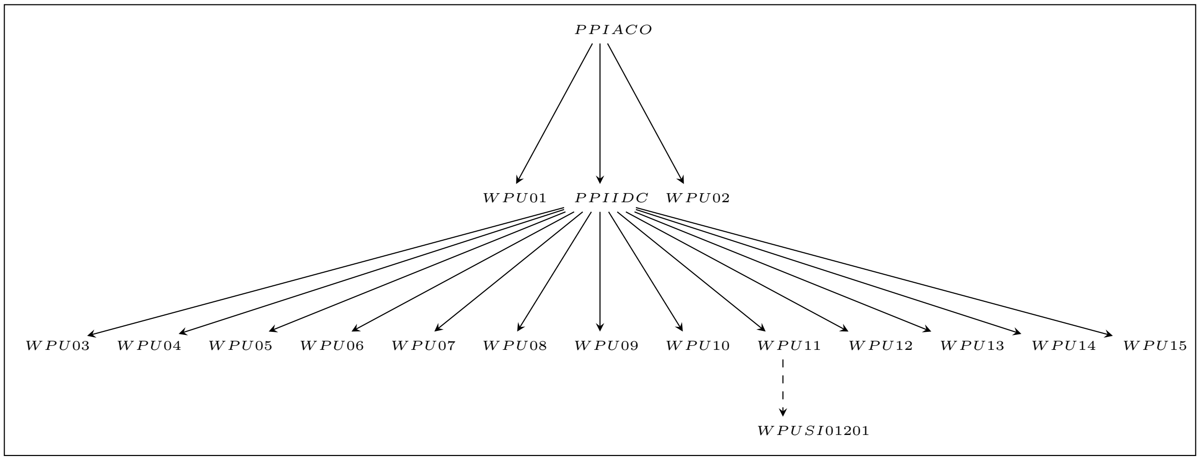


Figure 4: Disaggregation of PPI

With this in mind, we identify sixteen disaggregated PPI series described in Table 7.[8]

Table 7: Data Description: Disaggregate Models

Series Name	Sample Length	Pneumonic	Mean	Std. Dev.	Minimum	Maximum
PPI by Commodity: Farm Products	$1947:Q1$–$2021:Q4$	WPU01	0.02	0.10	-0.28	0.51
PPI by Commodity: Processed Foods and Feeds	$1947:Q1$–$2021:Q4$	WPU02	0.03	0.05	-0.12	0.32
PPI by Commodity: Textile Products and Apparel	$1947:Q1$–$2021:Q4$	WPU03	0.02	0.04	-0.15	0.22
PPI by Commodity: Hides, Skins, Leather and Related Products	$1947:Q1$–$2021:Q4$	WPU04	0.02	0.08	-0.30	0.33
PPI by Commodity: Fuels and Related Products and Power	$1947:Q1$–$2021:Q4$	WPU05	0.04	0.14	-0.51	0.51
PPI by Commodity: Chemicals and Allied Products	$1947:Q1$–$2021:Q4$	WPU06	0.03	0.07	-0.13	0.42
PPI by Commodity: Rubber and Plastic Products	$1947:Q1$–$2021:Q4$	WPU07	0.03	0.06	-0.14	0.42
PPI by Commodity: Lumber and Wood Products	$1947:Q1$–$2021:Q4$	WPU08	0.03	0.08	-0.23	0.41
PPI by Commodity: Pulp, Paper and Allied Products	$1947:Q1$–$2021:Q4$	WPU09	0.03	0.05	-0.07	0.28
PPI by Commodity: Metals and Metal Products	$1947:Q1$–$2021:Q4$	WPU10	0.04	0.07	-0.23	0.39
PPI by Commodity: Machinery and Equipment	$1947:Q1$–$2021:Q4$	WPU11	0.03	0.04	-0.02	0.22
PPI by Commodity: Furniture and Household Durables	$1947:Q1$–$2021:Q4$	WPU12	0.02	0.03	-0.03	0.16
PPI by Commodity: Nonmetallic Mineral Products	$1947:Q1$–$2021:Q4$	WPU13	0.03	0.04	-0.03	0.22
PPI by Commodity: Miscellaneous Products	$1947:Q1$–$2021:Q4$	WPU15	0.03	0.04	-0.06	0.27
PPI by Commodity: Industrial Commodities	$1947:Q1$–$2021:Q4$	PPIIDC	0.03	0.05	-0.18	0.25
PPI by Commodity: Special Indexes: Construction Materials	$1947:Q1$–$2021:Q4$	WPUSI01201	0.03	0.05	-0.09	0.30

Note: All data are sourced from the BLS and descriptive statistics are described in terms of growth rates.

It should be noted that aggregate PPI can be decomposed down to WPU01 through WPU15. WPU01 through WPU15 can be decomposed further into several subcategories. PPIIDC is an aggregation of WPU03 through WPU15, in effect capturing all producer prices not related to farm or food products, which can be somewhat more volatile. Unfortunately, weights are only available in annual buckets from the year 2001 onward. The tradeoff we face is the weights of each respective index for the length of the overall series available. Given the dimensionality of models and the quarterly buckets we leverage, the degrees of freedom associated with our models are best preserved opting for a longer time-series and forgoing the weights.

Similar to the aggregate PPI series, we note that the means of each disaggregate are fairly similar, but differ in their extreme values and volatility. In particular, we note that the minimums and maximums of each

[8] WPU14, transportation equipment, only has data extending as far back as 1969:Q1, hence its exclusion relative to each other series that extend as far back as 1947:Q1. Additionally, the inclusion of WPUSI01201 is to illustrate the possible extensions of this study by exploring further levels of disaggregation of each key commodity group.

commodity are very heterogeneous. Commodities like WPU05 and WPU10 for instance have extremes in either direction that exceed that of the average PPI series. This underscores the importance of disaggregation. The use of the aggregate PPI series clearly mutes the effects that specific commodity groups might have on inflation pass-through alone.

Exploiting this heterogeneity has potential to reveal underlying pass-through coefficients that may be more extreme than the aggregate series. Furthermore, the dynamics of each series might offer stronger predictive content in models aimed at forecasting CPI inflation that might not be captured by models that use the aggregate series. We acknowledge this heterogeneity may not guarantee new insights from each individual disaggregated series; however, if a handful of them have potential to be strictly more beneficial for forecasters and policymakers than aggregate PPI, then the exploration of these series overall will have been considered fruitful. In their log levels, the disaggregated series are each described by Figure 5.

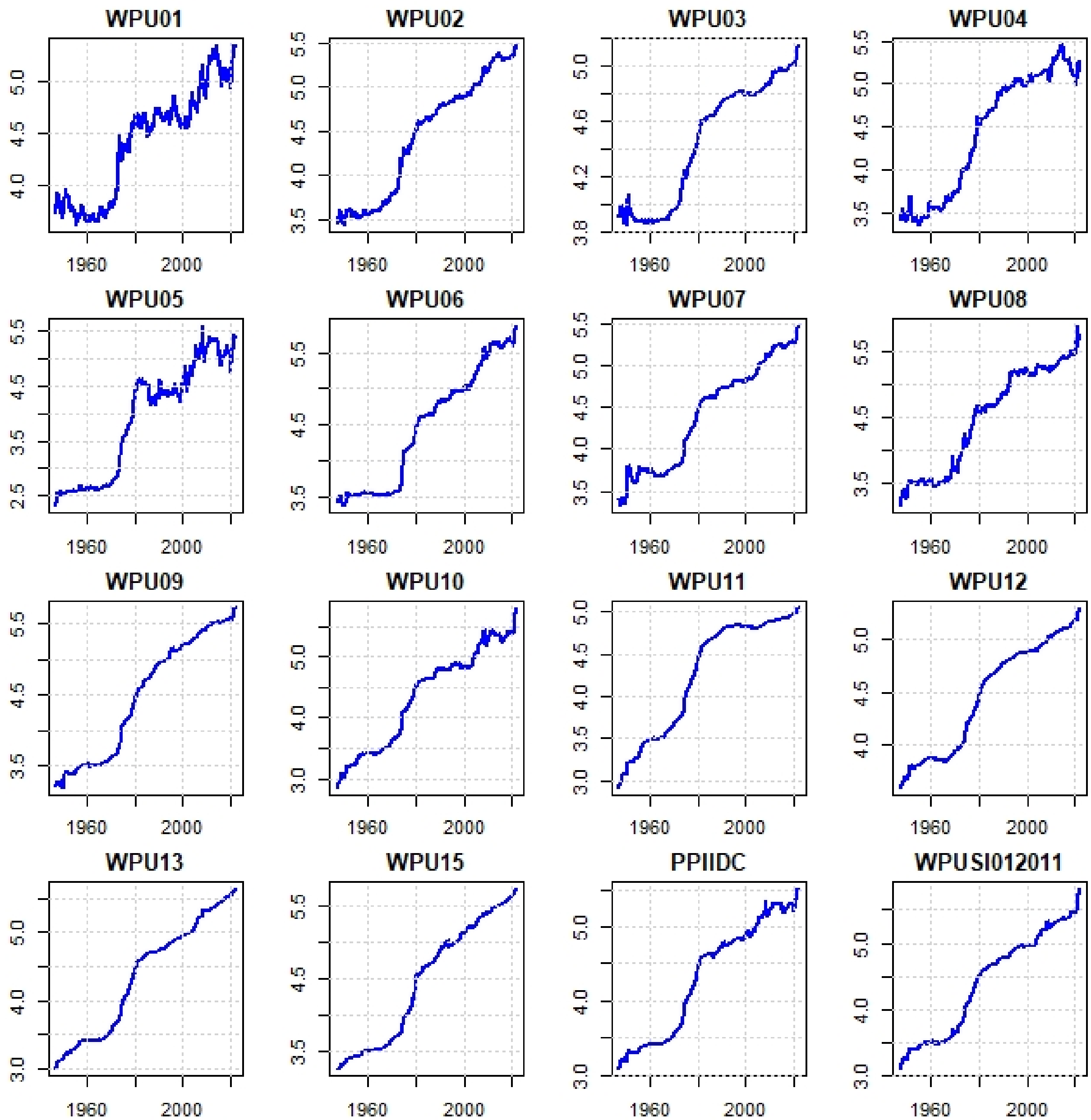


Figure 5: Disaggregate Data: Levels

Visible inspection provides reason to believe our data may be nonstationary or possess a unit root, so we test each series accordingly using the same ADF tests applied at the aggregate level. We conduct our ADF tests with neither a constant, nor a trend, and determine the maximum lag length to be applied to each series using the AIC criteria. The results of our unit route tests are detailed in Table 8.

Table 8: ADF Test Results

Variable	Levels		Transformed	
	T. Stat	5% Critical Value	T. Stat	5% Critical Value
$log(WPU01_t)$	1.32	−1.95	−5.63*	−1.95
$log(WPU02_t)$	2.90	−1.95	−4.10*	−1.95
$log(WPU03_t)$	3.13	−1.95	−6.16*	−1.95
$log(WPU04_t)$	2.10	−1.95	−4.02*	−1.95
$log(WPU05_t)$	1.77	−1.95	−7.20*	−1.95
$log(WPU06_t)$	3.46	−1.95	−6.11*	−1.95
$log(WPU07_t)$	2.92	−1.95	−6.98*	−1.95
$log(WPU08_t)$	3.10	−1.95	−8.15*	−1.95
$log(WPU09_t)$	3.96	−1.95	−4.19*	−1.95
$log(WPU10_t)$	3.50	−1.95	−6.98*	−1.95
$log(WPU11_t)$	3.15	−1.95	−3.63*	−1.95
$log(WPU12_t)$	3.58	−1.95	−1.49	−1.95
$log(WPU13_t)$	2.80	−1.95	−1.84	−1.95
$log(WPU15_t)$	4.29	−1.95	−4.17*	−1.95
$log(PPIIDC_t)$	3.47	−1.95	−2.71*	−1.95
$log(WPUSI01201_t)$	5.60	−1.95	−3.30*	−1.95

Note: * indicates rejection of the null hypothesis at a level of $\leqslant$ 5%.

We observe that after transformation of the data that the vast majority of our disaggregated data test as stationary with the exception of WPU12 and WPU13. Graphically, our data when converted to their stationary states possess the following visual characteristics. Figure 6 graphically illustrates the disaggregated data after being transformed to growth rates.

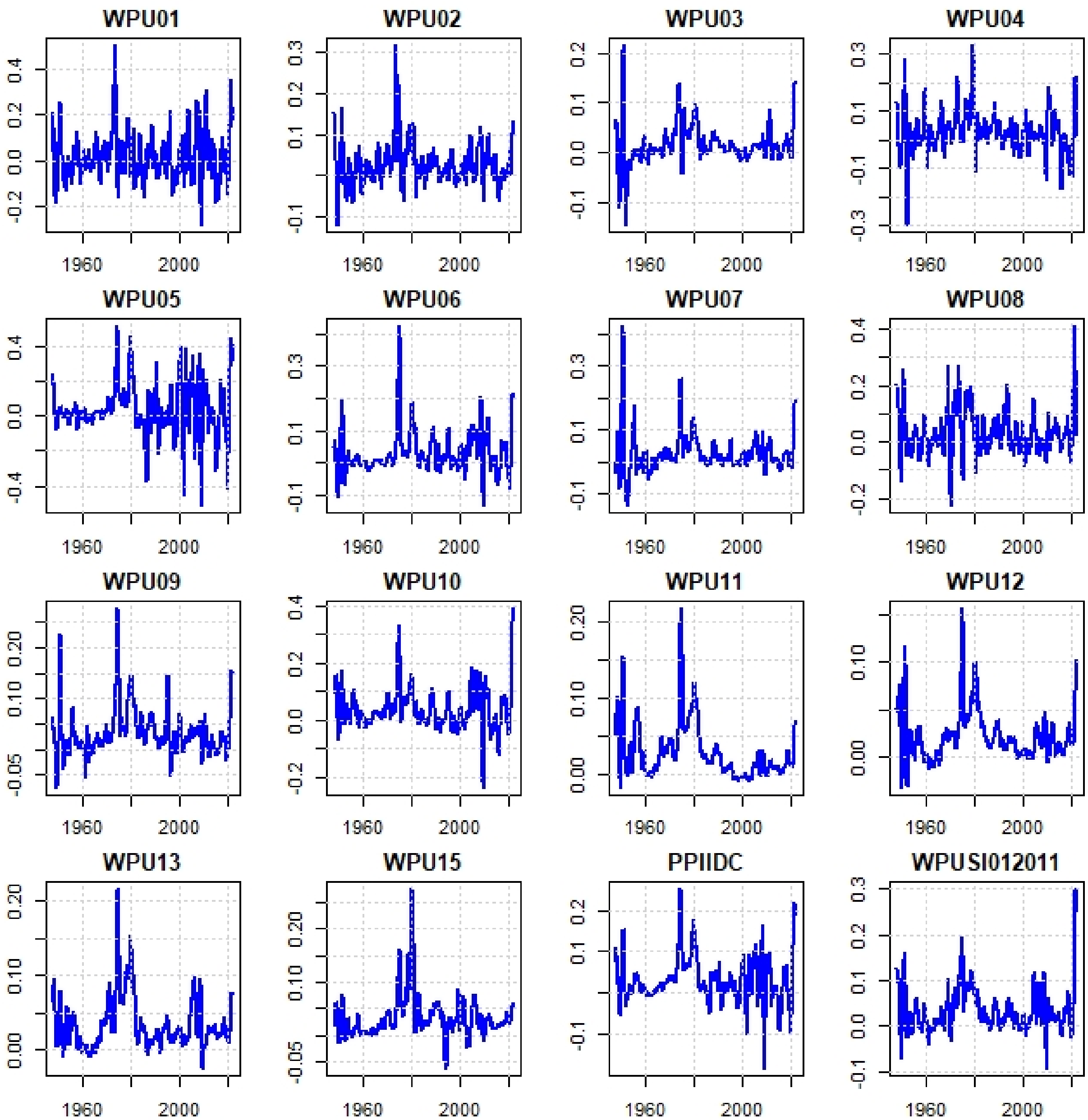


Figure 6: Disaggregate Data: Transformed

Going forward, the importance of our series being strictly I(1) is paramount, especially for constructing valid VEC models and conducting causality tests. Given that WPU12 and WPU13 are not convincingly I(1) based on our above ADF tests, we omit them from further analysis.

Full Sample Estimation

We begin our full sample analysis of disaggregate PPI inflation pass through by creating long-run models for each disaggregate index described broadly by equation (1.8).

$$\log(CPI_t) = \alpha + \beta_1 \log(GDP_t) + \xi_j \log(PPI_{jt}) + \varepsilon_{jt} \quad (1.8)$$

Where ξ_j is the long-run pass through coefficient of some disaggregated PPI index, j. We have fourteen disaggregated producer price indices in log levels, thus, we run this model and report results for fourteen long-run models.

Table 9: Long-Run Models: Disaggregated PPI

Dependent Variable: $\log(CPI_t)$			
Augmentation j	Constant (α)	β_1	ξ_j
WPU01	−5.477***	0.837***	0.554***
	(0.115)	(0.028)	(0.035)
WPU02	−2.817***	0.355***	0.917***
	(0.141)	(0.030)	(0.030)
WPU03	−5.475***	0.547***	1.131***
	(0.042)	(0.013)	(0.021)
WPU04	−3.244***	0.496***	0.733***
	(0.130)	(0.026)	(0.025)
WPU05	−3.203***	0.692***	0.375***
	(0.172)	(0.026)	(0.017)
WPU06	−3.089***	0.540***	0.604***
	(0.142)	(0.026)	(0.022)
WPU07	−3.486***	0.436***	0.911***
	(0.130)	(0.030)	(0.033)
WPU08	−2.593***	0.416***	0.736***
	(0.171)	(0.033)	(0.029)
WPU09	−1.258***	0.211***	0.848***
	(0.074)	(0.014)	(0.011)
WPU10	−2.254***	0.395***	0.725***
	(0.232)	(0.043)	(0.036)
WPU11	−4.207***	0.695***	0.573***
	(0.176)	(0.036)	(0.037)
WPU15	−1.355***	0.208***	0.878***
	(0.087)	(0.016)	(0.014)
PPIIDC	−1.955***	0.304***	0.848***
	(0.118)	(0.022)	(0.020)
WPUSI01201	−1.659***	0.231***	0.911***
	(0.151)	(0.029)	(0.026)

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

We note quite readily that the long-run pass-through coefficients, ξ_j, vary quite dramatically, especially when compared to the long-run PPI pass-through coefficient, β_2, from the aggregate model. Assuming, there is evidence of at least one cointegrating equation between these disaggregated series and aggregate CPI inflation, we can expect the short-run pass-through to also differ over the full sample for each j disaggregate index compared to its aggregate counterpart. Similar to our aggregate model, we test the stationarity of the residuals (ECT_{jt}) associated with each of our long-run models. The results are in Table 10.

Table 10: ADF Test Results: ECT_{jt}

Residual (ECT_{jt})	T. Stat	5% Critical Value
WPU01	−2.28*	−1.95
WPU02	−1.90	−1.95
WPU03	−4.66*	−1.95
WPU04	−3.36*	−1.95
WPU05	−2.57*	−1.95
WPU06	−1.38	−1.95
WPU07	−3.60*	−1.95
WPU08	−3.46*	−1.95
WPU09	−4.74*	−1.95
WPU10	−2.62*	−1.95
WPU11	−2.84*	−1.95
WPU15	−1.99*	−1.95
PPIIDC	−2.92*	−1.95
WPUSI01201	−1.54	−1.95

Note: * indicates rejection of the null hypothesis at a level of $\leqslant 5\%$.

We note that the residuals from the long-run models with WPU02, WPU06, and WPUSI01201 test as non-stationary from the likely presence of a unit root. Additionally, the residuals from each long-run model are presented graphically in Figure 7.

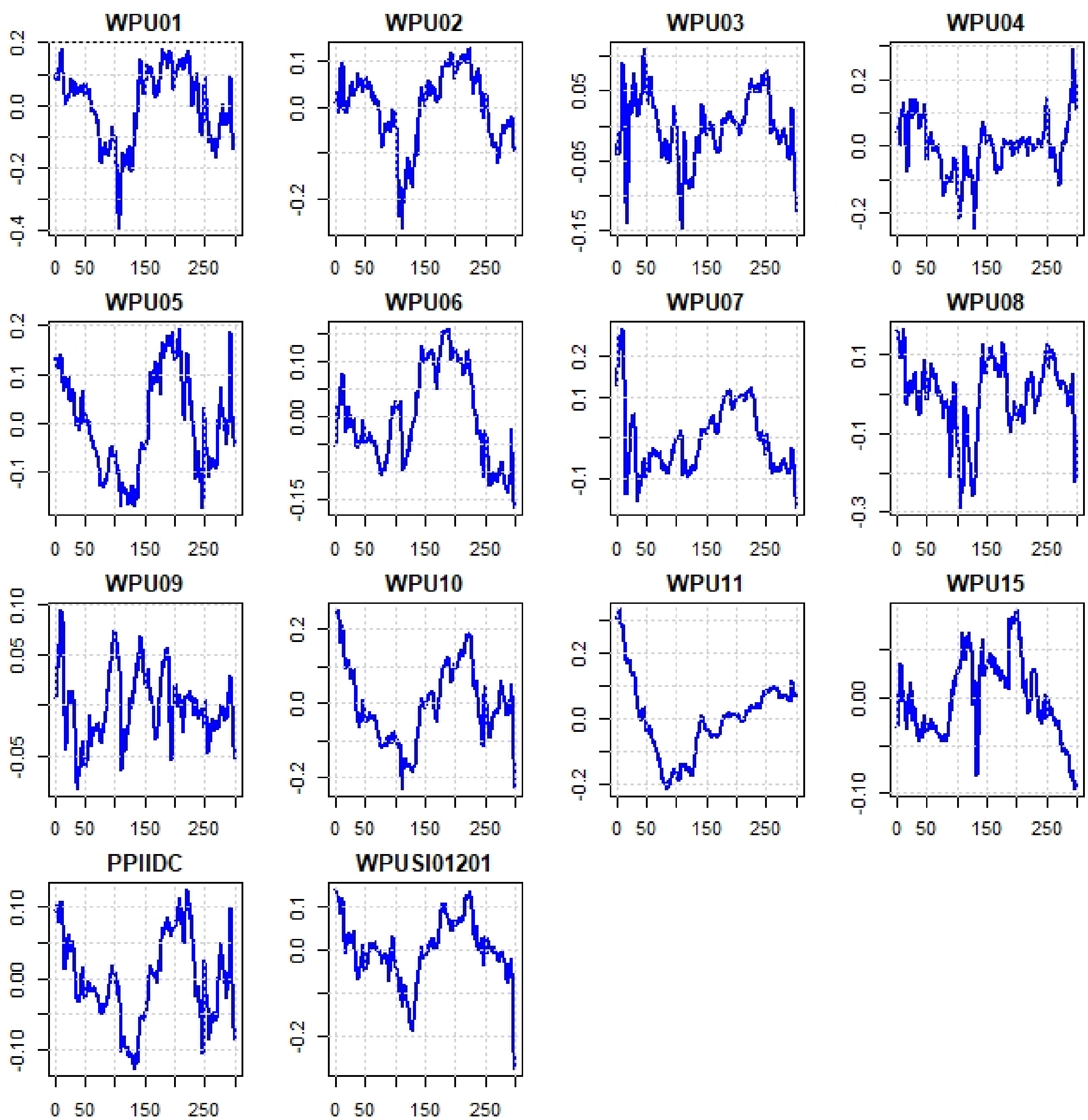


Figure 7: ECT: Disaggregate Models

Given the results of the ADF tests on the residuals for our disaggregated long-run models, we refrain from estimating short-run error correction models from the long-run models augmented with WPU02, WPU06, and WPUSI01201. With this in mind, we estimate eleven error-correction models using the same structure as the augmented Phillips Curve ECM presented in our aggregate model results. In the disaggregate, short-run pass-through from PPI j to CPI is expressed as $\theta_{1j} + (\psi_j \times \xi_j)$.

Table 11: Short-Run Models: Disaggregate

	Dependent Variable: π_t^{CPI}										
Variable/Augmentation	WPU01	WPU03	WPU04	WPU05	WPU07	WPU08	WPU09	WPU10	WPU11	WPU15	PPIIDC
α_j	0.003***	0.004***	0.003***	0.003*	0.003***	0.004***	0.003**	0.003***	0.004***	0.003***	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
π_{t-1}^{CPI}	1.240***	1.156***	1.224***	1.269***	1.236***	1.196***	1.178***	1.162***	1.171***	1.222***	1.132***
	(0.068)	(0.083)	(0.067)	(0.082)	(0.063)	(0.064)	(0.094)	(0.086)	(0.082)	(0.064)	(0.101)
π_{t-2}^{CPI}	−0.312***	−0.315**	−0.230***	−0.268*	−0.350***	−0.261**	−0.265*	−0.145	−0.232*	−0.288***	−0.089
	(0.078)	(0.095)	(0.080)	(0.110)	(0.083)	(0.079)	(0.111)	(0.130)	(0.097)	(0.079)	(0.140)
π_{t-3}^{CPI}	0.114	0.207*	0.048	0.067	0.168*	0.058	0.082	−0.008	0.071	0.039	0.098
	(0.082)	(0.098)	(0.083)	(0.132)	(0.082)	(0.073)	(0.085)	(0.128)	(0.069)	(0.075)	(0.140)
π_{t-4}^{CPI}	−0.134*	−0.180*	−0.099.	−0.125	−0.141*	−0.107*	−0.132.	−0.115.	−0.164*	−0.158**	−0.210**
	(0.055)	(0.080)	(0.056)	(0.077)	(0.070)	(0.054)	(0.073)	(0.067)	(0.063)	(0.057)	(0.077)
$(y_{t-1} - \bar{y}_{t-1})$	0.025	0.010	0.013	0.027	0.034	−0.007	0.043	0.028	0.024	0.031	0.024
	(0.037)	(0.033)	(0.034)	(0.036)	(0.033)	(0.030)	(0.032)	(0.032)	(0.033)	(0.035)	(0.040)
π_{jt-1}^{PPI}	0.006	0.113*	0.034*	−0.003	0.060*	0.012	0.141*	0.086***	0.210.	0.096**	0.069.
	(0.009)	(0.052)	(0.014)	(0.007)	(0.025)	(0.013)	(0.071)	(0.022)	(0.118)	(0.033)	(0.038)
π_{jt-2}^{PPI}	0.006	−0.034	−0.040	−0.006	0.013	0.005	−0.151	−0.163***	−0.293	−0.090.	−0.114*
	(0.012)	(0.121)	(0.029)	(0.011)	(0.054)	(0.012)	(0.129)	(0.043)	(0.194)	(0.051)	(0.045)
π_{jt-3}^{PPI}	−0.017	−0.162	0.011	−0.002	−0.137*	−0.011	0.023	0.092*	0.053	0.017	−0.002
	(0.013)	(0.149)	(0.029)	(0.010)	(0.061)	(0.021)	(0.091)	(0.043)	(0.140)	(0.050)	(0.046)
π_{jt-4}^{PPI}	0.001	0.116	0.004	−0.005	0.082**	−0.001	0.035	−0.011	0.084	0.069*	0.022
	(0.008)	(0.073)	(0.017)	(0.008)	(0.027)	(0.013)	(0.042)	(0.024)	(0.070)	(0.031)	(0.026)
ECT_{jt-1}	−0.012**	−0.040**	−0.015	−0.011*	−0.002	−0.028**	0.030.	−0.014*	−0.014*	0.019	−0.017.
	(0.004)	(0.012)	(0.010)	(0.005)	(0.005)	(0.009)	(0.017)	(0.006)	(0.001)	(0.013)	(0.009)
SRPT	0.000	0.067	0.024	−0.007	0.057	−0.009	0.166	0.076	0.202	0.113	0.054
	(0.009)	(0.052)	(0.014)	(0.007)	(0.025)	(0.013)	(0.071)	(0.022)	(0.118)	(0.033)	(0.038)
Adjusted R^2	0.93	0.94	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
Observations	292	292	292.00	292	292	292	292	292	292	292	292

Note: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; $^{\cdot}p < 0.10$

We note that at the surface level, disaggregated pass-through coefficient estimates vary dramatically category-to-category. Over the full sample, it would seem that WPU11 and WPU15 both pass-through at levels greater than 0.10 while WPU01 and WPU08 have negative pass-through coefficients (albeit, they are also quite close to zero). Some other categories such as WPU03, and WPU07 present pass-through estimates close to the aggregate pass-through coefficient. These results, however varied, present some evidence of the additional information that can be gleaned from the disaggregation of PPI down to its individual commodity groups.

These results are critical for several reasons. Each SRPT coefficient quantifies the importance of commodity-level producer inflation and its pass-through to consumer prices. We note that in several cases, disaggregated PPI SRPT is higher than aggregate PPI SRPT. A critical takeaway is that aggregate PPI pass-through sometimes underestimates producer price pass-through given the extreme heterogeneity among several producer price indices that present stronger rates of pass-through and higher levels of volatility. If the production view theory of producer prices leading consumer prices holds true, the results from Table 11 encourage practitioners to take pause and ask: *which* producer prices?

Time-Varying Estimation

As with our aggregate model, we assume that there is one distinct structural break in the CPI inflation time series and estimate a time-varying augmented Phillips Curve with each disaggregate PPI series in the same spirit as the time-varying model presented in the aggregate section of this paper. Given that we have not changed or altered CPI or CPI inflation between both the aggregate and disaggregate models, we will utilize the same structural break date of 1993:Q1. Results are described in Table 12.

Table 12: Time-Varying Models: Disaggregate

	Dependent Variable: π_t^{CPI}										
Variable/Augmentation	WPU01	WPU03	WPU04	WPU05	WPU07	WPU08	WPU09	WPU10	WPU11	WPU15	PPIIDC
α_j	0.003***	0.005***	0.003***	0.003**	0.003***	0.004***	0.004***	0.003***	0.003***	0.005***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
π_{t-1}^{CPI}	1.232***	1.135***	1.221***	1.253***	1.232***	1.210***	1.155***	1.152***	1.166***	1.171***	1.012***
	(0.071)	(0.085)	(0.074)	(0.089)	(0.069)	(0.061)	(0.087)	(0.095)	(0.083)	(0.069)	(0.126)
π_{t-2}^{CPI}	−0.317***	−0.308**	−0.282*	−0.293**	−0.293***	−0.261***	−0.215.	−0.161	−0.203.	−0.270***	−0.083
	(0.087)	(0.092)	(0.125)	(0.109)	(0.073)	(0.075)	(0.111)	(0.126)	(0.117)	(0.080)	(0.147)
π_{t-3}^{CPI}	0.089	0.201*	0.028	0.063	0.106	0.045	0.061	−0.005	0.069	0.030	0.161
	(0.082)	(0.100)	(0.115)	(0.126)	(0.099)	(0.072)	(0.094)	(0.109)	(0.113)	(0.077)	(0.135)
π_{t-4}^{CPI}	−0.102.	−0.184*	−0.082	−0.118	−0.134.	−0.110*	−0.153*	−0.105	−0.175*	−0.180**	−0.234**
	(0.054)	(0.071)	(0.062)	(0.076)	(0.077)	(0.049)	(0.077)	(0.064)	(0.077)	(0.060)	(0.081)
$(y_{t-1} - \bar{y}_{t-1})$	0.020	0.014	0.013	0.038	0.036	−0.017	0.034	0.018	0.024	0.043	0.030
	(0.033)	(0.032)	(0.032)	(0.031)	(0.032)	(0.031)	(0.032)	(0.031)	(0.031)	(0.036)	(0.031)
π_{jt-1}^{PPI}	0.028*	0.120**	0.073***	0.017	0.085***	0.037*	0.189**	0.174***	0.240*	0.137***	0.267***
	(0.014)	(0.044)	(0.019)	(0.013)	(0.019)	(0.015)	(0.069)	(0.044)	(0.099)	(0.037)	(0.075)
π_{jt-2}^{PPI}	−0.017	0.010	−0.084**	−0.009	−0.024	−0.002	−0.211	−0.254***	−0.35*	−0.104*	−0.291**
	(0.014)	(0.095)	(0.028)	(0.014)	(0.056)	(0.027)	(0.141)	(0.069)	(0.175)	(0.052)	(0.096)
π_{jt-3}^{PPI}	0.011	−0.227.	0.041	−0.010	−0.115	−0.035	0.023	0.120*	0.081	0.003	−0.005
	(0.013)	(0.136)	(0.035)	(0.016)	(0.075)	(0.026)	(0.116)	(0.058)	(0.160)	(0.056)	(0.085)
π_{jt-4}^{PPI}	−0.011	0.158*	0.001	0.003	0.076*	0.020.	0.056	−0.011	0.076	0.102**	0.071
	(0.009)	(0.070)	(0.022)	(0.011)	(0.031)	(0.012)	(0.054)	(0.032)	(0.076)	(0.037)	(0.049)
$D_{t-1} \times \pi_{jt-1}^{PPI}$	−0.039*	0.084	−0.092**	−0.028*	−0.205*	−0.056**	−0.212.	−0.129**	−0.443.	−0.144**	−0.235***
	(0.015)	(0.186)	(0.029)	(0.013)	(0.092)	(0.019)	(0.108)	(0.048)	(0.230)	(0.045)	(0.064)
$D_{t-2} \times \pi_{jt-2}^{PPI}$	0.045*	−0.446	0.120**	0.008	0.259.	0.002	0.228	0.146.	0.525*	0.084	0.223*
	(0.019)	(0.339)	(0.045)	(0.015)	(0.150)	(0.037)	(0.194)	(0.086)	(0.261)	(0.065)	(0.094)
$D_{t-3} \times \pi_{jt-3}^{PPI}$	−0.052**	0.520.	−0.086	0.011	−0.179	0.080.	0.048	−0.056	−0.163	0.047	−0.005
	(0.018)	(0.297)	(0.058)	(0.016)	(0.170)	(0.047)	(0.169)	(0.071)	(0.259)	(0.085)	(0.085)
$D_{t-4} \times \pi_{jt-4}^{PPI}$	0.022.	−0.253*	0.013	−0.013	0.082	−0.051.	−0.115	0.011	0.107	−0.093*	−0.056
	(0.013)	(0.125)	(0.037)	(0.013)	(0.094)	(0.030)	(0.078)	(0.033)	(0.196)	(0.041)	(0.047)
ECT_{jt-1}	−0.009*	−0.033**	−0.010	−0.010.	−0.001	−0.024**	0.032.	−0.011.	−0.013*	0.021	−0.011
	(0.005)	(0.013)	(0.010)	(0.005)	(0.006)	(0.009)	(0.016)	(0.007)	(0.006)	(0.013)	(0.010)
ω_1	0.023	0.082	0.066	0.013	0.084	0.019	0.217	0.166	0.232	0.156	0.258
	(0.029)	(0.230)	(0.048)	(0.026)	(0.111)	(0.034)	(0.177)	(0.092)	(0.329)	(0.082)	(0.139)
ω_2	−0.015	0.166	−0.027	−0.015	−0.122	−0.037	0.005	0.038	−0.211	0.012	0.023
	(0.014)	(0.044)	(0.019)	(0.013)	(0.019)	(0.015)	(0.069)	(0.044)	(0.099)	(0.037)	(0.075)
Adjusted R^2	0.93	0.94	0.93	0.93	0.94	0.94	0.93	0.94	0.93	0.93	0.94
Observations	292	292	292	292	292	292	292	292	292	292	292

Note: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; $^{\cdot}p < 0.10$

We observe that prior to the estimated 1993:Q1 breakpoint, almost all PPI series exhibited much larger pass-through coefficients relative to their post-break estimates. Some attributable reasons as to why this might be the case can be traced back to improvements in monetary policy, but also, events like China joining the World Trade Organization (WTO) in 1999. Given that trade with China has dramatically reduced the costs of imported materials and goods, it stands to reason that certain commodity groups, particularly ones that are imported heavily from China and other international sources, are expectedly expressing lower pass-through coefficients in the post-break period.

Commodities such as plastics, rubbers, metals, machinery, electrical equipment, plastics, and miscellaneous goods are prime examples of commodities that fall into this category; however, textiles appears to express much higher pass-through after the breakpoint. One theory with regards to the behavior of textiles (WPU03) is that the onset of the Covid-19 pandemic greatly increased the demand (and therefore, price level) associated with face masks, gaiters, and other personal protective equipment that fall into the textiles commodity group.

Summary of Pass-Through Coefficients

We take the time to briefly summarize and compare pass-through coefficients between our aggregate and disaggregate settings. Table 13 reports these coefficients in a succinct, compact form.

Table 13: Summary: Pass-Through Coefficients

Coefficient	SRPT	ω_1	ω_2
Aggregate	0.070	0.299	0.069
WPU01	0.000	0.023	−0.015
WPU03	0.069	0.082	0.166
WPU04	0.024	0.066	−0.027
WPU05	−0.007	0.013	−0.015
WPU07	0.057	0.084	−0.122
WPU08	−0.009	0.019	−0.037
WPU09	0.166	0.217	0.005
WPU10	0.076	0.166	0.038
WPU11	0.202	0.232	−0.211
WPU15	0.113	0.156	0.012
PPIIDC	0.054	0.258	0.023

We note a few things immediately. First, the short-run pass-through for the full sample with no time variation tends to be mostly positive between both the aggregate coefficient and each respective disaggregate coefficient, however, the magnitude of each pass-through coefficient in the disaggregate models can vary substantially. Secondly, in the models with time variation, the pre-break pass-through coefficients are all positive, and in all but one case, are larger than their post-break coefficient estimates for the same PPI of interest (this would support that prior to 1993, the production chain view of inflation pass-through held across both aggregate and disaggregated PPIs).

Finally, we note that the presence of negative pass-through is much greater in the post-break era than pre-break era. This raises questions about the degree to which the production chain view of pass-through holds, particularly when one disaggregates PPI down to the commodity level. The prevailing notion is that over a long period, producer prices lead consumer prices, however, our post-break pass-through coefficients call to mind the possibility that this theory has not held as strongly for specific commodity groups in recent decades.

Relative to our null hypothesis, we observe that over the course of the full sample, that there are several disaggregated price indices that express a stronger pass-through rate than the aggregated PPI, thus, we can reject our null hypothesis and favor our alternative that at least one of the disaggregated PPI series provides a greater degree of inference through its respective pass-through coefficients at least when one analyzes the full sample with no time variation. We note, however, that when time-variation is introduced, even with a singular endogenous breakpoint, that the aggregate PPI tends to express higher pass-through than the vast majority of the disaggregated series with the exception of WPU03 (textile products and apparel), which expresses a dramatically higher pass-through coefficient than the aggregate after the breakpoint. Overall, with time variation, there is stronger evidence to support the inference that the aggregate series can provide relative to the disaggregated series, however, at least one of those disaggregated series provides more pass-through inference than the aggregate allowing us to also reject our null hypothesis in favor our same alternative hypothesis for time-varying estimation as well as full sample estimation.

Another possible explanation for the general decline in pass-through after the breakpoint would be the explicit inflation targets the Fed has been committing too since 2012; however, despite the fact that an explicit inflation target was not formalized until 2012, papers like Shapiro & Wilson (2019) provide narrative evidence of discussions of an inflation target dating as far back as 1994. Shapiro & Wilson (2019) further highlight the escalation of interest in an inflation target by the stated preferences of Federal Open Market Committee (FOMC) participants. Between 2000 and 2007, narrative evidence supports a de facto inflation target of 1.5 percent while after the Great Recession, a more formal inflation of two percent became the norm. Anchoring inflation around these targets perhaps not only manifests lower inflation expectations in consumers, but mitigates inflation pass-through within the production chain as well.

It should be noted that only 11 out of 16 disaggregated indices were tested in this analysis based predominantly on data availability to compare to the aggregate analysis. If one were to continue this analysis of disaggregated producer price indices, one could work with a smaller time window for comparison, but leverage not only commodity-specific indices, but also industry-specific indices along with a variety of other commodity-specific indices that simply have not been available for as long as the ones utilized in this study. Furthermore, reducing the sample period and using higher frequency data to leverage each commodity groups' respective weight would also provide a richer analysis beyond the scope of this paper.

1.6 Rolling Regressions

Following the methodology of Chen (2009), we construct rolling regressions to capture the changes in our pass-through coefficient over rolling windows. While we can capture pass-through change using a one-time structural break, pass-through likely changes in waves, thus, rolling regressions should be sufficient for uncovering the dynamics of the pass-through coefficient over each rolling window. We use a rolling window of 40 quarters (ten years) and estimate the coefficients for our long-run models and error correction models one-step at a time for each roll up through the end of the sample period. We start by presenting the results for our aggregate model, and then for our disaggregate models.

Aggregate Model

The our aggregate model, we construct two rolling regressions from which our pass-through coefficients are constructed. First, we reconstruct our long-run regression using a rolling window of 40 quarters, thus, the first reported coefficient estimates start in 1956:Q4, and the last estimated coefficients are reported in 2021:Q4. A second short-run, error-correction model is reconstructed using the same rolling window with coefficient estimates reported beginning in 1958:Q4, and ending in 2021:Q4.

For all reported model coefficient estimates, we reconstruct our pass-through coefficient: $\theta_1 + (\psi \times \beta_2)$ for each rolling window estimation from 1956:Q4 through 2021:Q4. Thus the pass-through coefficient reported for 1956:Q4 is constructed using the β_2 coefficient from a long-run model that uses data from 1947:Q1 through 1956:Q4 and using the ψ and θ_1 coefficients estimated from the error correction model estimated over the same window over the same horizon. With this in mind, the pass-through coefficient for 1957:Q1 is constructed from the same model parameters, but uses data from 1947:Q2 through 1957:Q1. This pattern continues for each subsequent period up through the end of the sample.

Graphically, we present our the following rolling window coefficients: β_2, ψ, θ_1, and our SRPT coefficient in Figure 8.

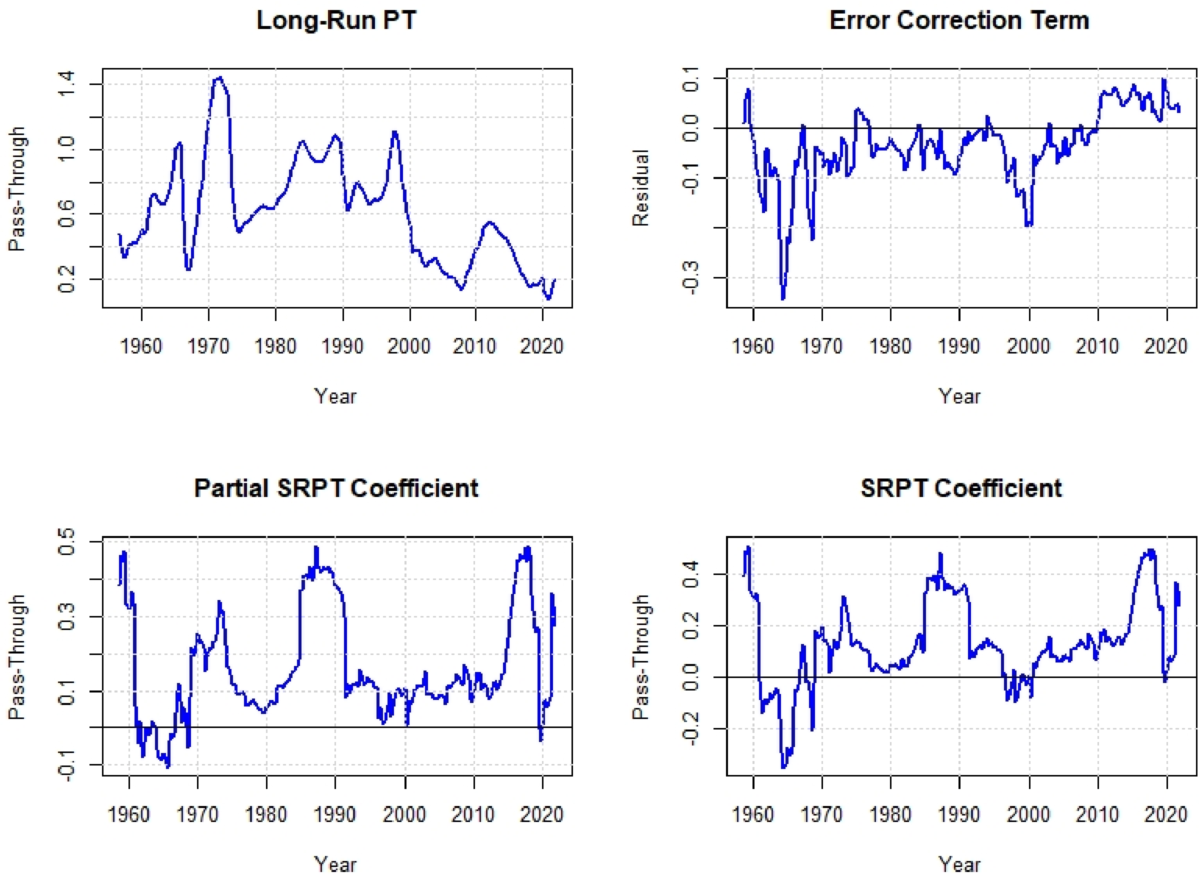


Figure 8: Rolling SRPT: Aggregate

We calculate the mean pass-through from our rolling models to be roughly 0.13, which is comparably larger than our full-sample estimation of pass-through, 0.07. These findings and visual characteristics tend to reinforce the production chain view of PPI inflation pass-through in the United States. We note that the shape of our rolling pass-through estimates tend to follow the same shape as CPI inflation, and that there is evidence over some periods of significant negative pass-through, which would be evidence of the derived-demand view, or CPI pass-through to PPI inflation.

These results illustrate a dominance of the production chain view theory of pass-through and directly quantifies pass-through over various subsamples in our data. This is of value to forecasters, and policymakers alike who seek to better understand the nature of producer price pass-through, particularly in periods of economic downturn.

Disaggregate Models

We repeat our analysis of rolling short-run pass-through coefficients leveraging our disaggregate error correction models. For all eleven disaggregated producer price indices, we utilize rolling windows of forty periods. Rolling coefficient estimates are reported starting in 1958:Q1 through 2021:Q4. Graphically, each disaggregated index's rolling pass-through coefficient is reported graphically in Figure 9.

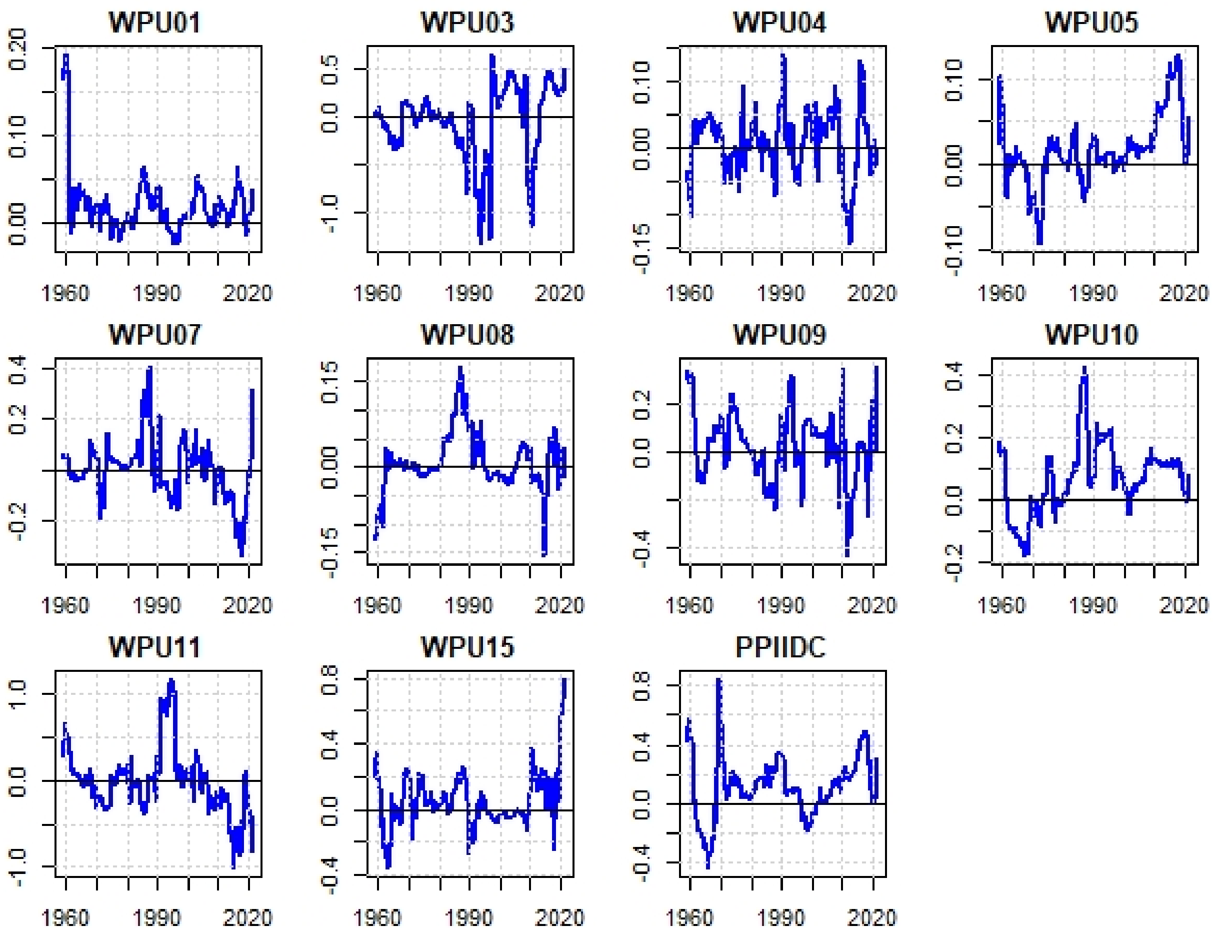


Figure 9: Rolling SRPT: Disaggregate

Unlike our rolling estimates for aggregate PPI, whose pattern mimics CPI inflation over the same period, that the disaggregate indices are generally far more volatile. WPU03, WPU04, WPU07, and WPU09 in particular swing frequently from positive-to-negative (and negative-to-positive) pass-through and tend to be much more volatile overall than the other indices. Reported in Table 14 are the means and standard deviations associated with each SRPT coefficient from the rolling regression series.

Table 14: Rolling SRPT Moments: Disaggregate

Augmentation	Mean	Standard Deviation
WPU01	0.022	0.036
WPU03	−0.063	0.389
WPU04	0.006	0.050
WPU05	0.018	0.040
WPU07	0.004	0.121
WPU08	0.010	0.052
WPU09	0.027	0.143
WPU10	0.071	0.114
WPU11	−0.007	0.387
WPU15	0.054	0.169
PPIIDC	0.119	0.204

We note that the mean SRPT in a rolling environment are fairly low and, in most cases, close to zero. However, the standard deviations can be quite extreme with WPU03, WPU09, and PPIIDC expressing a standard deviation of around 0.39, 0.14, and 0.20, respectively. These results contrast sharply with the aggregate rolling results. In the aggregate, average SRPT tends to be around 0.12, but with a standard deviation of around 0.06. In the disaggregate, we observe generally weaker mean levels of SRPT, but with much higher individual standard deviations. An implication of this is that on average, SRPT seems to be stronger in the aggregate, however, the disaggregated series do well to describe the most extreme swings in SRPT.

1.7 Vector Models

Vector models are the more traditional approach for capturing pass-through in the literature. What they might lack in direct quantification of pass-through, they offset by their ease of use for causality testing. We start this analysis by considering only variables that tested as stationary in our previous analysis, which includes all disaggregated PPI's with the exception of WPU12 and WPU13. We begin with an aggregate bivariate VAR with CPI and PPI and then construct additional disaggregated bivariate VARs with CPI and each j disaggregated PPI.

1.7.1 Aggregate Vector Model

We start with an aggregate bivariate vector model. As with all time series analysis, we begin with the unit root tests associated with each of our variables: CPI_t, and PPI_t. We note from earlier that both these variables test non-stationary in their levels, but stationary in their growth rates forms where $\pi_t^{CPI} = \log(CPI_t/CPI_{t-4})$ and $\pi_t^{PPI} = \log(PPI_t/PPI_{t-4})$. We note from our earlier results and from previous studies in the literature that both CPI_t and PPI_t are cointegrated, however, as these are bivariate models (omitting GDP), we present a new set of Johansen test results for a bivariate vector of $Z_t = [\log(CPI_t), \log(PPI_t)]^T$ in Table 15.

Table 15: Johansen Test Results: Bivariate Aggregate Model

Rank	T. Statistic	10% Crit.	5% Crit.	1% Crit.
$r = 0$	46.91	17.85	19.96	24.60
$r \leqslant 1$	6.68	7.52	9.24	12.97

These results inform us that we can strongly reject the null hypothesis of no cointegrating relationships and favor an alternative hypothesis that there is at least one cointegrating relationship. We proceed to test for the optimum lag length of this bivariate system and using the AIC criteria elect a lag length of $\rho = 6$. We use these results to guide the construction of a bivariate VAR in levels with exogenous regressors to conduct Toda-Yamamoto causality tests.

Aggregate Causality Testing

For causality testing, we follow the trend in the literature of leveraging the causality test proposed in Toda & Yamamoto (1995). The Toda & Yamamoto (1995) causality test considers a VAR in levels with a maximum, r_{Max}, cointegrating relationships, and ρ optimum lags considered as prescribed by the AIC criteria. An augmented $VARX(\rho + r_{Max})$ is then estimated where the VARX is constructed with the optimum lag length of ρ plus additional lag lengths equal to r_{Max} treated as exogenous regressors. An modified Wald test is then applied that adopts a χ^2 function with degrees of freedom equal to the number of exogenous lagged variables.

By artificially augmenting the order of a bivariate VAR, we account for the existence of causality and direction of causality in both the short-run and the long-run. From a performance standpoint, because the VARX is constructed with the endogenous variables in their levels, rather than their first differences, the test is strictly more efficient than Granger (1969) causality tests by treating each variable as seemingly unrelated to one another.[9]

Because our Z_t vector possesses an optimum lag length of $\rho = 6$ and there are $r_{Max} = 1$ cointegrating relationships, we augment a VAR with the seventh lags of our Z_t vector as exogenous regressors. Formally, our $VARX(6 + 1)$ is described by equation (1.9).

$$Z_t = \alpha + \sum_{i=1}^{6} \Gamma_i Z_{t-i} + \Theta X_{t-7} + \varepsilon_t \tag{1.9}$$

[9] Another handy feature of Toda & Yamamoto (1995) causality testing is that a rejection of the null hypothesis that variable a does not cause variable b simultaneously confirms cointegration between variables a and b as well; however, failure to reject the null in a bivariate system is not sufficient for determining $r = 0$ cointegrating relationships.

In this model our contemporaneous endogenous vector is $Z_t = [\log(CPI_t), \log(PPI_t)]^T$ and our augmented exogenous vector is $X_{t-7} = [\log(CPI_{t-7}), \log(PPI_{t-7})]^T$. We summarize our aggregate results in Table 16.

Table 16: Toda-Yamamoto Causality Test Results: Aggregate

Vector	Null Hypothesis	Lag (ρ)	$\rho + r_{Max}$	F–Stat	P–Value	Causal Direction
$Z_t = [\log(CPI_t), \log(PPI_t)]^T$	H_0: CPI does not cause PPI H_0: PPI does not cause CPI	$\rho = 6$	$\rho + r_{Max} = 6 + 1 = 7$	3.71 3.93	0.000 0.000	$CPI \longleftrightarrow PPI$

We note that we reject the null hypotheses associated with both causality tests and find that there is bidirectional causality associated with US CPI and PPI in the aggregate. This result is not surprising and, in some ways, acts as a replication of results well-known within the literature. For the purpose of our analysis, Table 16 serves as a benchmark from which we can compare causality and causal direction from other bivariate VARX models containing CPI and a given commodity-specific disaggregated PPI series.

1.7.2 Disaggregate Vector Models

Given that we know our variables are first-difference stationary, and subsequently I(1), we begin analysis of our disaggregate models by first presenting the results of Johansen cointegration tests. Each bivariate vector tested consists of the CPI level and a PPI level associated with one of our disaggregated variables. The results are presented in Table 17.

Table 17: Johansen Test Results: Disaggregate Models

Vector	Hypothesis	T. Statistic	10% Crit.	5% Crit.	1% Crit.
$Z_t = [\log(CPI_t), \log(WPU01_t)]^T$	H_0: $r = 0$ H_A: $r > 0$	43.77	17.85	19.96	24.60
$Z_t = [\log(CPI_t), \log(WPU03_t)]^T$	H_0: $r = 0$ H_A: $r > 0$	53.67	17.85	19.96	24.60
$Z_t = [\log(CPI_t), \log(WPU04_t)]^T$	H_0: $r = 0$ H_A: $r > 0$	40.17	17.85	19.96	24.60
$Z_t = [\log(CPI_t), \log(WPU05_t)]^T$	H_0: $r = 0$ H_A: $r > 0$	42.27	17.85	19.96	24.60
$Z_t = [\log(CPI_t), \log(WPU07_t)]^T$	H_0: $r = 0$ H_A: $r > 0$	43.67	17.85	19.96	24.60
$Z_t = [\log(CPI_t), \log(WPU08_t)]^T$	H_0: $r = 0$ H_A: $r > 0$	55.74	17.85	19.96	24.60
$Z_t = [\log(CPI_t), \log(WPU09_t)]^T$	H_0: $r = 0$ H_A: $r > 0$	56.12	17.85	19.96	24.60
$Z_t = [\log(CPI_t), \log(WPU10_t)]^T$	H_0: $r = 0$ H_A: $r > 0$	46.33	17.85	19.96	24.60
$Z_t = [\log(CPI_t), \log(WPU11_t)]^T$	H_0: $r = 0$ H_A: $r > 0$	51.61	17.85	19.96	24.60
$Z_t = [\log(CPI_t), \log(WPU15_t)]^T$	H_0: $r = 0$ H_A: $r > 0$	33.06	17.85	19.96	24.60
$Z_t = [\log(CPI_t), \log(PPIIDC_t)]^T$	H_0: $r = 0$ H_A: $r > 0$	45.79	17.85	19.96	24.60

In all cases, we find that for each bivariate vector of I(1) indices, there is evidence supporting a maximum of one cointegrating relationship.

Disaggregate Causality Testing

We proceed with Toda & Yamamoto (1995) causality tests of the disaggregated bivariate vectors. Following the same process as our aggregate models, we identify the optimum lag length for each vector in accordance with the AIC criteria, and then consider exogenous lagged vectors of one lag greater than the optimum. We then test for causal direction and report corresponding F-statistics and p-values. We consider a null hypotheses that no causal relationship exists between PPI or CPI nor CPI and PPI (where PPI in this case is a specific disaggregated index). A rejection of the null hypothesis and causal direction is reported if the p-value is significant at a level of 5% or less. We report these results in Table 18.

Table 18: Toda-Yamamoto Causality Test Results: Disaggregate

Vector	Null Hypothesis	Lag (ρ)	$\rho + r_{Max}$	F–Stat	P–Value	Causal Direction
$Z_t = [\log(CPI_t), \log(WPU01_t)]^T$	H_0: CPI does not cause WPU01 H_0: WPU01 does not cause CPI	$\rho = 6$	$\rho + r_{Max} = 7$	4.57 2.39	0.00 0.03	$CPI \longleftrightarrow WPU01$
$Z_t = [\log(CPI_t), \log(WPU03_t)]^T$	H_0: CPI does not cause WPU03 H_0: WPU03 does not cause CPI	$\rho = 12$	$\rho + r_{Max} = 13$	2.33 2.28	0.01 0.01	$CPI \longleftrightarrow WPU03$
$Z_t = [\log(CPI_t), \log(WPU04_t)]^T$	H_0: CPI does not cause WPU04 H_0: WPU04 does not cause CPI	$\rho = 6$	$\rho + r_{Max} = 7$	4.21 2.52	0.00 0.02	$CPI \longleftrightarrow WPU04$
$Z_t = [\log(CPI_t), \log(WPU05_t)]^T$	H_0: CPI does not cause WPU05 H_0: WPU05 does not cause CPI	$\rho = 6$	$\rho + r_{Max} = 7$	0.34 1.30	0.92 0.26	No Causality
$Z_t = [\log(CPI_t), \log(WPU07_t)]^T$	H_0: CPI does not cause WPU07 H_0: WPU07 does not cause CPI	$\rho = 6$	$\rho + r_{Max} = 7$	3.61 5.88	0.00 0.00	$CPI \longleftrightarrow WPU07$
$Z_t = [\log(CPI_t), \log(WPU08_t)]^T$	H_0: CPI does not cause WPU08 H_0: WPU08 does not cause CPI	$\rho = 12$	$\rho + r_{Max} = 13$	1.83 1.23	0.04 0.26	$CPI \longrightarrow WPU08$
$Z_t = [\log(CPI_t), \log(WPU09_t)]^T$	H_0: CPI does not cause WPU09 H_0: WPU09 does not cause CPI	$\rho = 6$	$\rho + r_{Max} = 7$	2.74 3.81	0.01 0.00	$CPI \longleftrightarrow WPU09$
$Z_t = [\log(CPI_t), \log(WPU10_t)]^T$	H_0: CPI does not cause WPU10 H_0: WPU10 does not cause CPI	$\rho = 11$	$\rho + r_{Max} = 12$	1.67 2.73	0.08 0.00	$CPI \longleftarrow WPU10$
$Z_t = [\log(CPI_t), \log(WPU11_t)]^T$	H_0: CPI does not cause WPU11 H_0: WPU11 does not cause CPI	$\rho = 12$	$\rho + r_{Max} = 13$	2.62 1.26	0.00 0.24	$CPI \longrightarrow WPU11$
$Z_t = [\log(CPI_t), \log(WPU15_t)]^T$	H_0: CPI does not cause WPU15 H_0: WPU15 does not cause CPI	$\rho = 11$	$\rho + r_{Max} = 12$	3.24 1.70	0.00 0.07	$CPI \longrightarrow WPU15$
$Z_t = [\log(CPI_t), \log(PPIIDC_t)]^T$	H_0: CPI does not cause PPIIDC H_0: PPIIDC does not cause CPI	$\rho = 6$	$\rho + r_{Max} = 7$	1.36 3.49	0.23 0.00	$CPI \longleftarrow PPIIDC$

Recall that in the aggregate, the relationship between CPI and PPI is bidirectional. We find that for most disaggregated PPI series that the relationship is also bidirectional, specifically for WPU01, WPU03, WPU04, WPU07, and WPU09. We note, however, that unidirectional causality also exists. Specifically, we note that WPU10 (metals and metal products) and PPIIDC (industrial commodities) both pass-through to CPI, but CPI does not pass-through to these specific commodities.

Interestingly, we note that evidence of derived-demand pass-through exists, specifically for WPU08, WPU11, and WPU15, which implies that increases in producer prices to not flow to consumer price for these commodities, rather producers would expect higher prices for these commodities if the consumer prices increase. Finally, we note that there is no causal relationship in either direction for WPU05 (fuels and related products and power). Part of this could be due to regulatory powers that limit the pricing power of providers of energy and power, thus, increases in producer prices for energy may not flow perfectly through to consumer prices in the aggregate.

Given that aggregate PT of PPI is bidirectional, there is an argument to be made that CPI movements cause PPI and visa versa. As a result of this, PPI does not always lead CPI. However, from our disaggregated models, we note that several disaggregated series exert unidirectional causality. With more confidence than the aggregate PPI series, we can state that WPU10 and PPIIDC lead consumer prices over our full sample.

1.7.3 Time-Varying Causal Direction

As established, there is a convenient one-time structural break in CPI inflation in 1993:Q1. Given that pass-through can vary with time, it is not unreasonable to think that causal direction can, too, change with some subsample analysis. Thus, we rerun our Toda & Yamamoto (1995) causality tests, but consider the possibility that causal direction can vary before-and-after said structural break. In a sense, while we believe we have addressed the question of *which* producer prices lead consumer prices, one can think of this extension as a preliminary attempt at answering *when* do producer prices lead consumer prices? Table 19 illustrates our time-varying causal direction test results.[10]

Table 19: Time-Varying TY Causality

Model	Pre-Break	Post-Break
$Z_t = [\log(CPI_t), \log(PPI_t)]^T$	$CPI \longleftrightarrow PPI$	$CPI \longleftrightarrow PPI$
$Z_t = [\log(CPI_t), \log(WPU01_t)]^T$	$CPI \longrightarrow WPU01$	$CPI \longrightarrow WPU01$
$Z_t = [\log(CPI_t), \log(WPU03_t)]^T$	$CPI \longrightarrow WPU03$	No Causality
$Z_t = [\log(CPI_t), \log(WPU04_t)]^T$	No Causality	No Causality
$Z_t = [\log(CPI_t), \log(WPU05_t)]^T$	No Causality	$CPI \longleftarrow WPU05$
$Z_t = [\log(CPI_t), \log(WPU07_t)]^T$	$CPI \longleftrightarrow WPU07$	$CPI \longrightarrow WPU07$
$Z_t = [\log(CPI_t), \log(WPU08_t)]^T$	$CPI \longleftarrow WPU08$	No Causality
$Z_t = [\log(CPI_t), \log(WPU09_t)]^T$	$CPI \longrightarrow WPU09$	No Causality
$Z_t = [\log(CPI_t), \log(WPU10_t)]^T$	$CPI \longleftarrow WPU10$	$CPI \longleftarrow WPU10$
$Z_t = [\log(CPI_t), \log(WPU11_t)]^T$	$CPI \longrightarrow WPU11$	$CPI \longrightarrow WPU11$
$Z_t = [\log(CPI_t), \log(WPU15_t)]^T$	$CPI \longrightarrow WPU15$	$CPI \longrightarrow WPU15$
$Z_t = [\log(CPI_t), \log(PPIIDC_t)]^T$	$CPI \longleftrightarrow PPIIDC$	$CPI \longleftrightarrow PPIIDC$

Note: For brevity, we suppress p-values, F-statistics and lag lengths. These are available upon request.

We note some considerable heterogeneity in our time-varying Toda & Yamamoto (1995) causality test results. Firstly, relative to our full sample tests, there are significantly more instances where causality cannot be statistically confirmed (which is say we fail to reject the null hypothesis of one of our endogenous variables "causing" the other). We also see several instances in the cases of WPU03, WPU07, WPU08 and WPU09 where causality either fails to be confirmed or switches directions between the break periods.

We also note that PPIIDC leads CPI in our full sample model, but is bidirectional when splitting the full sample into two distinct break periods. Finally, we do note that aggregate PPI is bidirectional in both periods, consistent with its corresponding full sample model. We also see that WPU10 leads CPI in both break periods, once more consistent with its full sample counterpart.

Ultimately, 19 confirms that causal direction of PPI, particularly when disaggregated, is in and of itself, varying in nature. The question of *when do producer prices lead consumer prices* is worth exploring beyond the scope of this paper, but helps us understand the question *which producer prices lead consumer prices*, in particular by exploring how causal direction or the strength of the production view theory of producer price inflation changes with time.[11]

[10] For compactness, we display causal direction if the corresponding Toda & Yamamoto (1995) causality test is significant at a level of $\leqslant 0.05$.

[11] One technical caveat to this analysis would be that breakpoint tests are highly sensitive to the level of aggregation and the length of the data in question, thus, our results in Table 19 may be spurious. Future work would do well to more rigorously evaluate this secondary question of *when* rather than *which* producer prices lead consumer prices. See Bai & Perron (1998),

1.8 Predictive Content

Following in the form of T. E. Clark et al. (1995), we construct a series of rolling one-step-ahead forecasts from several VAR models. We motivate this section by acknowledging that while quantifying SRPT is important and heterogenous depending on whether one utilizes aggregate PPI or a given disaggregated PPI, the value in disaggregating also should manifest itself in forecasting precision of CPI inflation. If there is value to utilizing one or many disaggregate PPI series, then it is important that they also improve in the forecasting quality of CPI relative to aggregate PPI and a model that omits PPI altogether.

We stress that fundamentally, the rolling forecasts utilized herein are based on *in-sample* forecasting techniques, and therefore should not be taken as benchmarks for out-of-sample forecasting potential. In a sense, we ask ourselves if the fits of forecasting models with a disaggregate PPI series are stronger compared to some benchmark models consistent with T. E. Clark et al. (1995). With these caveats in mind, each VAR model takes on the following unrestricted general form described by equation (1.10).

$$Z_t = \alpha + \delta t + \sum_{i=1}^{4} \beta_i Z_{t-i} + \varepsilon_t \tag{1.10}$$

Where α is a constant and δt is a linear time trend. Z_t is an $n \times 1$ vector of endogenous variables that are estimated simultaneously. Z_{t-i} is a vector of lagged endogenous variables with i capturing the associated lag order. Based on the AIC criteria, all models are estimated with a maximum of four lags.

We start by estimating two baseline models. The first is a five-variable VAR that includes CPI inflation, real GDP growth, the 10-month treasury bill rate (to control for financial market conditions), and the wage growth of all non-supervisory manufacturing employees (to account for labor market conditions). The second includes the same variables as the first baseline model, but with the addition of aggregate PPI inflation. We motivate our second baseline as a result of T. E. Clark et al. (1995) which posited that if PPI leads CPI, then the forecast quality of a model with PPI should outperform a model that omits its inclusion. With these baselines established, we estimate eleven VAR models, but instead of utilizing aggregate PPI inflation, we substitute it with PPI inflation of a given disaggregated commodity.

The forecasting methodology also follows the spirit of T. E. Clark et al. (1995). We opt for utilizing rolling one-step-ahead forecasts. We then forecast over specific subsamples within our data and evaluate the MAE associated with each forecast. For this study, we select four distinct recessionary periods in US history: the 1973 oil crisis (1973:Q4-1975:Q1), the Iranian Revolution (1981:Q3-1982:Q4), the collapse of the Dot-Com bubble (2001:Q1-2001:Q4), and the Financial Crisis of 2007 (2007:Q4-2009:Q2). We also evaluate the forecast quality over a restricted subsample from 1973:Q1 through 2021:Q4. To illustrate this procedure,

Bai & Perron (2003), and Carrion-i Silvestre et al. (2009) for exposition on these caveats.

suppose we wish to forecast from 1973:Q4 through 1975:Q1, we would begin our rolling forecast using only data available up until 1973:Q3 and then construct one-step-ahead forecasts for the era of interest.

To evaluate forecast quality, the mean absolute error (MAE) criteria is then used to compare the predicted values of our forecast to the actual values in our data for that specific range of interest.[12] This exercise is repeated five times for thirteen different VAR models—two baselines and eleven using the disaggregated PPI series that have been utilized in the previous sections. The ratios of the MAE from our rolling forecasts with disaggregated indices and a model with no PPI are reported in Table 20. A ratio > 1 implies a forecast from a VAR with an augmentation from a disaggregated index is inferior to a model that does not include PPI at all. In theory, if a given PPI truly does lead CPI, then its inclusion in an elementary inflation forecasting model should at the very least outperform a model that omits it altogether.

Table 20: Rolling Forecasts: MAE Ratio Compared to No PPI

Forecast Variable: π_t^{CPI}					
Augmentation	1973:Q4–1975:Q1	1981:Q3–1982:Q4	2001:Q1–2001:Q4	2007:Q4–2009:Q2	1973:Q1–2021:Q4
WPU01	1.27	1.12	1.26	0.96	1.35
WPU03	0.84	1.05	0.88	1.12	1.09
WPU04	0.97	0.99	0.96	1.07	1.01
WPU05	1.08	0.87	0.86	1.00	1.07
WPU07	0.58	0.95	0.82	0.98	0.75
WPU08	0.93	1.08	1.09	1.06	1.05
WPU09	0.44	0.96	1.11	1.00	0.87
WPU10	1.06	1.17	0.94	1.19	1.16
WPU11	1.27	1.15	0.90	0.97	1.07
WPU15	0.78	1.00	1.04	1.01	0.90
PPIIDC	0.51	1.00	0.40	1.18	0.96

We observe some heterogeneity in our initial results. Looking across each subsample, we see that more than half of the models with disaggregated PPI inflation outperform or perform as well as a model with no inflation from 1973:Q4-1975:Q1, 1981:Q3-1982:Q4, and 2001:Q1-2001:Q4. Looking at the Financial Crisis subsample and the longer sample covering the Productivity Slowdown through the present, we see weaker performance across our disaggregated models. Looking across each individual model, we find that a VAR including WPU07 inflation outperforms a model with no PPI inflation across all subsamples. We note that has a similarly strong performance relative to a model without PPI with the exception of the Financial Crisis era. Finally, relative performance varies considerably depending on the subsample, however, some disaggregated models perform consistently poorer than others including models augmented with WPU01 and WPU08. With these results in mind, we turn toward comparing the relative performances of our disaggregated forecast models to a model that uses aggregate PPI inflation. Relative performance metrics are reported in Table 21.

[12]The specific formulation is $MAE = \frac{\sum_{i=1}^{n} |a_i - b_i|}{n}$, where a_i is our actual data versus b_i, which are our predicted values.

Table 21: Rolling Forecasts: MAE Ratio Compared to Aggregate PPI

Forecast Variable: π_t^{CPI}					
Augmentation	1973:Q4–1975:Q1	1981:Q3–1982:Q4	2001:Q1–2001:Q4	2007:Q4–2009:Q2	1973:Q1–2021:Q4
WPU01	1.60	1.16	2.48	0.78	1.24
WPU03	1.05	1.08	1.74	0.91	1.00
WPU04	1.22	1.02	1.90	0.87	0.93
WPU05	1.35	0.89	1.70	0.81	0.98
WPU07	0.73	0.98	1.62	0.80	0.69
WPU08	1.17	1.11	2.15	0.86	0.97
WPU09	0.56	0.99	2.18	0.81	0.80
WPU10	1.34	1.21	1.86	0.97	1.07
WPU11	1.59	1.18	1.77	0.79	0.99
WPU15	0.98	1.03	2.05	0.82	0.83
PPIIDC	0.65	1.04	0.78	0.96	0.88

Once more, we begin our analysis of relative performance by looking across time. We observe that in many cases, it is difficult to outperform a model with aggregate inflation in earlier subsamples, however, over the Financial Crisis and the restricted full sample, we see strong relative performance of models with disaggregated series. Over the Financial Crisis, every single model with a given disaggregated PPI outperforms a model with aggregate PPI. We also see that over the restricted full sample, all but three disaggregate models similarly outperform a baseline model with aggregate PPI inflation. We observe that disaggregate models struggle most to outperform a model with aggregate PPI during the Dot-Com bubble. The model with PPIIDC, however, does outperform a model with aggregate PPI. Looking across each model, we see that once more that models with WPU07, WPU09 and PPIIDC perform considerably stronger than other models with a different augmentation. We also note that WPU01 and WPU03 perform poorly across each subsample relative to other augmentations.

Overall, when evaluating our models with a given disaggregate PPI series, we observe that WPU07, WPU09, WPU15, and PPIIDC consistently perform as-well-as or outperform at least one baseline model. WPU07 and WPU09 consistently match or outperform both baselines, but interestingly, both inadequately forecast the Dot-Com bubble era. PPIIDC is the only disaggregate that outperforms both baselines specifically during the Dot-Com bubble era.

Looking across subsamples, we see that our disaggregated series are consistently better at forecasting inflation during the Financial Crisis than models without PPI or with aggregate PPI. The disaggregated series models also do incredibly well at forecasting over the entirety of our restricted full sample, and the 1973 oil crisis. Our disaggregated series seem to struggle the most relative to our baselines in forecasting inflation during the Iranian Revolution and dot-com bubble. Overall, however, there are many instances where a given disaggregated PPI offers strict improvements in forecasting quality over aggregate PPI across our short-term recessionary periods of interest and the restricted full sample.

1.9 Discussion on the Economic Interpretation of Pass-Through

Abstracting from our disaggregated results, and focusing on our aggregate SRPT coefficient of 0.07, we stress that the economic interpretation and mechanisms driving aggregate SRPT are not clear within our reduced form setting, nor are they particularly clear in the literature, which is heavily reliant on reduced form econometric methodologies. In terms of qualifying what transmission process could propagate SRPT, consider a variation of the Berndt & Wood (1975) KLEM model of production described by equation (1.11).

$$Y_t^f = f(K_t, L_t, E_t, M_t) \tag{1.11}$$

Where Y_t^f is the output of final goods and services in the economy, K_t is the capital stock, L_t is the labor input to production, E_t is energy, and M_t are intermediate material inputs to production. Assume that there is a stage of production wherein intermediate materials, M_t are produced using capital, labor, and energy as inputs:

$$M_t = g(K_t, L_t, E_t) \tag{1.12}$$

Under the assumption that input materials, M_t, are procured at some price P_t^P, and given that all output in the economy is simply the product of the price level and final goods and services output, $\Pi_t = P_t^C \times Y_t^f$, we can begin to digest wherein producer price pass-through can enter into play. While not modeled explicitly in our simple framework, markups from intermediate goods producers in imperfectly competitive markets can enable some pass-through of P_t^P to P_t^C–in reality, this is probably the most appealing source of pass-through captured via our results. Beyond markups, there is potential for the costs associated with both labor and capital, w_t and r_t, respectively, at all levels of production to increase and pass-through to consumer prices.

If there is heterogeneity in the labor and capital components between intermediate and final goods producers (rather than treating them as homogenous between stages of fabrication as is the case described above), then it is possible as well that asymmetric shocks to wages and the rental rate of capital in intermediate goods markets can pass-through along the production chain to consumer prices through no fault of the final goods producer. Clearly, shocks to energy, E_t, at the intermediate goods producer level (such as oil shocks) can pass-through to P_t^P as well. This pass-through is covered at length in the oil price pass-through literature; however, if one cannot disentangle asymmetries in the responses of intermediate and final goods producers to oil price shocks, then it is possible that our SRPT rate of 7% may be confounded by oil price pass-through as well.[13]

[13] On a related note, shocks to productivity and utilization rates of inputs to production can also push P_t^P pass-through higher, although identifying productivity shocks is difficult, and often are identified as oil price shocks in reality.

Adding in complexity, if we allow for final goods producers to also obtain input materials to production from international trade partners, then M_t could have two corresponding prices: a domestic producer price P_t^P, as well a price from the international market for the same M_t good at some exchange rate. Adding this complexity means that observed shocks in P_t^P can be, in some cases, offset by favorable exchange rates and access to foreign markets for M_t; much like E_t, this extrapolation of where M_t can be procured brushes up against the exchange rate pass-through literature, further confounding our reduced form interpretation of pass-through.

A final–although less clear–source of pass-through can arise from supply chain shocks or global supply chain pressures. While supply chain pressure is not an input to production, it likely is an important environmental factor that directly affects production markets at both the intermediate and final goods level. Consider Figure 10 below, which provides graphical evidence of the correspondence between producer price inflation, and global supply chain pressures as measured by the New York Fed Global Supply Chain Disruption Index.

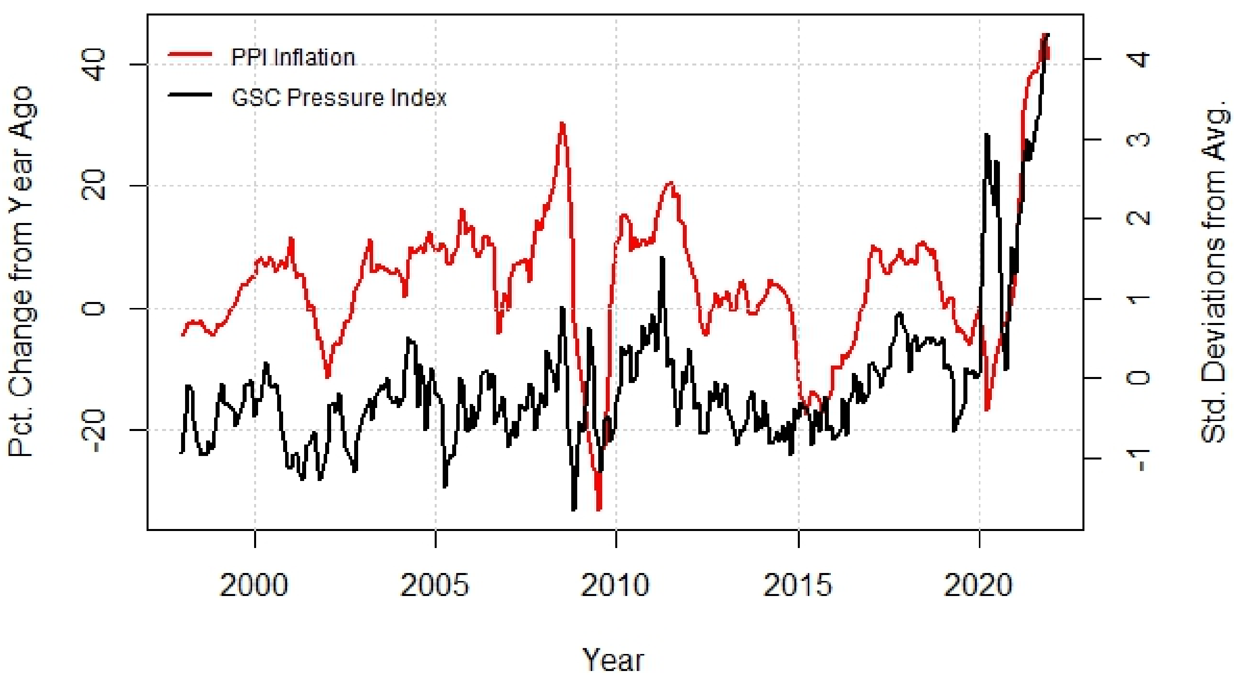


Figure 10: Supply Chain Pressures & Producer Price Inflation

The comovement between both producer price inflation and supply chain disruptions is quite compelling, particularly over the past decade of our sample. Further research would do well to map supply chain pressures to the simple framework described herein.

Overall, the complexity of the stage-of-fabrication nature of most final goods production along with the confounding roles that both exchange rates, and energy prices have in general equilibrium necessarily muddle the mixtures of mechanisms that *could* be driving PPI inflation, and subsequent pass-through of PPI to CPI. A more rigourous study would do well to incorporate producer price, oil price, and exchange rates into a unifying structural model wherein pass-through can be both better described, and its transmission mechanisms more clearly identified than what reduced form approaches alone are capable of. The historical treatment of the pass-through literatures as separate from one another motivates a more compelling approach–structural and reduced form–that can capture pass-through across the production chain.

1.10 Conclusion

The goal of this analysis was to first and foremost quantify both the short-run and long-run pass-through of producer prices to consumer price level. Beyond considering the SRPT of aggregate PPI to CPI, we consider the possibility that the aggregate PPI series can mute the true underlying commodities that drive stronger rates of pass-through than what the aggregate can capture alone. To directly quantify PT in the aggregate, we leverage a two-step error correction framework. The first step is a long-run spurious regression of the CPI level against real GDP and the aggregate PPI level. The second step is to estimate a short-run augmented Phillips Curve leveraging the error-correction term from the long-run model. To this end, we estimate a SRPT of aggregate PPI inflation to the CPI level of around 7 percent.

With an aggregate benchmark quantified, we turn toward a mix of disaggregated commodity-specific PPI series and evaluate their SRPT coefficients relative to an estimated Phillips curve model with aggregate PPI. We find significant heterogeneity in SRPT. We observe that in particular, SRPT from WPU09, WPU11, and WPU15 are significantly higher compared to the aggregate PPI series at roughly 17, 20, and 11 percent each, respectively. We estimate a one-time structural break in 1993:Q1, and find that pre-break SRPT is significantly higher than post-break SRPT for aggregate PPI and most disaggregate PPI series with the only exception being WPU03 (textile products and apparel). One could speculate for this specific commodity that the heightened demand for masks and other PPE items during the Covid-19 pandemic is the driving factor behind this asymmetric result.

Leveraging the cointegrated nature of our time series of interests, we construct a series of bivariate vector models (VECM and VARX) to test for causality and identify causal direction. We observe via Toda-Yamamoto causality that bidirectional causality exists between aggregate PPI and CPI; however, the causal nature of CPI and a given disaggregate PPI series is quite heterogeneous. WPU01, WPU03, WPU04, WPU07, and WPU09 all exhibit a bidirectional relationship with the price level while WPU10 and PPIIDC unidirectionally cause changes to the price level. Interestingly, CPI unidirectionally causes changes to the levels of WPU08, WPU11, and WPU15. An outlier in this analysis is WPU05, which has no causal relationship with the CPI level.

Finally, we test the forecast quality of CPI in a series of VAR models with controls for both financial conditions and labor market conditions. Results indicate that all the disaggregate PPI series perform as well or better than a baseline model without PPI inflation or with only aggregate PPI inflation during the Financial Crisis era. On average, models with disaggregate PPI series outperform the benchmark models over the course of the full sample. WPU07 and WPU09 in particular outperform both benchmark models over the full sample and each subsample recession with the exception of the Dot-Com bubble.

Overall, with just a handful of long disaggregate PPI series, we confirm our hypothesis stemming from our motivation that the aggregate PPI series while valuable does mute and obscure underlying commodities that have stronger causal ties with CPI and more efficient predictive capabilities. If policymakers are keen on attaining price stability and adhere to the production chain view of producer price pass-through, then utilizing aggregate PPI as a leading indicator for CPI is not necessarily the most efficient choice. The results of this paper advocate for the consideration of a few choice disaggregate indices. WPU07, WPU09, and PPIIDC have more potency than aggregate PPI on average across our results, and thus warrant further investigative studies and utilization as an alternative to the aggregate PPI series.

A final note is that our disaggregate series chosen for this study were determined largely on the length of the series themselves. In reality, there are thousands of commodity PPI series and industry PPI series as well as PPI series constructed based on final demand. The shortcoming of many of these series is that there lengths seldom go back earlier than 1980, thus the use of extensive disaggregate PPI series depends in part on the questions and hypotheses of a given researcher.

Finally, we only evaluate CPI inflation using the traditional CPI level; however, one could imagine a study tertiary to this one that looks at core goods inflation (CPI inflation less food and energy). The predictive content and usefulness of disaggregate PPI series could differ depending on the measure of inflation utilized, thus, a natural extension of this study would be a replication utilizing core inflation as the outcome variable of interest.

1.11 Appendix A: Auxiliary Results

We consider some auxiliary results. We consider two specific modifications to our baseline augmented Phillips Curve model, and present key evidence of relevant pass-through changes with such modifications. We consider three specific deviations from our original analysis: first, we restrict our sample to only go as far out as 2019:Q4, thus reducing bias that could arise from the Covid-19 pandemic, which presented otherwise unprecedented challenges to the supply-side of the economy and shocks in the demand-side of the economy; secondly, we consider a reestimation of our Phillips Curve models using the CBO estimated output gap, rather than the calculated output gap we use and compare results.

1.11.1 Omitting Covid-19

A reasonable critique to our benchmark model is that the most recent subset of our sample includes the Covid-19 recession and several subsequent events that have notably distorted global value chains and other key intermediaries in the production-side of the economy. While this portion of our sample is small, the potential for extreme values in both CPI inflation and PPI inflation might upwardly bias our SRPT coefficient estimates. Thus, we proceed to drop observations after 2019:Q4 and reestimate our baseline augmented Phillips Curve in both the aggregate PPI and for all disaggregate indices.

Aggregate

We present aggregate SRPT without Covid-19 years in Table 22.

Table 22: Short-Run Model: Aggregate (Without Covid-19)

Dependent Variable: π_t^{CPI}		
Variable	Estimate	Standard Error
α	0.003**	(0.001)
π_{t-1}^{CPI}	1.108***	(0.093)
π_{t-2}^{CPI}	−0.087	(0.128)
π_{t-3}^{CPI}	0.134	(0.138)
π_{t-4}^{CPI}	−0.221**	(0.085)
$(y_{t-1} - \bar{y}_{t-1})$	0.063.	(0.036)
π_{t-1}^{PPI}	0.085*	(0.042)
π_{t-2}^{PPI}	−0.135***	(0.039)
π_{t-3}^{PPI}	−0.009	(0.050)
π_{t-4}^{PPI}	0.030	(0.024)
ECT_{t-1}	−0.019*	(0.009)
SRPT	$\theta_1 + (\psi \times \beta_2) = 0.07$	(0.042)
Adjusted R^2	0.94	
Observations	284	

Note: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; $\cdot p < 0.10$

We observe that the SRPT without Covid-19 in our sample rounds up to 0.07, which is matches our original aggregate SRPT coefficient with Covid-19 in the sample. We can conclude that Covid-19's inclusion in the sample does not have a strong effect in distorting the aggregate SRPT.

Disaggregate

We now consider cutting off our sample size for our disaggregate models. The results are tabulated in Table 23.

Table 23: Short-Run Models: Disaggregate (Without Covid-19)

	Dependent Variable: π_t^{CPI}										
Variable/Augmentation	WPU01	WPU03	WPU04	WPU05	WPU07	WPU08	WPU09	WPU10	WPU11	WPU15	PPIIDC
α_j	0.003**	0.004***	0.004**	0.003*	0.002**	0.003***	0.003**	0.003***	0.004***	0.003***	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
π_{t-1}^{CPI}	1.258***	1.209***	1.237***	1.281***	1.257***	1.232***	1.215***	1.203***	1.184***	1.230***	1.156***
	(0.069)	(0.076)	(0.072)	(0.083)	(0.065)	(0.063)	(0.095)	(0.080)	(0.080)	(0.069)	(0.097)
π_{t-2}^{CPI}	−0.349***	−0.394***	−0.331***	−0.307**	−0.398***	−0.321***	−0.325**	−0.202.	−0.276*	−0.337***	−0.128
	(0.074)	(0.077)	(0.078)	(0.111)	(0.093)	(0.089)	(0.116)	(0.113)	(0.109)	(0.074)	(0.143)
π_{t-3}^{CPI}	0.116	0.216*	0.067	0.095	0.176.	0.069	0.103	−0.005	0.082	0.061	0.127
	(0.084)	(0.097)	(0.089)	(0.132)	(0.097)	(0.093)	(0.088)	(0.113)	(0.105)	(0.077)	(0.135)
π_{t-4}^{CPI}	−0.112*	−0.157*	−0.082	−0.129.	−0.119	−0.091	−0.123	−0.094	−0.143*	−0.144**	−0.220**
	(0.053)	(0.079)	(0.052)	(0.076)	(0.072)	(0.056)	(0.075)	(0.066)	(0.072)	(0.055)	(0.078)
$(y_{t-1} - \bar{y}_{t-1})$	0.059.	0.039	0.036	0.053	0.065*	0.020	0.069*	0.058*	0.054.	0.063*	0.051
	(0.031)	(0.032)	(0.030)	(0.035)	(0.030)	(0.030)	(0.032)	(0.029)	(0.029)	(0.032)	(0.039)
π_{jt-1}^{PPI}	0.008	0.078.	0.038**	−0.001	0.057*	0.024.	0.111	0.084***	0.205.	0.094**	0.072.
	(0.009)	(0.040)	(0.013)	(0.007)	(0.025)	(0.012)	(0.073)	(0.025)	(0.107)	(0.035)	(0.044)
π_{jt-2}^{PPI}	0.001	0.042	−0.064	−0.009	0.015	−0.009	−0.105	−0.176***	−0.289	−0.085	−0.13**
	(0.012)	(0.096)	(0.021)	(0.011)	(0.054)	(0.022)	(0.135)	(0.042)	(0.184)	(0.053)	(0.046)
π_{jt-3}^{PPI}	−0.015	−0.228.	0.036	−0.004	−0.133*	−0.017	−0.009	0.111**	0.064	0.021	0.003
	(0.013)	(0.132)	(0.024)	(0.010)	(0.057)	(0.020)	(0.095)	(0.039)	(0.159)	(0.051)	(0.046)
π_{jt-4}^{PPI}	0.001	0.143*	−0.006	0.000	0.025	0.011	0.042	0.021	0.079	0.033	0.023
	(0.008)	(0.066)	(0.163)	(0.007)	(0.025)	(0.011)	(0.042)	(0.021)	(0.079)	(0.033)	(0.023)
ECT_{jt-1}	−0.011*	−0.035**	−0.026*	−0.011*	−0.001	−0.026**	0.032.	−0.012.	−0.015**	0.030*	−0.016.
	(0.004)	(0.012)	(0.010)	(0.005)	(0.006)	(0.008)	(0.017)	(0.006)	(0.006)	(0.014)	(0.009)
SRPT	0.002	0.046	0.008	−0.010	0.057	0.008	0.140	0.075	0.193	0.116	0.063
	(0.009)	(0.040)	(0.013)	(0.007)	(0.025)	(0.012)	(0.073)	(0.025)	(0.107)	(0.035)	(0.044)
Adjusted R^2	0.94	0.94	0.94	0.93	0.94	0.94	0.94	0.94	0.94	0.94	0.94
Observations	284	284	284	284	284	284	284	284	284	284	284

Note: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; $\cdot p < 0.10$

We pay special attention to our SRPT coefficients, which all bear similar magnitudes to our baseline model (though, as expected, there is a trivially small upward bias of Covid-19 when comparing to our baseline). What is interesting, however, is that the output gap becomes statistically significant in several of the reestimated models. Clearly, Covid-19 had a strong dampening effect on the output gap's predictive content in our baseline model. While this statistical significance has no practical effect on the SRPT estimates, this result does provide an additional layer of inference with regards to Covid-19's effect on the components of an estimated augmented Phillips Curve.

1.11.2 CBO Output Gap

We consider the possibility that deviating output from its HP-filtered trend may not be an adequate approximation of the output gap. We also acknowledge that while the HP filter works as an adequate means to detrend series, it is not without its faults, many of which are well-described in Hamilton (2018). An alternative to consider is the output gap as estimated from the Congressional Budget Office (CBO), which is calculated as $(GDP_t - PGDP_t)/PGDP_t$ where real GDP is the same variable as utilized earlier, however, real potential GDP ($PGDP_t$) results from CBO estimates. Our derived output gap compares with the CBO output gap graphically in Figure 11.

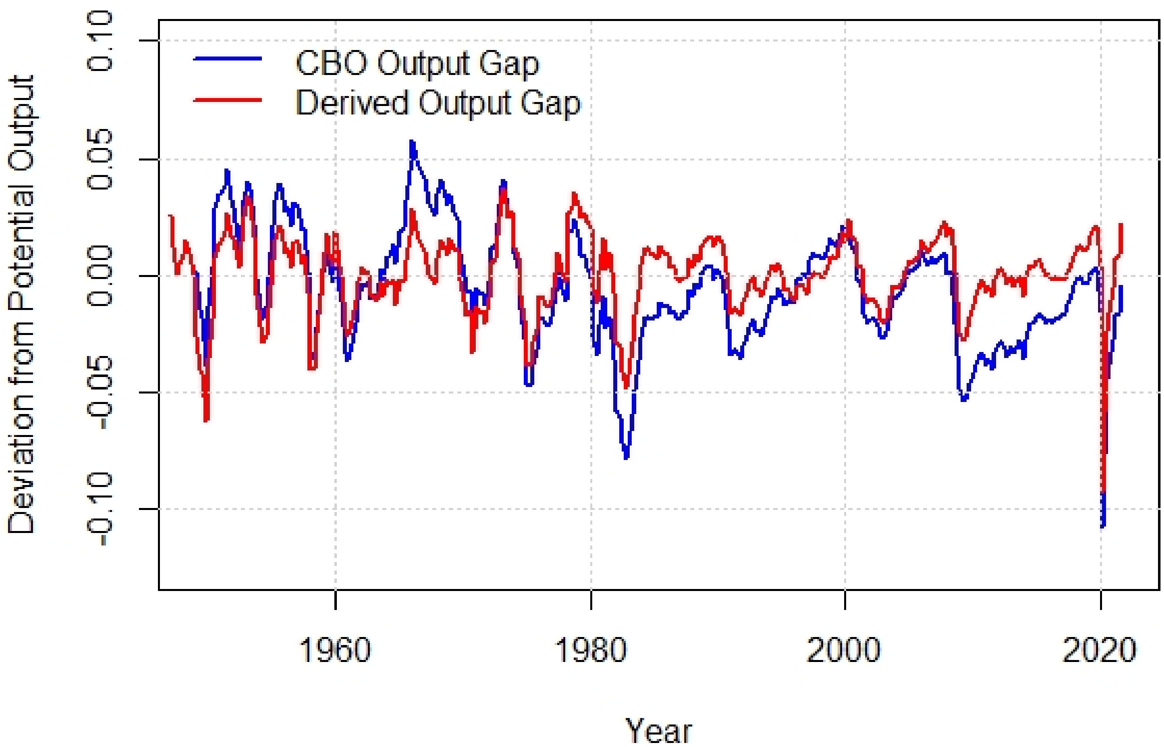


Figure 11: Comparison of Output Gap Estimates

In terms of the relative relationship to each series, there is a positive correlation of about 75 percent. We can observe that the standard HP-filtered derived output gap (the standard in the literature) tends to reflect a smaller relative gap than the CBO estimate. Because of this, and despite a high correlation, our output gap could be seen as overly optimistic. We present a new set of aggregate and disaggregate results with a change-of-variable for the output gap. Specifically, we reestimate our Phillips Curve models with the CBO gap rather than our own derived gap.

Aggregate

We estimate a new aggregate short-run Phillips Curve model with the CBO output gap and derive new pass-through coefficients from both short-run estimates and long-run estimates (estimated with CBO output gap). Reported in Table 24 are the Phillips Curve model results and subsequent SRPT coefficient.

Table 24: Short-Run Model: Aggregate (CBO Output Gap)

Dependent Variable: π_t^{CPI}		
Variable	Estimate	Standard Error
α	0.003***	(0.001)
π_{t-1}^{CPI}	1.027***	(0.097)
π_{t-2}^{CPI}	−0.029	(0.133)
π_{t-3}^{CPI}	0.152	(0.136)
π_{t-4}^{CPI}	−0.214*	(0.086)
$(y_{t-1} - \bar{y}_{t-1})$	0.035.	(0.019)
π_{t-1}^{PPI}	0.094**	(0.036)
π_{t-2}^{PPI}	−0.119**	(0.042)
π_{t-3}^{PPI}	−0.019	(0.046)
π_{t-4}^{PPI}	0.020	(0.026)
ECT_{t-1}	−0.021*	(0.008)
SRPT	$\theta_1 + (\psi \times \beta_2) = 0.07$	(0.036)
Adjusted R^2	0.93	
Observations	292	

Note: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; $\cdot p < 0.10$

We note that our SRPT coefficient is precisely 0.08, which is comparable to our original estimate of 0.07. The output gap definition in this model positively effects SRPT relative to our original model, albeit by at marginal level.

Disaggregate

We finally consider the change-of-variable for our output gap measurements within the context of our disaggregated short-run models. SRPT coefficients are derived from the short–run model estimates and the corresponding long-run models (estimated with the CBO gap). The results for the short-run model and subsequent SRPT coefficients are presented in Table 25.

Table 25: Short-Run Models: Disaggregate (CBO Output Gap)

	Dependent Variable: π_t^{CPI}										
Variable/Augmentation	WPU01	WPU03	WPU04	WPU05	WPU07	WPU08	WPU09	WPU10	WPU11	WPU15	PPIIDC
α_j	0.003***	0.004***	0.003***	0.002*	0.002**	0.003***	0.003**	0.003***	0.003***	0.003***	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
π_{t-1}^{CPI}	1.223***	1.129***	1.203***	1.247***	1.217***	1.180***	1.140***	1.137***	1.147***	1.201***	1.079***
	(0.071)	(0.082)	(0.066)	(0.091)	(0.061)	(0.061)	(0.088)	(0.076)	(0.078)	(0.065)	(0.095)
π_{t-2}^{CPI}	−0.311***	−0.300**	−0.282***	−0.255*	−0.335***	−0.268**	−0.259*	−0.133	−0.208*	−0.279***	−0.066
	(0.082)	(0.093)	(0.078)	(0.115)	(0.084)	(0.085)	(0.106)	(0.108)	(0.100)	(0.077)	(0.138)
π_{t-3}^{CPI}	0.130	0.213*	0.079	0.088	0.168.	0.082	0.106	0.019	0.100	0.046	0.151
	(0.082)	(0.101)	(0.077)	(0.141)	(0.098)	(0.087)	(0.085)	(0.112)	(0.083)	(0.081)	(0.136)
π_{t-4}^{CPI}	−0.134*	−0.173*	−0.100.	−0.137	−0.126.	−0.120*	−0.123	−0.120.	−0.182*	−0.139*	−0.221**
	(0.059)	(0.086)	(0.060)	(0.086)	(0.075)	(0.055)	(0.077)	(0.066)	(0.070)	(0.059)	(0.082)
$(y_{t-1} - \bar{y}_{t-1})$	0.015	0.014	0.013	0.025	0.028	−0.020	0.035.	0.021	−0.003	0.029	0.030
	(0.021)	(0.018)	(0.021)	(0.020)	(0.021)	(0.023)	(0.019)	(0.020)	(0.021)	(0.020)	(0.020)
π_{jt-1}^{PPI}	0.007	0.113*	0.036*	−0.001	0.051*	0.012	0.144*	0.090***	0.272**	0.102**	0.079*
	(0.009)	(0.052)	(0.015)	(0.007)	(0.023)	(0.014)	(0.069)	(0.021)	(0.099)	(0.033)	(0.034)
π_{jt-2}^{PPI}	0.007	−0.032	−0.029	−0.007	0.043	0.008	−0.142	−0.156***	−0.377*	−0.089.	−0.113*
	(0.011)	(0.119)	(0.029)	(0.011)	(0.052)	(0.018)	(0.126)	(0.042)	(0.157)	(0.050)	(0.047)
π_{jt-3}^{PPI}	−0.017	−0.153	−0.004	−0.003	−0.166**	−0.011	0.008	0.081.	0.077	0.011	−0.016
	(0.012)	(0.149)	(0.028)	(0.010)	(0.057)	(0.019)	(0.091)	(0.041)	(0.123)	(0.049)	(0.046)
π_{jt-4}^{PPI}	0.002	0.109	0.016	−0.005	0.088**	0.002	0.036	−0.009	0.081	0.064*	0.024
	(0.008)	(0.074)	(0.015)	(0.008)	(0.028)	(0.013)	(0.043)	(0.021)	(0.064)	(0.027)	(0.025)
ECT_{jt-1}	−0.010*	−0.038**	−0.011	−0.010*	−0.001	−0.028**	0.031.	−0.012.	−0.013*	0.018	−0.016.
	(0.005)	(0.013)	(0.010)	(0.005)	(0.005)	(0.011)	(0.017)	(0.006)	(0.006)	(0.013)	(0.009)
SRPT	0.001	0.079	0.023	−0.009	0.051	−0.015	0.172	0.081	0.261	0.114	0.069
	(0.009)	(0.052)	(0.015)	(0.007)	(0.023)	(0.014)	(0.069)	(0.021)	(0.099)	(0.033)	(0.034)
Adjusted R^2	0.93	0.94	0.93	0.93	0.94	0.94	0.93	0.99	0.93	0.93	0.93
Observations	292	292	292	292	292	292	292	292	292	292	292

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.10$

We observe results that fairly consistent with our baseline model. Deviations in the SRPT coefficients tend to be higher than our baseline in this model, which is sensible given the deviation between our output gap and the CBO gap. Overall, however, our main results still hold for each subsequent disaggregated PPI.

1.12 Appendix B: Augmented Quantity Equation

While we argue that the production view theory of producer prices leading consumer prices is a plausible long-run relationship if cointegration exists between CPI and PPI, there may be those who are skeptical of the long-run dynamics of the price level outside of the quantity theory of money. We use this space to provide a similar pass-through application of producer-to-consumer prices using an augmented quantity equation, rather than an augmented Phillips curve.[14]

Equation (1.13) describes the quantity equation in its textbook representation.

$$M_t V_t = P_t^C Y_t \tag{1.13}$$

In logs, we can rewrite equation (1.13) in logs as equation (1.14).

$$\log(M_t) + \log(V_t) = \log(P_t^C) + \log(Y_t) \tag{1.14}$$

[14] For the purpose of this exercise, we look to compare SRPT between these functional forms using only the aggregate PPI series.

We can rearrange equation (1.14) with the price level on the lefthandside and augmented with producer prices to form a regression equation described by (1.15). This is our long-run equation.

$$\log(P_t^C) = \beta_0 + \beta_1 \log(M_t) + \beta_2 \log(Y_t) + \beta_3 \log(V_t) + \beta_4 \log(P_t^P) + \epsilon_t \tag{1.15}$$

We can convert our core variables to growth rates relative to a year ago to capture our short-run relationships necessary to operationalize our error-correction framework. Equation (1.16) describes our short-run augmented quantity equation.

$$\pi_t^{CPI} = \alpha_0 + \alpha_1 \Delta M_{t-1} + \alpha_2 \Delta Y_{t-1} + \alpha_3 \Delta V_{t-1} + \alpha_4 \pi_t^{PPI} + \psi ECT_{t-1} + \varepsilon_t \tag{1.16}$$

Equations (1.15) and (1.16) taken together described our augmented quantity equation as an error correction model. We report the estimates for equation (1.15) in Table 26.

Table 26: Equation (1.15) Estimation Results

Dependent Variable: $log(P_t^C)$		
Variable	Estimate	Std. Error
β_0	-2.57***	(0.16)
$\log(M_t)$	-1.20***	(0.13)
$\log(Y_t)$	1.55***	(0.12)
$\log(V_t)$	-1.01***	(0.14)
$\log(P_t^P)$	0.69***	(0.02)
Adjusted R^2	0.99	
Observations	255	

Note: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; $\cdot p < 0.10$

We take note that our long-run coefficient in this model, $\beta_4 = 0.69$. With this in mind, we turn towards the estimation of equation (1.16). Table 27 reports these results.

Table 27: Equation (1.16) Estimation Results

Dependent Variable: π_t^{CPI}		
Variable	Estimate	Std. Error
α_0	0.03***	(0.00)
ΔM_{t-1}	-0.60***	(0.15)
ΔY_{t-1}	0.42*	(0.17)
ΔV_{t-1}	-0.50**	(0.16)
π_t^{PPI}	0.32***	(0.03)
ECT_{t-1}	0.05	(0.03)
Adjusted R^2	0.64	
Observations	251	

Note: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; $\cdot p < 0.10$

With these results in mind, we can estimate SRPT of producer prices in a similar vein as our augmented

Phillips Curve: $\alpha_4 + (\psi \times \beta_4) = 0.35$. When we estimate our augmented Phillips Curve over the same sample period, we arrive at a SRPT rate of roughly 0.08, which is considerably smaller than what the augmented quantity equation would predict.[15] One may ask, which is more preferable? Given that our dependent variable is not changing between either model, we can evaluate each model's fit via its adjusted R^2 and F-statistic.[16]

Looking at our long-run models, both possess an R^2 of around 0.99, which is unsurprising given how well the quantity equation is known to capture the long-run dynamics of money and prices. The augmented quantity equation in levels has an F-statistic of 17280 while our augmented Phillips Curve possesses an F-statistic of 19050. Ultimately, our long-run models produce comparable fits.

When evaluating our short-run models and their specifications, we find that our augmented Phillips Curve possesses an R^2 of around 0.95 and a corresponding F-statistic of roughly 407.8. On the other hand, our augmented quantity equation in the short-run posses an R^2 of roughly 0.64 and F-statistic of 88.28. Clearly, the goodness-of-fit from our augmented Phillips Curve in the aggregate is, on the surface level, better fit for use in an error-correction framework than an analogous augmented quantity equation. This finding reaffirms the use of the Phillips Curve as a functional form in the pass-through literature and gives some indication that our SRPT estimates from our main results are more reliable than similar SRPT estimates derived from an alternative functional form like the quantity equation.

None of this is to say that the quantity equation is without application. It has proven to be a sound and undeniable theory for capturing the long-run dynamics of the price level. Beyond this study, there is potential for deeper empirical treatments that can leverage the quantity equation to describe pass-through phenomena.

1.13 Appendix C: Impulse-Response Functions & Forecast Error Variance Decomposition

While not essential to our analysis, nor our main results, one can also measure pass-through via impulse-response functions and through a forecast error variance decomposition (FEVD) exercise. Mean responses from a PPI shock to CPI can tell us how strong a PPI price shock can pass through to consumer prices. An FEVD can tell provide us insights with regards to the amount of variance in h period ahead CPI inflation forecasts that can be explained by a given producer price index. In essence, the FEVD can shed light on

[15] Explicit results from our comparable augmented Phillips Curve are available upon request. For brevity, we discuss our key results. Furthermore, we estimate our augmented Phillips Curve with one lag for the righthandside variables rather than 4 to ensure the sample length for each model are identical to one another.

[16] There are, of course, more rigorous ways to compare goodness-of-fit, but so as to not go beyond the scope of this supplemental section, we stick with simple regression diagnostics for naive, but straightforward, comparisons.

how long PPI movements can explain CPI movements when forecasting outward.

1.13.1 Aggregate Model

Impulse-Response Functions

Using the features of a bivariate VEC model, we generate a series of impulse response functions, and graph their projections for $h = 8$ periods ahead.[17] The full sample is considered when constructing these models and generating said IRF's. In order to generate these results, our VEC model is converted functionally into a VAR model in levels. Thus, our projections are also in levels. Standard errors are graphically reported at a 90% confidence interval with dashed red lines denoting the upper and lower bounds, respectively. Figure 12 presents these results.

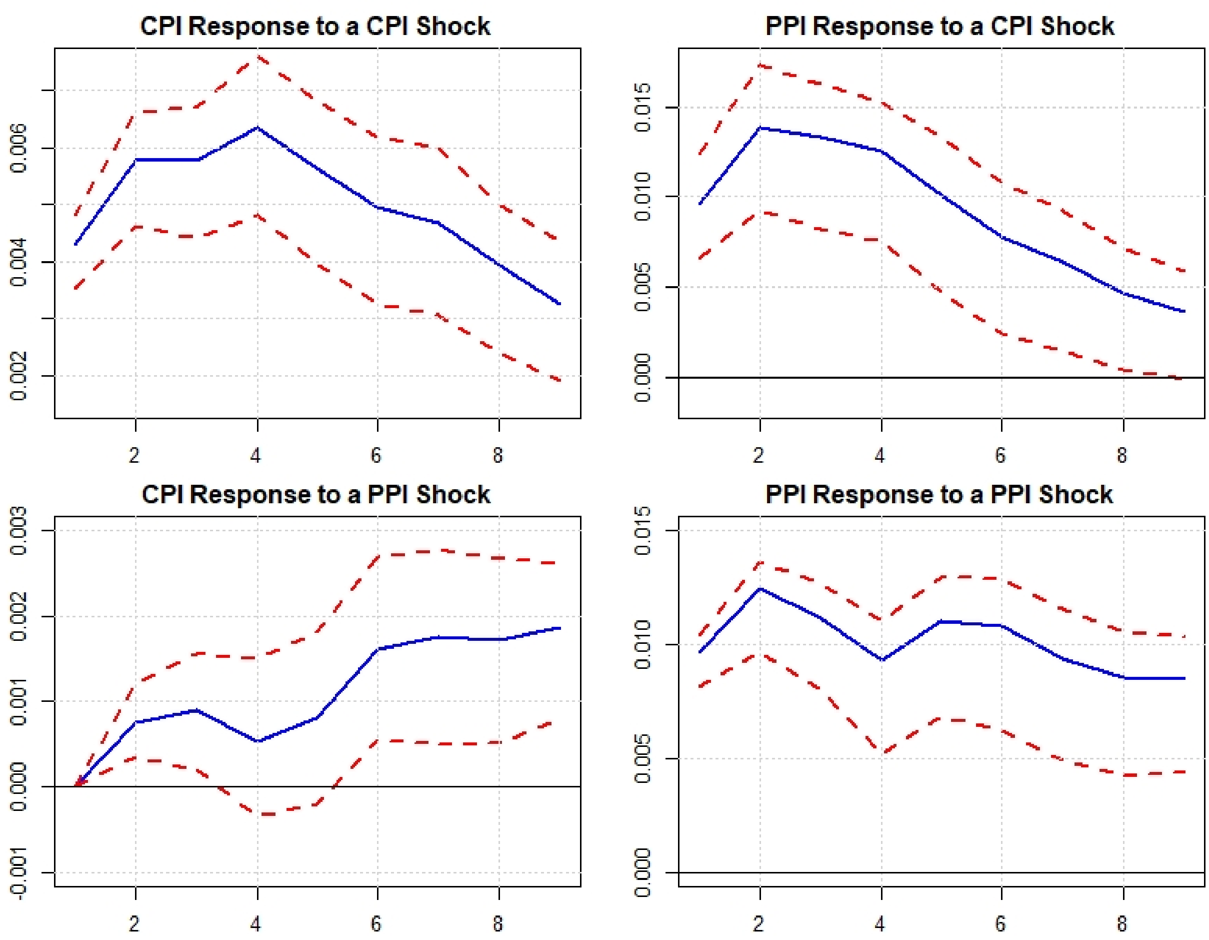


Figure 12: Impulse Response Functions: Aggregate

We observe that in all cases, shocks from both CPI and PPI produce positive and significant responses to one another. This supports the evidence of bidirectional causality provided by our TY causality test. It

[17]Our bivariate VEC model consists of a vector containing both consumer and producer prices in their levels.

should be noted, however, that a CPI shock appears to be more transitory in nature. For both CPI and PPI, a shock to CPI has a pronounced positive responses that tapers off over the course of the projection. Interestingly, the magnitude of the responses seems to be larger in the case of PPI responses to a CPI shock than a CPI response to the same shock. We note that in the case of a PPI shock that in both cases, CPI and PPI do not tend to taper off after the initial shock, rather the shock persists and the indices generally remain at a higher level than prior to the shock. CPI responds at a weaker magnitude than PPI to a PPI shock, and furthermore, the statistical significance of CPI's response tapers off a little before the fourth step ahead, and then begins to rise to a statistically significant level once more just before the sixth step ahead.

Variance Decomposition

One final exercise we perform on our aggregate data is the tabulation of the forecast error variance decomposition (FEVD). The FEVD allows us to explain for h periods ahead the amount of variance in projection of either CPI or PPI that can be explained by itself and by the other variable. We construct our aggregate FEVD from our VEC model in a VAR form, thus, the FEVD reported is explaining the variance in the levels, rather than growth rates. Table 28 reports the results of this exercise.

Table 28: FEVD: Aggregate

h Ahead	CPI FEVD		PPI FEVD	
	CPI	PPI	CPI	PPI
$h = 1$	1.0000000	0.000000000	0.5392128	0.4607872
$h = 2$	0.9945431	0.005456939	0.5942954	0.4057046
$h = 3$	0.9930874	0.006912556	0.6363652	0.3636348
$h = 4$	0.9953275	0.004672462	0.6815670	0.3184330
$h = 5$	0.9961430	0.003857026	0.6921565	0.3078435
$h = 6$	0.9953847	0.004615298	0.6990935	0.3009065
$h = 7$	0.9945587	0.005441270	0.7086651	0.2913349
$h = 8$	0.9935990	0.006400968	0.7146914	0.2853086

We observe from our FEVD exercise that most of the variance in the CPI level can be explained by its own dynamics. The level of CPI seldom varies with the level of PPI despite bidirectional causality between the two. For PPI on the other hand, it would seem that for early h periods ahead, CPI explains more of the variance in PPI than PPI itself for all eight periods ahead. This goes against the production chain theory of pass-through and tends to support derived-demand theory of pass-through in the aggregate for the US, rather than bidirectional causality, which is more evident in our IRFs and causality tests.

1.13.2 Disaggregate Models

Impulse-Response Functions

Our bivariate impulse response functions are generated from a series of bivariate VEC models that are converted to VARs in levels. We present how the CPI level responds to shocks from the level of each PPI series individually, and how each PPI series responds to a CPI shock.

PPI Shocks

Here, we examine the shocks each individual disaggregated PPI series elicits and how CPI responds. Thus, there are eleven graphical impulse-response functions described by Figure 13.

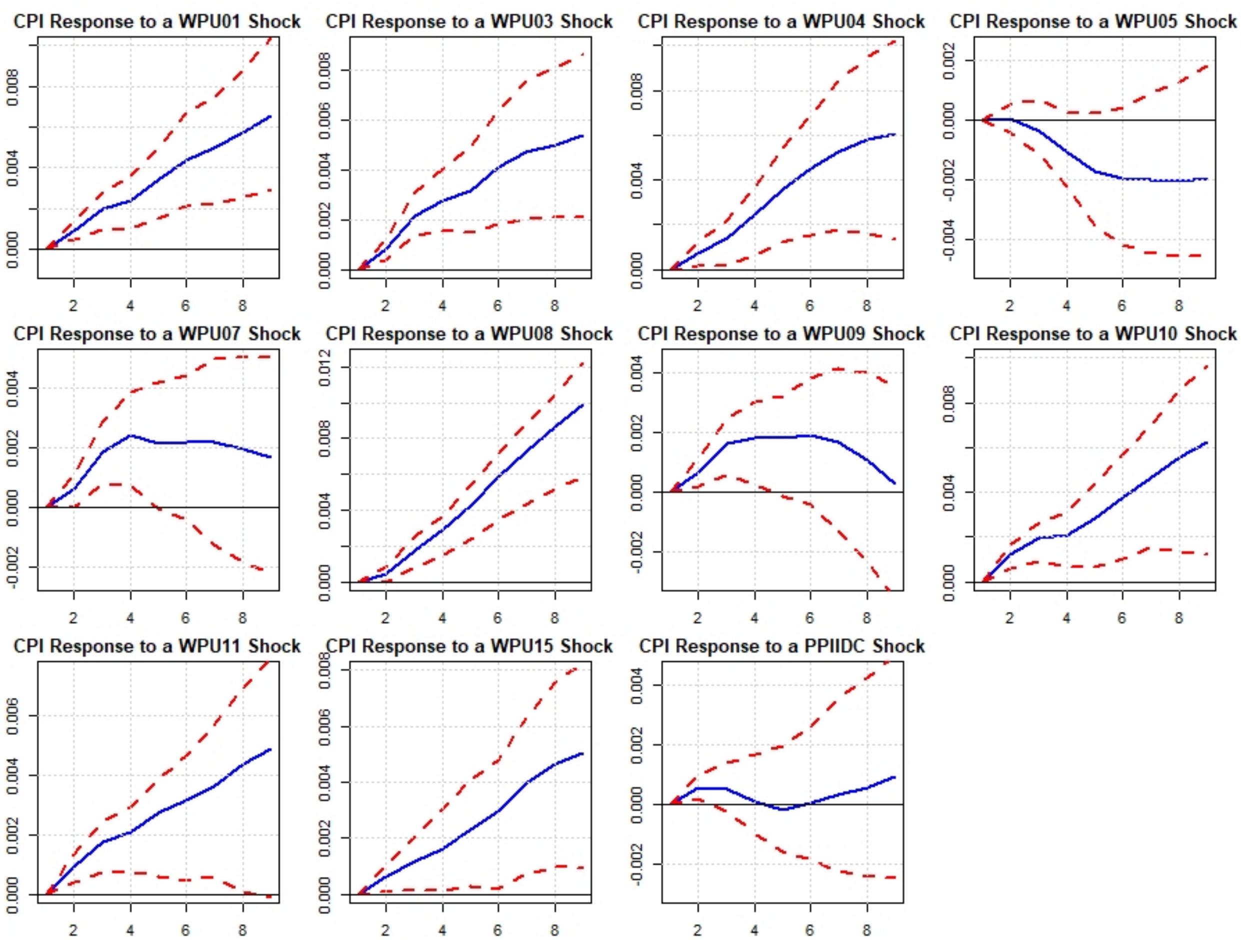


Figure 13: Impulse Response Functions: Disaggregate PPI Shocks

We observe that CPI responds positively and significantly in all cases of disaggregate PPI shocks with the exception of shocks originating from WPU05, and PPIIDC. We note that for almost all statistically significant shocks, there is a tendency for CPI's response to exhibit persistence and settle at a higher level than it was prior to the shock. The exceptions to this persistence are CPI responses in the cases of WPU05, and PPIIDC, which are not different from zero, and WPU07, WPU09, and WPU11, in which CPI exhibits a transitory, but significant response.

CPI Shocks

Here, we examine shocks elicited by CPI and how each disaggregated PPI series responds to said shock. Eleven graphical impulse response functions for each PPI series are presented in Figure 14.

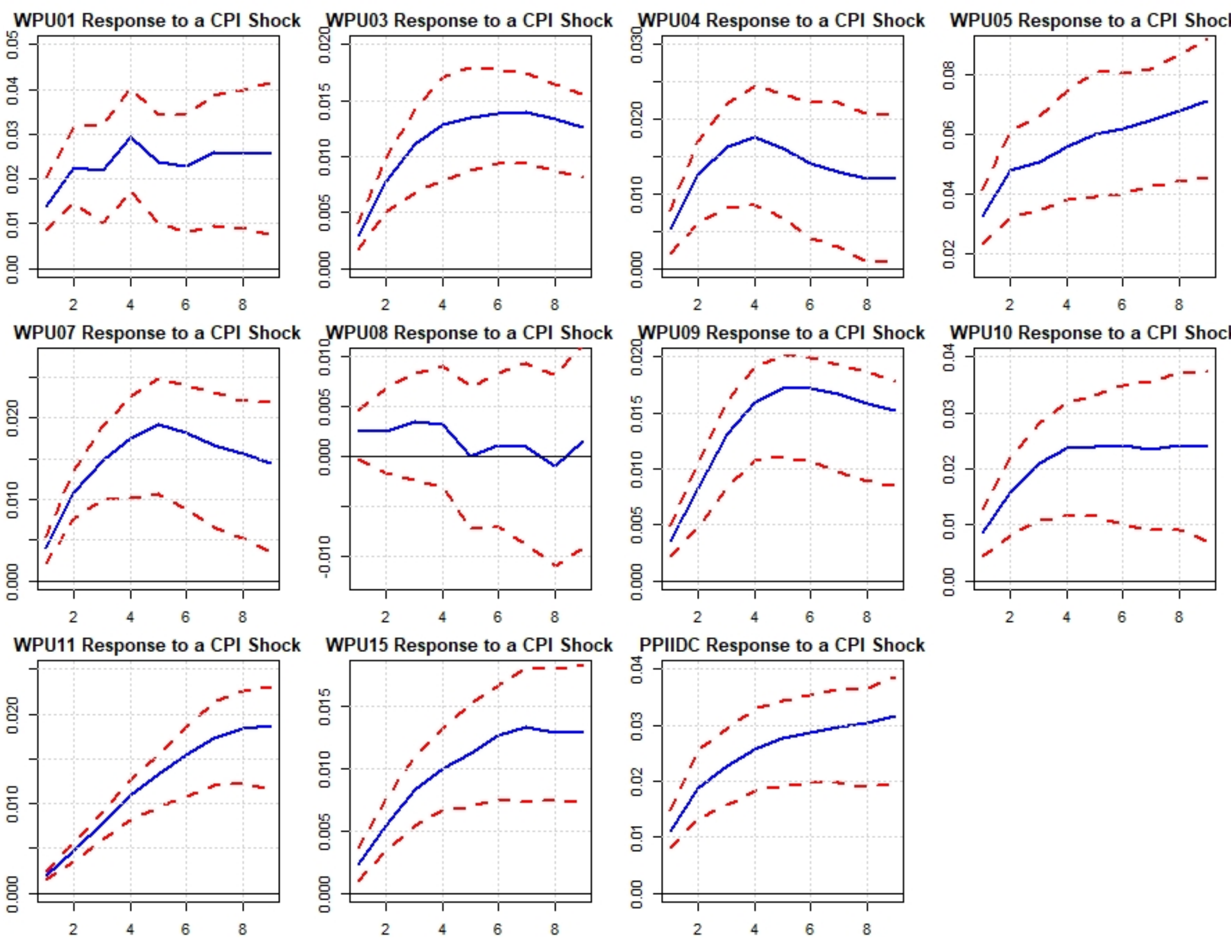


Figure 14: Impulse Response Functions: Disaggregate CPI Shocks

We note that in all cases each disaggregated PPI responds to a CPI shock positively and significantly with strong evidence of persistence. The only exception to this being WPU08, which is not different from zero. Taken with the CPI responses to a given shock from a disaggregated PPI, it would seem the IRFs generally support bidirectional causality with some exceptions of unidirectional causality.

Variance Decomposition

We conduct a variance decomposition exercise for each bivariate vector of CPI and a given disaggregated PPI. The first table below presents each vector and its respective forecast variables and reports the forecast variance that can be explained by CPI. The second table presents the same vectors, but reports the forecast variance that can be explained by a given disaggregated PPI. Each FEVD is reported for eight quarters ahead($t+1 ... t+8$). Table 29 presents these findings.

Table 29: FEVD: Disaggregate–Variance Explained by CPI

Vector	Forecast Variable	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$
$Z_t = [\log(CPI_t), \log(WPU01_t)]^T$	CPI	1.0000000	0.9932896	0.9814053	0.9800549	0.9717469	0.9561446	0.9438685	0.9328514
	WPU01	0.00000000	0.00671041	0.01859471	0.01994509	0.02825311	0.04385538	0.05613150	0.06714857
$Z_t = [\log(CPI_t), \log(WPU03_t)]^T$	CPI	1.0000000	0.9959037	0.9744527	0.9656421	0.9624066	0.9548616	0.9487162	0.9448564
	WPU03	0.000000000	0.004096258	0.025547251	0.034357939	0.037593407	0.045138447	0.051283799	0.055143566
$Z_t = [\log(CPI_t), \log(WPU04_t)]^T$	CPI	1.0000000	0.9909821	0.9863392	0.9794812	0.9655124	0.9496341	0.9332403	0.9166838
	WPU04	0.000000000	0.009017867	0.013660836	0.020518768	0.034487589	0.050365918	0.066759684	0.083316180
$Z_t = [\log(CPI_t), \log(WPU05_t)]^T$	CPI	1.0000000	0.9976403	0.9977115	0.9983863	0.9985158	0.9985265	0.9979646	0.9973441
	WPU05	0.000000000	0.002359664	0.002288530	0.001613708	0.001484173	0.001473499	0.002035388	0.002655931
$Z_t = [\log(CPI_t), \log(WPU07_t)]^T$	CPI	1.0000000	0.9895233	0.9443687	0.9106227	0.8973208	0.8852196	0.8735067	0.8623031
	WPU07	0.00000000	0.01047674	0.05563128	0.08937730	0.10267918	0.11478037	0.12649330	0.13769685
$Z_t = [\log(CPI_t), \log(WPU08_t)]^T$	CPI	1.0000000	0.9999336	0.9955404	0.9911719	0.9842363	0.9725734	0.9604847	0.9464984
	WPU08	0.000000000	6.644736e-05	4.459591e-03	8.828051e-03	1.576372e-02	2.742665e-02	3.951529e-02	5.350158e-02
$Z_t = [\log(CPI_t), \log(WPU09_t)]^T$	CPI	1.0000000	0.9902785	0.9574695	0.9375327	0.9204964	0.8971945	0.8768265	0.8627724
	WPU09	0.00000000	0.00972145	0.04253051	0.06246730	0.07950362	0.10280546	0.12317353	0.13722758
$Z_t = [\log(CPI_t), \log(WPU10_t)]^T$	CPI	1.0000000	0.9818189	0.9759156	0.9807351	0.9809655	0.9778818	0.9732265	0.9678511
	WPU10	0.00000000	0.01818105	0.02408444	0.01926494	0.01903449	0.02211820	0.02677345	0.03214887
$Z_t = [\log(CPI_t), \log(WPU11_t)]^T$	CPI	1.0000000	0.9889824	0.9757301	0.9734845	0.9698597	0.9682191	0.9685563	0.9671726
	WPU11	0.00000000	0.01101756	0.02426986	0.02651547	0.03014033	0.03178090	0.03144369	0.03282745
$Z_t = [\log(CPI_t), \log(WPU15_t)]^T$	CPI	1.0000000	0.9938027	0.9855789	0.9789228	0.9681924	0.9536615	0.9324464	0.9091795
	WPU15	0.000000000	0.006197281	0.014421064	0.021077175	0.031807600	0.046338465	0.067553567	0.090820477
$Z_t = [\log(CPI_t), \log(PPIIDC_t)]^T$	CPI	1.0000000	0.9886554	0.9856408	0.9888310	0.9900468	0.9863079	0.9803423	0.9749393
	PPIIDC	0.000000000	0.011344575	0.014359154	0.011168983	0.009953243	0.013692089	0.019657663	0.025060724

We observe that the predictive content of CPI does well to explain its own forecast variance, but is largely insufficient for explaining a given PPI's variance. This evidence suggests the degree to which CPI leads a given PPI is limited, if not, insignificant. We do note, however, that CPI is inadequate for accounting for the variance of a particular PPI in early forecast periods, but gains explanatory power (even if trivially) in later forecast periods in almost all cases. This is especially true in the case of WPU09 for which CPI explains zero percent of the variance one period ahead, but can explain up to almost 14% in later periods. With this in mind, we turn to our results presented in Table 30, which show a similar FEVD exercise, but looking at the variance explained by a given PPI, rather than CPI.

Table 30: FEVD: Disaggregate–Variance Explained by PPI

Vector	Forecast Variable	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$
$Z_t = [\log(CPI_t), \log(WPU01_t)]^T$	CPI	0.06741219	0.08039006	0.06710384	0.07153273	0.05910673	0.05316110	0.04993874	0.04785301
	WPU01	0.9325878	0.9196099	0.9328962	0.9284673	0.9408933	0.9468389	0.9500613	0.9521470
$Z_t = [\log(CPI_t), \log(WPU03_t)]^T$	CPI	0.1450808	0.2097028	0.2346337	0.2521460	0.2618637	0.2680743	0.2720980	0.2730622
	WPU03	0.8549192	0.7902972	0.7653663	0.7478540	0.7381363	0.7319257	0.7279020	0.7269378
$Z_t = [\log(CPI_t), \log(WPU04_t)]^T$	CPI	0.01588128	0.03594645	0.04652731	0.04352357	0.03719912	0.03998549	0.05079353	0.06419671
	WPU04	0.9841187	0.9640536	0.9534727	0.9564764	0.9628009	0.9600145	0.9492065	0.9358033
$Z_t = [\log(CPI_t), \log(WPU05_t)]^T$	CPI	0.3617848	0.3621495	0.3737311	0.3896815	0.3988494	0.4043953	0.4089034	0.4109407
	WPU05	0.6382152	0.6378505	0.6262689	0.6103185	0.6011506	0.5956047	0.5910966	0.5890593
$Z_t = [\log(CPI_t), \log(WPU07_t)]^T$	CPI	0.07644138	0.13260539	0.15394387	0.16200895	0.16572056	0.15902257	0.15070248	0.14485316
	WPU07	0.9235586	0.8673946	0.8460561	0.8379911	0.8342794	0.8409774	0.8492975	0.8551468
$Z_t = [\log(CPI_t), \log(WPU08_t)]^T$	CPI	0.001356478	0.004348352	0.007348413	0.021841709	0.042248489	0.062623187	0.090780611	0.125551229
	WPU08	0.9986435	0.9956516	0.9926516	0.9781583	0.9577515	0.9373768	0.9092194	0.8744488
$Z_t = [\log(CPI_t), \log(WPU09_t)]^T$	CPI	0.1471434	0.1959141	0.2535436	0.2927724	0.3050519	0.3038977	0.2998364	0.2953060
	WPU09	0.8528566	0.8040859	0.7464564	0.7072276	0.6949481	0.6961023	0.7001636	0.7046940
$Z_t = [\log(CPI_t), \log(WPU10_t)]^T$	CPI	0.2071219	0.2067706	0.2337875	0.2568649	0.2637020	0.2615387	0.2579429	0.2543923
	WPU10	0.7928781	0.7932294	0.7662125	0.7431351	0.7362980	0.7384613	0.7420571	0.7456077
$Z_t = [\log(CPI_t), \log(WPU11_t)]^T$	CPI	0.1261163	0.1652807	0.2161693	0.2744470	0.3181800	0.3559191	0.3851263	0.4033331
	WPU11	0.8738837	0.8347193	0.7838307	0.7255530	0.6818200	0.6440809	0.6148737	0.5966669
$Z_t = [\log(CPI_t), \log(WPU15_t)]^T$	CPI	0.04444907	0.10921943	0.18365710	0.22583202	0.25058129	0.28002270	0.29630683	0.29676327
	WPU15	0.9555509	0.8907806	0.8163429	0.7741680	0.7494187	0.7199773	0.7036932	0.7032367
$Z_t = [\log(CPI_t), \log(PPIIDC_t)]^T$	CPI	0.4686874	0.4780951	0.5071314	0.5326654	0.5380912	0.5322445	0.5259144	0.5184839
	PPIIDC	0.5313126	0.5219049	0.4928686	0.4673346	0.4619088	0.4677555	0.4740856	0.4815161

We note that the amount of variance explained by a given PPI is varying across both CPI forecast variables for a given vector. In the cases of WPU03, WPU05, WPU07, WPU09, WPU10, WPU11, and PPIIDC, we see that a given PPI can explain a significant amount of variance in CPI over the forecast horizon. Contrasting this with the variance explained by CPI for the same vectors, there is a stronger case to suggest that PPI and disaggregated PPI's provide more predictive information about CPI than CPI can provide for a given PPI.

Chapter 2 Learning by Doing, Productivity, and Growth: New Evidence on the Link between Micro and Macro Data

Scott Schuh, Brad Humphreys, Corey J.M. Williams

2.1 Introduction

Prior research used improvements in athletic performance over time to understand the relationship between changes in output and factors driving technological progress like learning-by-doing (LBD). A large and rich theoretical literature on the relationship between LBD and economic growth (Arrow, 1962; Jovanovic, 1996; Solow, 1997) motivated this analysis. Early studies of contest outcomes in the modern Olympic Games (Fellner, 1969) and the Indianapolis 500 automobile race (Barzel, 1972) developed deterministic measures of the impact of LBD, a factor affecting worker productivity, over long time periods on contest winning times (Fellner), speeds (Barzel), and the inverse of time (miles per hour). Both argued that time and speed over constant contest distances and venues are more standardized and less prone to measurement error than alternative measures of technological progress or productivity, which are notoriously hard to measure.[18]

Despite a clear, plausible link between observable proxy variables for LBD and outcomes reflecting productivity growth, relatively little subsequent empirical research used data from this setting. Mantel Jr et al. (1995) extended the analysis of Indy 500 outcomes to capture the roles of car technology (changes embodied in capital) and organizer rules and regulations. Munasinghe et al. (2001) and Preston & Johnson (2015) extended the focus in this line of research to record-setting times instead of winning times in human races, drawing from work on statistical distributions of the record setting process dating back to Chandler (1952). Overall, this line of research offers clear, but limited evidence of the importance of micro-founded "*changes* in [the process of] technology itself, defining technology as useful knowledge pertaining to the art of production . . . usually in a disaggregated way in firms and industries" (Kennedy & Thirlwall, 1972).

The macroeconomic literature on economic growth emphasizes ". . . the *effects* of changes in technology, or more specifically the role of technical progress in the growth process . . . " and studies ". . . which attempt to quantify the rate of technical progress as a determinant of the growth of output" (Kennedy & Thirlwall,

[18]For examples, see Diewert (1980), Sichel (2019), and the *Journal of Economic Perspectives* symposia in Fall 1988 and Spring 2017.

1972).[19] Macroeconomists measure aggregate technological change as either total factor productivity (TFP) constructed from the Solow (1956) and Swan (1956) growth model, or simply labor productivity (LP), defined as real output divided by labor input. Since the first generation of real business cycle models (Kydland & Prescott, 1982; Long Jr & Plosser, 1983), macroeconomists generally agreed that TFP is driven by a long-run stochastic trend.[20]

Thus far, little research empirically analyzed linkages between microfoundations like LBD and the macro implications of TFP for economic growth. However, some recent research integrates analysis of microfounded technological change and productivity with its aggregate implications in the general macroeconomy. See, for example Foster et al. (2001) and Morrison (2012). For examples of TFP and LP measurement, especially problems with measuring output and inputs, see and Syverson (2017). We empirically investigate the relationship between macro TFP and micro LBD using data from sports contests, an important extension to this literature.

This paper makes two contributions toward linking micro LBD and macro TFP. First, it updates and expands the auto racing data analyzed in previous studies. Relative to the existing literature, we extend the Indy 500 outcome data by nearly three decades (1993-2021), during which aggregate United States productivity growth fluctuated significantly, and include new race day control variables in the analysis.

Also, data on NHRA drag racing are analyzed for robustness and because these races are much shorter and potentially less susceptible to race-specific effects than the Indy 500. A second contribution applies, expands, and modernizes econometric models of long-run trends used in the literature to analyze outcomes of athletic competitions. Deterministic trend models from the early literature are estimated with new race day controls and modern econometric methods and compared with existing deterministic trend models of TFP. Jointly estimated stochastic trend models of speed and TFP from the contemporary literature also are presented and compared with results from deterministic trend models.

Analysis of updated data on automobile racing speeds show that after the early 1970's LBD continued to impact outcomes at roughly the same level as prior to the 1970's in both foot racing (track and field) and automobile racing. This result holds for both length-of-time and cumulative output specifications of LBD trends, but the best specification of these models (log-log versus semi-log) remains unclear. The data show that the impact of LBD in auto racing varies significantly across eras and drivers, and that the best drivers exhibit faster LBD than the average driver. However, the vast majority of decline in auto racing elapsed times can be explained by technological improvement in capital (race cars), rather than by driver-specific LBD; trends in automobile race speeds are better explained by trends in TFP than by trends in LBD.

[19] See the earlier survey by Hahn & Matthews (1964) for additional details about classifying the literature.
[20] See King et al. (1987), Engle & Granger (1987), Stock & Watson (1988) and Wickens (1996).

Productivity (TFP and LP) growth rates varied more following the 1973 Productivity Slowdown than in prior periods, but growth rates in Indy 500 and NHRA racing speeds vary less than aggregate productivity.

The results show that automobile racing speeds share a common long-run trend with TFP. Coefficients on the full sample deterministic time trends of speeds and TFP are statistically equal, and the hypothesis of cointegration (a stochastic trend) between speeds and TFP cannot be rejected. However, neither exogenous nor endogenous estimated subsample breakpoints of deterministic trend models align well among speeds or between speeds and TFP. Furthermore, estimated subsample deterministic trends do not match the major breakpoints in U.S. productivity growth (1973, mid-1990's, and mid-2000's) well.[21] Also, joint tests of deterministic and stochastic trends (Bai & Perron, 1998, 2003) support the latter. Perhaps most importantly, bivariate vector error-correction models (VECM) of TFP and speed estimated using the method of Johansen (1988, 1991, 1995) reveal the presence of only one cointegrating vector and significant dynamic adjustment of race speeds to TFP but not vice versa. This result supports the basic stochastic growth RBC model with aggregate TFP that diffuses throughout the economy, rather than micro LBD bubbling up to the macroeconomy. Over the full sample, TFP grows roughly twice as fast as auto racing speeds, suggesting that not all of TFP diffuses into auto racing performances.

Three specific results emerge from the empirical analysis. First, coefficients on the regression controls for various Indy 500 race day events generally have expected signs and magnitudes, especially past precipitation and crash incidents. But most are statistically insignificant and many controls have little or no trend, so they do contribute substantively to the observed changes over time. Secondly, we use central moment measures of speeds to avoid the impact of extreme values (winning or record speeds) in estimation and to match the aggregate (average) TFP data. Tests for normality and autocorrelation in residuals are mixed, neither rejecting or supporting our empirical approach. Small (time series) sample sizes and limited availability of data for all race cars in a year likely limit test clarity. Third, long-run cointegration coefficients are roughly consistent across speeds (about 0.4-0.5 per year). But the VECM coefficients for the actual races (Indy 500 and NHRA) are four times larger in absolute value than for the Indy 500 qualifying laps (-0.4 versus -0.1).

The econometric results in this paper motivate future research. Unexploited richness and heterogeneity in micro data offer opportunities for deeper, more insightful specifications of the structural diffusion of technological change and productivity plus improved measurement of the process. Better identification of race day effects on the diffusion process is needed. The results raise questions about the transmission mechanism of TFP to auto racing and other industries. Finally, at the industry level, modeling the demand for speed by consumers of auto racing entertainment is a missing micro foundation.

[21] See Romer (1987), Baily et al. (1988), and B. E. Hansen (2001) for exposition on consensus trend breaks in US productivity as well as competing views on the sources of economic slowdown, particularly from 1973 through the present day.

2.2 Existing Literature

This paper extends and combines two distinct branches of literature. One focuses on the microeconomics of technical change manifested in athletic performances over time. The other focuses on the macroeconomic effects of aggregate technical progress on output growth.

2.2.1 Technological Change and Sports Outcomes

A body of literature analyzes outcomes from sports competitions to understand the rate of technological progress over time. These papers posit that outcomes in sports contests reflect both increases in athlete (worker) productivity and improvements in capital like cars and engines in motor racing, skis in skiing, and swimsuits in swimming contests. These papers all point out a number of advantages inherent in sports data relative to other more commonly used macroeconomic data. These advantages include clean and consistently measured output variables and implicit or explicit controls for factors that affect output unobservable in other settings. For example, the Indianapolis 500 car race has been held annually at the same time of year on the same track under similar conditions for more than a century.

Fellner (1969) performed the first empirical analysis of progress using data from sports contests. This paper focused on understanding the role of "learning by doing," (LBD) defined by decreasing production costs over time when capital and output remain unchanged, in driving technological progress. Fellner (1969) developed two competing empirical proxy variables for learning by doing in automobile racing. The first used cumulative production as a proxy for increased learning by doing, reflecting the idea that doing more of some productive activity generated more LBD. The second used the passage of time as a proxy for LBD, reflecting the idea that performing some productive activity for a longer period of time generated more LBD.

Fellner (1969) analyzed data on the winning performance in the modern Summer Olympic Games held from 1896 to 1964 to assess the role played by performing activities longer in driving technological progress. Fellner (1969) estimated log-log regression models with winning outcomes as the dependent variable and the number of Olympic Games where the event was held as the explanatory variable. He focused on 11 men's Olympic events where capital remained unchanged. The paper reported evidence that the winning outcomes in all these events changed significantly with the number of Olympic Games where each was contested. Rates of increase averaged 5% to 7% in most events although swimming (14.4%) and Discuss (28.7%) increased more quickly. The results supported the idea that performing a productive activity for a longer period of time, a form of LBD, can explain observed technological progress.

Barzel (1972) analyzed data from automobile racing, winning speeds and times at the Indianapolis (Indy) 500, an open wheeled car race conducted annually since 1911. This paper exploited exogenous variation in the amount of time over which this race was contested, generated by breaks in competition from the two World Wars, to revisit the length of time LBD proxy employed by Fellner (1969). Barzel (1972) estimated log-log regression models using both winning speed and winning race time ($\frac{miles}{speed}$) as dependent variables and the number of previous races contested, a time trend, as the main explanatory variable. These models also included separate intercepts and time trend slopes for the pre-WWI and post-WWII periods, making the interwar period the omitted category. Data spanned the 1911 to 1969 competitions. The results from the winning speed model showed a faster rate of technological progress in the pre-WWI era, about 3.1% per year, followed by a slower rate of increase in the post-WWII era. The model using winning race time as the dependent variable generated similar results.

Mantel Jr et al. (1995) explored the role that car characteristics and race-organizer imposed characteristics played in determining Indy 500, and thus proxy for the rate of technological progress. The paper estimated linear regression models with separate slopes and intercepts for three discrete time periods: 1920-1922 and 1930-1937 (front engine cars with driver and riding mechanic), 1923-1929 and 1938-1960 (front engine cars with one driver), and 1961-1992 (rear engine cars). Mantel Jr et al. (1995) found a result opposite to Barzel (1972) in that the rate of change in the average qualifying speed in the post-1961 period was roughly double that in the earlier two time periods.

A few other papers analyzed the frequency of record setting in athletic events as a proxy for output. Munasinghe et al. (2001) analyzed the process describing record setting in track and field competitions. They compared record setting in two types of track and field competitions: competitions open to anyone in the world (world record times, Olympic record times, Milrose Games record times) and competitions open to a restricted group of athletes (the US record time and the New Jersey state track and field record times).

Munasinghe et al. (2001) argued that analyzing record breaks, and not actual performance offers advantages. In terms of methods, these outcomes can be analyzed nonparametrically, avoiding any strong assumptions about the underlying distribution of performance or changes in this distribution over time. In terms of measurement, Munasinghe et al. (2001) argued that records better reflect discrete changes by optimizing agents.

The results in Munasinghe et al. (2001) showed that the rate of technological progress, as reflected in record setting, remained constant over the 1900 to 1992 period, and that globalization did not affect the rate of change. The results also indicated that LBD played a role in explaining the rate of change, since the rate of record setting times for less experienced high school athletes fell below the rate in contests involving more experienced athletes in the world level competitions.

Preston & Johnson (2015) analyzed the frequency of record setting in the context of competitive swimming. This paper focused on the impact of innovations in swimsuit technology. They analyzed variation in the number of records set in a calendar year over the period 1969-2009. This variable contains many zeros (mean 1.1, min 0, max 5). The paper found that the number of new swim suit innovations introduced in a year was correlated with the number of swimming records set in that year. A new swimsuit innovation was associated with an increase of about 1/3 of a record in that year. But this sort of "counting" of innovations assumes homogeneity of the impact of each innovation on performance.

While these papers make methodological advances by focusing on record setting, we return to the performance based measures analyzed by Fellner (1969) and Barzel (1972). While record breaking may be a better indicator of technological progress, annual performance in the Indy 500 generates a time series that can be easily compared to other macroeconomic measures of technological change, like Total Factor Productivity and labor productivity, used in the broad macroeconomic literature on technological change. This allows us to link our results more closely to this broader literature than Munasinghe et al. (2001) and Preston & Johnson (2015).

2.2.2 Aggregate Technical Change

There is an extensive literature concerning the estimation and theory of productivity and, by extension, technological progress. At the core of this literature, there are three overarching themes of note: the theoretical underpinnings of productivity [or lack thereof], the [mis]measurement of productivity aggregates in econometric and empirical studies, and the subsample trends associated with TFP and LP.

An excellent primer to the overarching themes of the literature can be found in Hulten (2001), which provides an extensive history of total factor productivity as a measurement of productivity and technical progress within the field. In particular, Hulten (2001) provides an excellent synthesis and critique of the Solow residual, which has long been the staple for empirically measuring the aggregate level of productivity in a given setting. Related to our motivation, the authors highlight that the Solow residual does not directly and is likely capturing only the components of productivity that are directly affected by innovation. Additionally, because the Solow residual is captured by what is "left over" after measurable inputs are accounted for, the residual can be thought of as a "measure of ignorance" in a manner of speaking (Hulten, 2001).[22]

Delving further into the problem of how TFP is measured, Syverson (2017) introduces the idea of the mismeasurement hypothesis, which possesses two potential forms. The first form that is relevant to our motivation and methodology posits that after the end of the New Economy boom, products that embodied technological progress expressed a lower share of utility nested in their prices than during the beginning of the New Economy boom, thus a "slowdown" in tech progress after the New Economy era may be mismeasured

[22] See Van Beveren (2012) for additional exposition of potential issues with the Solow residual as a measurement of technical progress and productivity.

by the prices of the goods that embody the technology.[23]

Much like measuring TFP, the underlying theory of TFP and technological progress is as difficult to define. That being said, there are a handful of relevant papers worth highlighting. For instance, Prescott (1998) provides a critique of the traditional viewpoint that differences in the capital stock account for large differences in output across countries. Across time and different nations, total factor productivity accounts the most for these differences. The author implores the field to consider the importance of a developing a theoretical underpinning for total factor productivity. In recent years, the literature has taken the critiques of Prescott (1998) well. For example, Nordhaus (2021) introduces a new growth model built on the idea of technological singularity. Effectively, the author proposes a singularity threshold where, once crossed, technological growth will be the driving pace of economic growth, rather than a residual effect of long-run economic growth.

A critique of the Nordhaus (2021) model can be seen when disaggregating economic production down to the industry-level, however. For instance, in the case of auto racing, technological growth is likely embodied in the capital stock (engine, chassis) of the race cars themselves, which may not benefit from the ever-decreasing costs of computation speed and power (Nordhaus, 2021). One such work that disaggregates down to industry-specific sources of TFP would be Foerster et al. (2019) who examine the variation in the trend of TFP at a sectoral level in the economy and relate sector-specific productivity slowdowns to aggregate productivity trends and GDP growth. When evaluating the decline in GDP growth in the post-WWII era, the authors find that such a decline is attributable to sector-specific factors, rather than aggregate factors. Similarly, Wolff (1985) performs a decomposition of productivity growth to investigate the Productivity Slowdown of the 1970's. Through this exercise, around one-fifth to one-sixth of the slowdown was attributable to a decline in overall productivity growth while the remainder was largely attributable to industry-specific sources.[24]

In terms of historical trends, productivity grew rapidly during the world wars and continued to grow up until the mid-1970's, which was the beginning of the Productivity Slowdown. TFP and LP growth were stagnant and, at times, negative during this era up through the late-1990's, at which point TFP and LP growth spiked during the New Economy boom, and then receded shortly into the early 2000's. There was a slow recovery of growth up until the Financial Crisis, at which point TFP and LP growth declined again, and has been recovering slowly since. There are a handful of selected papers that discuss these trends at length.

Closer to the present, D. M. Byrne et al. (2016) examine the declines in productivity aggregates after

[23]The second, though less relevant form, argues that it is possible that the price deflators of new technology are rising faster relative to their prices in the New Economy, thus leading output growth in nominal terms to be mismeasured (Syverson, 2017).

[24]For recent works looking at heterogeneity in TFP growth in the context of nations, rather than industries, see Boppart & Li (2021).

2004 (post Dot-Com era) and find that measurement error does not explain the slowdown. Furthermore, the authors differentiate non-market gains such as consumer benefits from smartphones and other IT services from market-sector production, and find that the slowdown in market sector production outweighs the gains in non-market activities. These findings suggest that ongoing improvements in IT services is not spilling over into other sectors and innovations to productivity are diffusing at a slower rate than they may have been prior to the slowdown.

A complementary analysis to D. M. Byrne et al. (2016) would be Fernald (2015), which evaluates the trends in US productivity and potential output over course of the Great Recession era, and finds that the bulk of the pre-recession productivity slowdown is attributable to industries and markets that produce or rely heavily on information technologies. In a manner of speaking, the "waning" growth effects of information technology is the primary factor contributing to the pre-recession slowdown. On a related note, Fernald et al. (2017) pay special attention to the post-Financial Crisis productivity recovery, and utilizing a growth accounting framework to explain the lackluster recovering since the trough of the Financial Crisis in 2009. The authors find that the slow recovery is attributable to the slowdown of total factor productivity, and a decline in the labor force participation rate.

2.3 Updated Evidence on LBD

This section first describes data sources and then presents and analyzes updated evidence on LBD in auto and foot racing contests.

2.3.1 Data

Auto racing data come from two sources. First, Indianapolis 500 data for 1911–2022 are from the Indianapolis Motor Speedway's historical archives, which extend a half century beyond Barzel (1972), and three decades beyond Mantel Jr et al. (1995). The Indy 500 data include race results by driver and by year, plus additional data on race day incidents, prize distribution, and so forth. Two speeds are available: 1) the Indy 500 race; and 2) the Indy pre-race qualifying trials that determine drivers' starting pole positions. Second, National Hot Rod Association (NHRA) Winternationals drag racing data for 1961–2022 are constructed from: ProQuest news articles containing race results and the results archive maintained by the NHRA.[25] Adjusted 100M dash times and mile-run record times are from the World Athletics organization.

Aggregate US productivity data come from two sources. First, annual productivity data for 1890–2018

[25] Unlike the Indy 500 (and marathons or mile races), which have many simultaneous competitors and take hours to complete, drag races have two parallel, simultaneous competitors and take seconds to complete. Mantel Jr et al. (1995) emphasize the importance of extensive changes to rules and regulations governing the more than two-hour Indy 500, and the number and severity of crashes varies widely by year—both of which influence the winning speed.

come from the Long-Term Productivity Database (LTPD) created by the Bank of France.[26] Two types of aggregate productivity data are available: 1) total factor productivity (TFP); and 2) labor productivity (LP), or output per hour. A second source of annual (TFP) and quarterly (LP) productivity in the non-farm business sector for 1948–2021 is Bureau of Labor Statistics (BLS) or FRED.

Potentially important differences arise in the measurement of athletic performances (time and speed) found in the literature. Most racing performances are *extreme value* measures (winner, world record) whereas TFP and LP are *central moments* (average across all agents). Thus, our Indy 500 race speeds are field *averages* to match TFP, but the NHRA drag race data are limited to winning speeds. Empirically, differences between trends in extreme value and central moment data are modest, but we test for evidence of non-normality in the deterministic trend models.

2.3.2 Athletic Performance and LBD

The literature contains several important empirical issues related to output measures and LBD that merit further discussion. Heterogeneity in inputs and outcomes *across* sports (track and field, swimming, autos) represents one key issue. Although research in this area focuses on outcomes over time, foot and swim racing depends on limited capital inputs (shoes, track surfaces, pools, bathing suits) while auto racing depends more heavily on capital inputs (cars, engines, suspension systems). Auto race drivers tend to have longer careers, which provides rich variation in individual LBD. Another issue is heterogeneity in inputs and outcomes *within* sports, where LBD may differ between short versus long distance events, or between events based on time versus distance (Fellner, 1969). Practical issues in productivity measurement also arise, where outcome choices include elapsed time versus speed (inverses of each other) and winning times of regularly scheduled events versus irregularly occurring record times. Finally, the functional forms of LBD proxy variables vary.

Performance and LBD Over Time

Fellner (1969) defined athletic performance, P_t, as the inverse of labor productivity, $P_t = L/Y$, where L is labor input (time) and Y is a fixed quantity of output, distance traveled. P_t represents the elapsed to cover a certain distance (from 100M to 500 miles).[27] Fellner (1969) argued that LBD manifests as declines in elapsed race times (increases in labor productivity) over many years athletes participate repeatedly in the same race. This implies that elapsed race times should decline over time.

Figure 15 plots time series of the log of elapsed times in two long distance competitions, the Indy 500 mile race and one-mile foot races, and two shorter distance races, the NHRA Winternationals quarter-mile

[26]We thank Dan Sichel for referring us to these data, which extend farther back in time (1890) than the Bureau of Labor Statistics (BLS) multi-factor productivity estimates (1948) and Penn World Tables TFP estimates (1954).

[27]All times are converted to a common time unit, one hour, for comparability across panels.

drag race and the 100-meter dash. These appear as black lines on Figure 15. The panels in Figure 15 are analogous to figures presented in Fellner (1969), except the auto races are annual winning times, and the track and field times are world records. The vertical lines denote the terminal years of data in Fellner (1969) (track and field) and Barzel (1972) (autos), both of which occurred shortly before the US productivity growth slowdown in 1973.

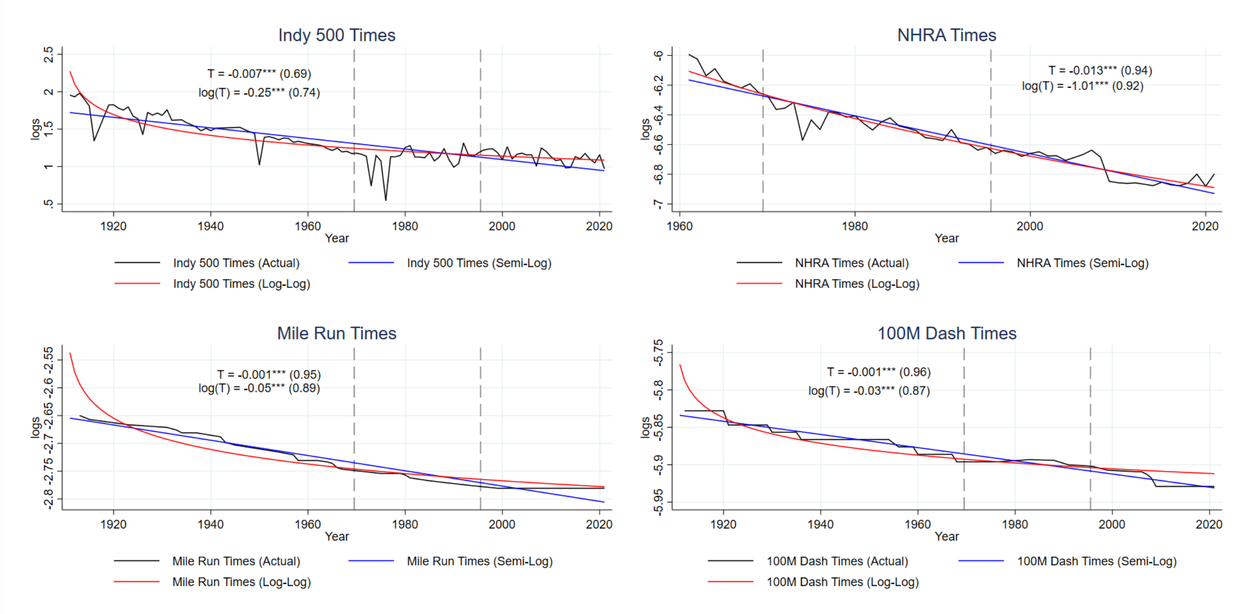


Figure 15: Elapsed Times in Short and Long Foot and Auto Races

Figure 15 contains evidence of widespread sustained declines in actual elapsed times across all foot and auto competitions for all distances, not only in periods analyzed in the early literature but also over the following decades for both winning and record times. The left panel shows the number of hours (Indy 500), or minutes (mile run) that the winning participant took to complete the event. In the case of the mile run, this is the world record time. Fellner (1969) noted that the number of units of time needed to complete a race reflects the number of units of labor inputs needed to produce one unit of output. If the world record time in the mile race fell from 4 minutes to 3 minutes and 50 seconds, the cost of producing output fell because fewer units of inputs were required to produce the same unit of output. The right panel shows the number of seconds the winning participant took to complete the NHRA drag race and 100 meter dash. Actual time elapsed in these short competitions also shows steady decreases over time.

Fellner (1969) hypothesized that competitors acquire productivity-enhancing human capital from past race experience (LBD) by repeatedly performing the same task over time. Fellner (1969) proposed two alternative model specifications for analyzing changes in athletic performance over time (T). Both assume that the passage of time represents a good proxy variable for increases in LBD. These models take the form:

$$\ln P_t = \beta_0 + \beta_1 f_i(T), \quad i = \{1, 2\}, \quad \beta_1 < 0$$

Where $f_1(T) = \ln T$ and $f_2(T) = T$. Again, these models assume that a time trend variable captures changes in LBD over time. The fitted values from these two model specifications, referred to as "log-log" and "semi-log," models respectively, appear in Figure 15 as red and blue lines respectively. Fellner (1969) hypothesized that the log-log model would best capture LBD in events that had very few "improvable dimensions" and were heavily dependent on "given facilities and a given model," but the semi-log model would better fit events "rich in improvable dimensions" (including technological progress generated by the macroeconomic growth path). From an econometric standpoint, he found short running events were best represented by log-log models and long running events (5,000 meters or more) by semi-log models.

Figure 15 shows that the semi-log and log-log trend lines also exhibit clear downward slopes over the entire sample period for all four events. Even with a large amount of additional data, the trend model regression results summarized on Figure 15 shows that the best fitting functional form for the model of elapsed times over the sample remains uncertain. Based on R^2, the model fits over the full sample are better for the semi-log model in three of the four events, although the differences are modest. Subsample results split at 1973 (not reported) show a consistent decline in the elapsed auto racing times since 1973 for the log-log specification, but little change in the pace for auto racing semi-log models or in any track and field models.[28]

The R^2 declines modestly (up to 6 percentage points) after 1973 for six of the eight models and the relative performance across models after 1973 differs from before 1973. Both results underscore the lack of clarity in the literature pre-1973 about the preferred time trend model specification. Most importantly, neither specification is likely the best fit for all time periods. Semi-log models predict future elapsed times will eventually reach zero or become negative, an impossible outcome. On the other hand, log-log models backcast implausibly slow elapsed times in the past.[29]

LBD and Cumulative Experience

The use of time trends to quantify changes in LBD represents a relatively crude approach. Other approaches exist in the literature. The seminal paper by Alchian (1963) assumed that LBD increases directly with cumulative output. In other words performing an activity more times regardless of time elapsed proxies

[28] Results available upon request.

[29] Of course, with more data both specifications would continue to look sensible in sample as the estimated curvature and slopes diminish in absolute. However, the asymptotic predictions in both directions of time would remain implausible. This result is analogous to the debate between Lucas Jr & Nicolini (2015) and P. Ireland (2015) over the functional form of money demand.

for LBD. More recently, Levitt et al. (2013) assumed that learning by doing at a microeconomic level is analogous to an inventory stored and accumulated over time.

In the context of data from the Indy 500, learning can be defined as a driver's participation in the race (actual or qualifying trial) in a particular year, Y_{it}.[30] In terms of empirical analysis let X_{it} denote the cumulative previous experience of driver i in year t (the total number of previous appearances in the Indy 500), and $X_t = (1/N^d)\sum_i^{N^d} X_{it}$ denote average cumulative experience of the entire race field in year t. Unlike most track and field events, some Indy 500 drivers accumulate decades of experience that offers unique measures of individual-specific LBD.

To analyze the relationship between cumulative experience and race outcomes, we use data on qualifying times for the Indy 500 and not finishing times due to data limitations. Qualifying times come from individual cars and drivers completing a single lap around the track in the weeks leading up to the actual race. Qualifying time data are available for all drivers who complete the trials, while times from the actual race are available only for the subset of drivers still racing when the winner received the checkered flag. This is often less than half the initial field. Qualifying times provide a simpler, standardized measure of learning how to drive the track but abstract from learning about how to negotiate the track crowded with other cars during the actual race. Qualifying times are in minutes per lap of 2.5 miles; 1.0 minute is 150 MPH.

Figure 16 plots average elapsed times in the Indy 500 qualifying trials by cumulative level of driver experience (number of previous appearances) over the full sample 1911-2022. The red line depicts the actual elapsed time of the best qualifying lap. As expected, cumulative experience lowers the qualifying times. During the first 15 years, the elapsed time declines by 0.3 minutes (or about 64 MPH) but then flattens during the subsequent 15 years. In auto racing, improvement in elapsed times is heavily influenced by technological progress in capital (race cars). The black line plots demeaned elapsed times (actual elapsed time in year t minus the average qualifying time for all drivers in year t) to abstract from improvements in car technology. Driver-specific LBD exhibits a qualitatively similar pattern, declining steadily for 15 years and then flattening afterward. The decline in driver-specific LBD is about 0.006 minutes (or about 0.9 MPH) during the first 15 years of a driver's career.

[30] As an empirical matter, completion of the race (all 200 laps and 500 miles) is the best definition of output. However, a driver completing only part of the race still acquires experience that can be measured in elapsed time or miles per hour for the distance completed.

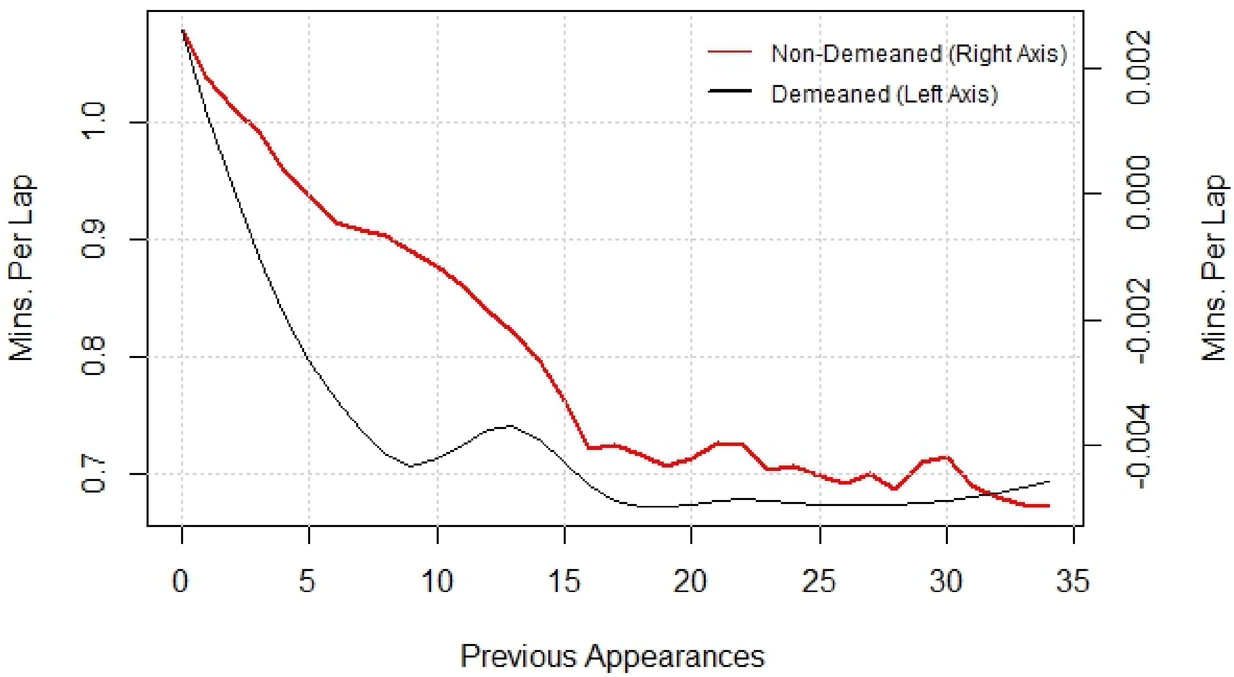


Figure 16: Indy 500 Qualifying Times by Driver Experience

Interestingly, the data suggest the pace of driver-specific (demeaned) LBD changes over calendar time (eras) and across drivers, as shown in Figure 17. The left panel shows that LBD was fast in 1911-1945 when qualifying times declined sharply with driver experience during the first five years. Since 1946, however, driver-specific LBD has been gradual during the first 10 years of experience. After 10 years, driver-specific LBD accelerated in 1946-1964 and, to a lesser extent, in 1965-1984. Since 1985, however, driver-specific LBD has remained stable or deteriorated after 15 years of experience. The right panel shows how LBD varies across of some of the top Indy Car drivers of all time.[31]

All top drivers, except Mario Andretti, exhibit faster LBD than the average driver (black line) during the first decades of their careers, as might be expected from their relative success. After the first decade, however, all of the top drivers see their LBD advantage fade; some even become relatively less adept than the average driver. Heterogeneity across drivers is considerable. Mario Andretti, in particular, began his career 0.025 minutes (about 0.4 MPH) faster than the average first-year driver, and maintained that advantage throughout his career. In contrast, Johnny Rutherford began his career somewhat slower than the average

[31] For a list of the drivers and their accomplishments, see The Bleacher Report and its ranking of the top 25 Indy car drivers.

first-year driver but improved significantly for 15 years, then lost his advantage after 20 years.[32]

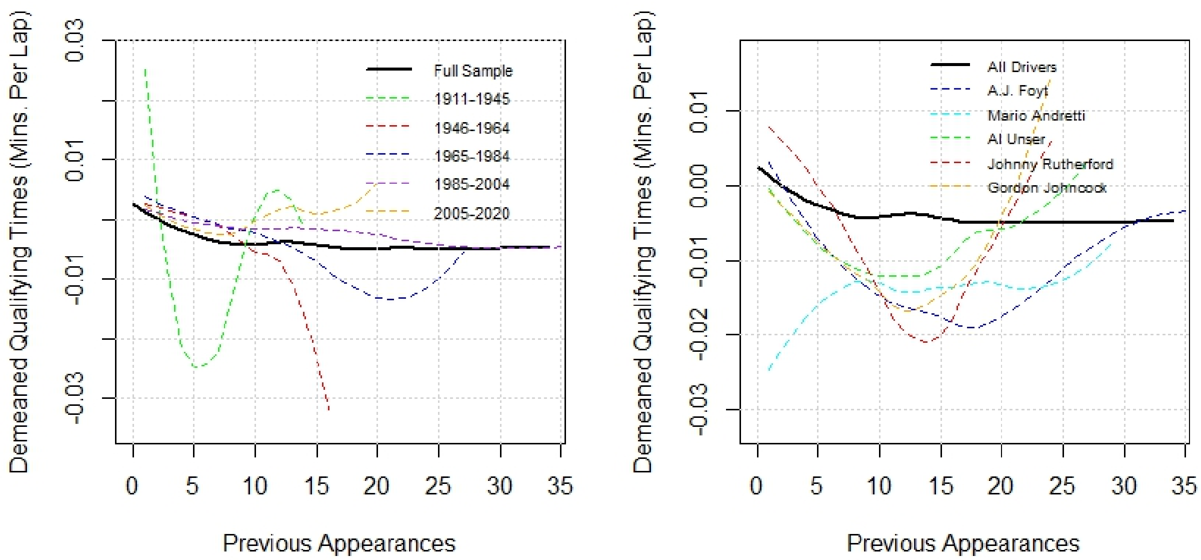


Figure 17: Indy 500 Qualifying Times (Demeaned) by Experience Across Eras and Drivers

2.3.3 Auto Racing Outcomes and Aggregate Productivity

Results in the preceding section suggest four initial conclusions. First, elapsed times in auto and foot races declined substantially over the entire 1911-2022 period. Second, LBD has a positive relationship with output over this period. Third, improved athletic performances in auto racing are mainly attributable to technological progress in capital (race cars) over time. Improvements in labor input stemming from LBD are evident but modest in absolute terms, generating an increase of about 1 MPH compared to 64 MPH from automobiles. For these reasons, factors besides LBD likely contribute to improvements in auto racing performances, and alternative functional specifications of technological progress may be more effective. Fourth, when athletic performance as measured by lower elapsed times (P_t) proxies for technological progress, the appropriate functional form of the relationship between technological progress and time is ambiguous. Neither the log-log nor semi-log model specifications clearly fits the data best overall.

We next investigate the relationship between racing outcomes and changes in aggregate US productivity. Figure 18 plots data on US TFP and three auto racing speed outcomes since 1911. We convert all variables to a common index (1947 = 100) and plot these indexes in natural logs to facilitate growth comparisons and highlight basic stylized facts. The dashed vertical lines identify the ends of the samples in Barzel (1972)

[32] The results in this subsection are similar in spirit to Levitt et al. (2013), which motivates specification of LBD in auto racing as $P_t = \beta X_t^{\gamma}$, where $\gamma < 0$ is effectively a learning rate. Appendix B provides more details and evidence on this approach.

and Mantel Jr et al. (1995) (1969 and 1992, respectively). Total factor productivity (TFP) represents a broader measure of technological change that embodies factors affecting the quality of race cars, pit crew servicing, and other elements determine race outcomes.[33] For this analysis, we switch from elapsed race times, $P = L/Y$, to speed, $S = P^{-1} = Y/L$, the distance driven divided by the elapsed time to better match changes in TFP, which increase over time. In this context, we interpret speed as analogous to labor productivity, or output per hour, in a conventional production function for goods and services.

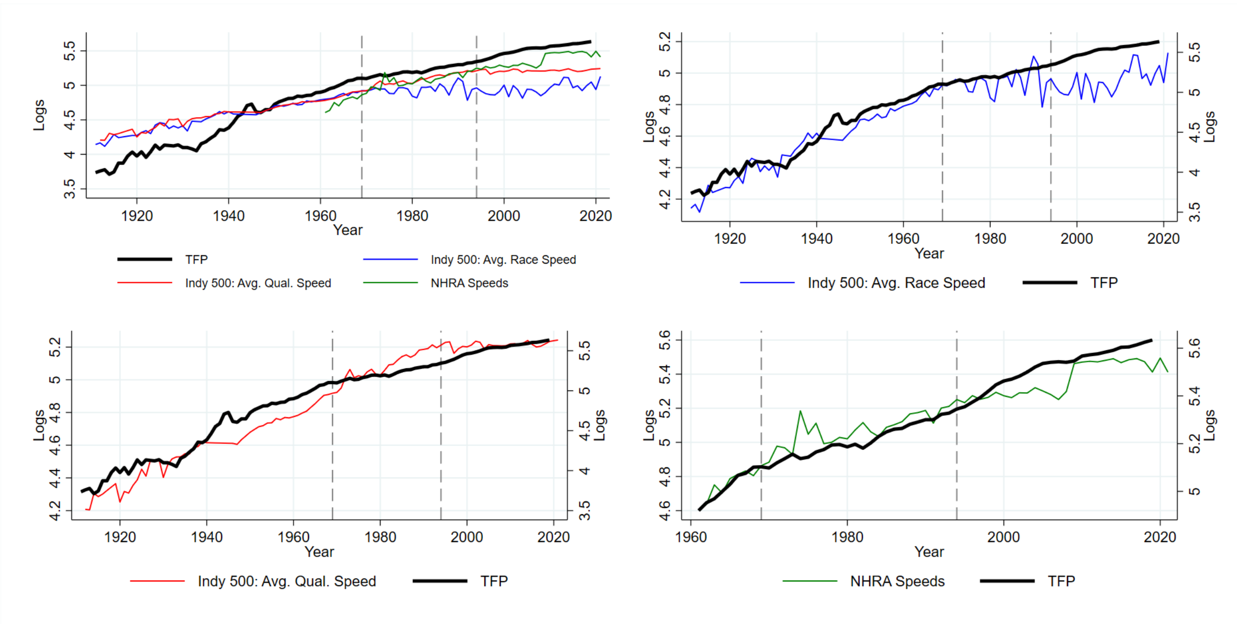


Figure 18: Total Factor Productivity and Auto Racing Speeds, 1911-2021

TFP grew about twice as fast (1.8 percent per year) as the Indy 500 race and qualifying speeds (0.8 and 1.0 percent per year, respectively) over the full sample, as shown in the upper left panel of Figure 18 with a common y-axis scale. Aggregate technology diffusing less than one-for-one into speeds suggests that some components of general aggregate technology are irrelevant for auto racing or an idiosyncratic part of technological change in auto racing offsets some TFP growth.

Indy 500 speeds grew at about the same rate until the 1970's, when race speeds began growing considerably slower than qualifying speeds. Since the 1960's, drag racing speeds grew about the same as Indy qualifying speeds, 0.7% versus 0.6% percent annual rates, respectively. A key difference between race and qualifying speeds is the former can reflect additional complications (in-race events like caution flags, strategic driving interactions, etc.) or race-specific rules and regulations (for which we do not have data) that influence their relative trends. The remaining three panels plot TFP pairwise with each speed using two y-axis scales to highlight the commonality of their trends, which are evident in the econometric models later.

[33] Long historical time series of TFP were unavailable at the time of earlier research in the literature.

Two important developments occurred since the early research on Indy 500 speeds. First, there have been no more World Wars, which Barzel (1972) interpreted as a primary cause of major disruptions to production and technology trends. Second, growth in U.S. labor productivity (BLS) has fluctuated frequently since World War II, as shown in the first two rows of Table 31. A major Productivity Slowdown occurred around 1973–almost exactly the end of the Barzel (1972) sample–after which growth rates of speed and productivity fell by more than half.[34] Productivity growth increased from the mid-1990's to mid-2000's–almost exactly after the end of the Mantel Jr et al. (1995) sample–and were about 50 percent higher. However, speeds did not increase commensurately, perhaps because auto racing less impacted investment in information and communication technology (ITC) than the rest of the economy.[35] In any case, the New Economy/Dot-Com boom was transitory and productivity growth declined in the mid-2000's to about its 1973-1994 average.[36] The literature debates whether this decline reflects a second slowdown or increased error from measuring productivity in service industries and complications from information technology.[37]

Table 31: Annual Growth Rates of Speeds and Productivity by Era

	Speeds		Productivity	
Era	Race	Qualifying	LP (BLS)	TFP
1911–1973	**1.42**	**1.56**	**N/A**	**2.28**
1911–1916	1.68	1.89	N/A	2.32
1919–1941	1.36	1.09	N/A	1.88
1946–1973	1.34	1.63	2.78	1.95
1974–2021	**0.38**	**0.48**	**1.93**	**1.09**
1974–1994	0.07	0.96	1.61	0.99
1995–2006	0.19	-0.16	2.72	1.56
2007–2021	1.52	0.23	1.73	0.70

2.4 Deterministic Trend Models

This section describes single-equation time-series models of long-run growth in auto racing speeds and technology based on deterministic trends that shift discretely over time. As noted in Section 2.2, this empirical approach was common prior to the 1980s when analyzing the relationship between technology and auto racing speeds and the Productivity Slowdown. This paper extends these literatures by formally testing whether deterministic trends in auto racing speeds coincide with trends in technical change measured by total factor productivity (TFP).

[34]See Griliches (1980) and Romer (1987) for more details on the Productivity Slowdown.
[35]For more on ITC capital investment, see Vu et al. (2020), Brynjolfsson & Hitt (2003), and Brynjolfsson & McAfee (2014) among others.
[36]See Jorgenson et al. (2007) and Jorgenson et al. (2008) for more details on the New Economy Boom/Bust. P. N. Ireland & Schuh (2008) argue these low-frequency TFP dynamics are best interpreted as a one-time shift in the level of investment-sector TFP in a two-sector RBC model.
[37]See Andrews et al. (2016) for exposition on productivity after the Financial Crisis.

2.4.1 Econometric Models

For simplicity, let S_t^k denote speeds for $k = \{r, q\}$, where r is the Indy race and q is Indy qualfying; lowercase variables s_t^k denote log levels. The econometric models for race speeds are described below by equation (2.1).

$$s_t^k = \alpha_{s0} + \beta_{s0}T + \left[\sum_{j=1}^{N^D} \alpha_{sj}D_j + \beta_j(D_j \times T) \right] + \gamma_s CONTROLS_t + \varepsilon_{st}^k \qquad (2.1)$$

Our models described generally by equation (2.1) are extended versions of the log transformation $f_2(S^k) = \ln(S_t^k) = s_t^k$ in Barzel (1972), which also correspond to the standard specification of deterministic models of technology and productivity.[38] T is a deterministic time trend; D_j are N^D dummy variables that capture exogenous shifts in the time trend T; and $CONTROLS$ is a set of variables that might influence estimation of the trend coefficients. As explained in Section 2.3.1, the data on race speeds are averages of the field so the regression error ε_{st} is assumed to be independently and identically distributed $N(0, \sigma^2_{\varepsilon_s^k})$, rather than following an extreme value distribution, which might be warranted with data based on winning speeds or world records. Omitted variables or unobservables that affect the trend in race speed are captured in the regression error.

We estimate equation (2.1) using data from 1911-2022, extending the period of analysis compared to earlier papers. Barzel (1972) (1911-1969) examined the impact of World Wars I and II impacted technical change through spillover improvements to automobile production. More recent data (1994-2022) incorporates the New Economy and Dot-Com boom-bust dynamics around the turn of the 21^{st} century, which influenced technical change and productivity growth in new ways. Equation (2.1) incorporates medium-term fluctuations in technical change and productivity growth using deterministic trends identified from the literature and summarized in Table 32. The interwar period (1917-1945) represents the omitted trend variable. Where applicable, $D_3' = D_3 + D_4 + D_5$ represents a dummy variable identifying the period following the Productivity Slowdown in 1974.

[38] Barzel (1972) also estimated transformations $f_1(S^k) = S^k$ and $f_3(S^k) = 500/S^k$, the latter being equivalent to the elapsed time for the 500-mile race (or adjusted for the number of miles actually completed); the results were not significantly different across transformations.

Table 32: Variable Definitions

Variable	Description	Units
T	Time Trend	$[1, ..., 112]$
D_1	1911-1916 (pre-WWI)	Dummy
D_2	1946-1973 (post-WWII)	Dummy
D_3	1974-1995 (Productivity Slowdown)	Dummy
D_4	1996-2005 (New Economy Boom)	Dummy
D_5	2006-2022 (Productivity Bust)	Dummy
$Prec_t$	Precipitation	Inches
Inc_t	Incidents Per Lap	$(0, 1]$
$Spread_t$	Deflated Prize Spread	logs
$Temp_t$	High Temperature	logs
Exp_t	Average Field Experience	logs
$Field_t$	Field Size	logs
$Rider_t$	Number of Vehicle Occupants	$\{1, 2\}$
$Pole_t$	Pole Position	logs

Equation (2.1) also includes a vector of $CONTROLS$ containing two types of variables that might influence trends in race speed. One type captures variation in race-day conditions, including: precipitation ($Prec_t$), which may reduce race speed even with improvements in tire technologies; the number of incidents per race lap, (Inc_t), which reflects caution laps run at lower speeds; $Spread_t$, which measures dispersion in the dollar value of driver winnings that induces greater competition among drivers (hence higher speeds); and $Temp$, the ambient temperature during the race.

The other type of variable captures trends in rules set by race organizers pertaining to improvements in car technologies and governing the actually running of the Indy 500 race, including: experience, Exp, which measures the total number of prior Indy 500 races run by field of drivers in year t; size of the field, $Field_t$, which is the total number of drivers; $Rider_t$, which measures the number of vehicle occupants; and pole position, $Pole$, which measures the extent to which the winning driver had to negotiate past other drivers.

Analogous econometric analyzing changes in technology, measured as total factor productivity (TFP), A_t, can be described by equation (2.2).

$$a_t = \alpha_{a0} + \beta_{a0}T + \left[\sum_{j=1}^{N^D} \alpha_{aj}D_j + \beta_j(D_j \times T)\right] + \varepsilon_{at} \tag{2.2}$$

Where $a_t = \ln(A_t)$, and the deterministic trends (T and D_j) are the same as for race speeds. $CONTROLS$ for auto racing presumably do not influence aggregate TFP (at least not directly), and thus are excluded.

Equations (2.1) and (2.2) are estimated with ordinary least squares (OLS). In addition to coefficient estimates for each equation, results from tests of the null hypothesis, $H_0 : \alpha_{sj} = \alpha_{aj}\ \forall\ j = \{1, \ldots, N^D\}$, are of central interest. These tests are conducted by jointly estimating bivariate speed and TFP models using seemingly unrelated regression (SUR) estimates of the common deterministic trends. Rejection of these null hypotheses provides evidence against the view that race speeds and technical change share common deterministic trends. The regression residuals also are tested for normality using the Jarque-Bera test to assess whether the coefficient estimates might be drawn from a non-normal distribution, perhaps due to imperfect averaging of the race speed data.

2.4.2 Deterministic Trend Results

Table 33 contains results for the deterministic trend models of Indy 500 speeds and TFP.[39] Coefficient estimates are multiplied by 100 to reflect annual percent changes. For each dependent variable, the table reports four sets of model estimates using samples of increasing length from left to right. Qualifying speed data begin in 1912 instead of 1911. The first sample (1911-1973) shows the effects of including $CONTROLS$, which were not originally in Barzel (1972).[40] The second sample (1911-1994) is similar to that in Mantel Jr et al. (1995) and includes the influence of the Productivity Slowdown. The final two samples (through 2005 and 2022) incorporate the effects of the boom-bust cycle in productivity around the beginning of the 21^{st} century.

The results in Table 33 reveal broad similarity, and general statistical significance in the deterministic trend estimates across dependent variables. Several key conclusions can be drawn from these results:

1. **Full-sample trends**—Estimated coefficients on the full-sample time trend (T) are economically and statistically the same. Both speeds and TFP grow about 1.6 to 1.7 percent per year on average after accounting for trend breaks and controls. *This result suggests that race speeds and TFP share a common trend, at least in the very long run*

2. **Subsample trends within dependent variables**—Estimated coefficients on subsample time trends (D_i) generally are stable (robust) across sample periods as more time series data are added. However, about half of the subsample trend estimates are statistically insignificant, especially during the earlier time periods. *This result suggests that adding considerably more time series data does not alter the results in an economically significant way*

3. **Subsample trends across speeds**—Estimated coefficients on subsample time trends (D_i) and across sample periods are qualitatively similar between race and qualifying speeds, although the difference in

[39] NHRA drag race speeds are excluded from this analysis because the data begin in 1961, and thus are too short for comparison with the full Indy 500 data. Unreported results for NHRA deterministic trend models show a small positive improvement in the elapsed time of winning cars over the productivity slowdown (D_3) of around 0.003 and a similar-in-magnitude improvement from the productivity bust onward (D_5) with zero improvement over the course of the Dot-Com boom/bust (D_4). Given limitations to our NHRA sample size, the signs and significance of these trend growth rates should be taken lightly.

[40] Estimates of the Barzel (1972) model over the original sample (through 1969) with *winning* race speeds but no controls are replicated exactly and available upon request. Estimates in Table 33 with *average* race speed are not economically or statistically significantly different from the original Barzel (1972) results.

quantitative magnitudes is sometimes economically significant. *This result indicates that trend breaks are not the same for race and qualifying speeds, so the choice of speed measure will likely influence the analysis of the relationship between speed and TFP*

4. **Subsample trends between speeds and TFP**—Estimated coefficients on subsample time trends (D_i) for TFP exhibit economically large differences with those for speeds, although the degree of statistical significance in the speed and TFP estimates for subsample trends is similar in general. *This result suggests that although speed and TFP share a clear common trend in the long run, the two variables exhibit quite difference trend breaks in subsamples*

5. **Controls**—Estimated coefficients on most of the controls for race speed are not statistically significant except for precipitation and number of incidents, which are both the expected sign (negative). Only some of the controls are available and relevant for qualifying speed; most notably, coefficients for the number of riders are economically large and statistically significant—more so than for race speed, interestingly. *These results suggest that the controls generally are not important determinants of the trend estimates, so their omission from earlier studies probably did not bias previous results*

6. **Regression diagnostics**—The adjusted R^2 statistics are uniformly high, as typically found in time series trend models and thus not surprising. The Durbin-Watson statistic reveals evidence of some autocorrelation in the residuals, especially for TFP. The Jarque-Bera test (of skewness and kurtosis matching a Normal distribution) suggests the hypothesis of Normality can be rejected at conventional levels except for the full-sample models of average speeds, but the chi-squared approximation is sensitive in small samples like this (about 110 observations) and susceptible to high rates of Type I errors. *These results suggest the deterministic models may not be ideal due to serial correlation and lack of sufficient observations to draw firm conclusions about normality. Correcting for these issues in estimation or using more advanced diagnostics may alter the results*

Table 33: Deterministic Trend Model Results

	Avg. Race Speed				Avg. Qual Speed				Total Factor Productivity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
T	1.67***	1.73 ***	1.70***	1.67***	1.71***	1.6***	1.55***	1.54***	1.71***	1.71***	1.71***	1.71***
	(0.22)	(0.27)	(0.27)	(0.29)	(0.11)	(0.11)	(0.11)	(0.10)	(0.16)	(0.13)	(0.12)	(0.12)
$D_1 \times T$	1.63	1.79	1.70	1.88	1.63*	1.41	1.32	1.31	-0.05	-0.05	-0.05	-0.05
	(1.04)	(1.40)	(1.47)	(1.59)	(0.80)	(0.85)	(0.85)	(0.80)	(1.21)	(1.01)	(0.95)	(0.89)
$D_2 \times T$	-0.48*	-0.45	-0.40	-0.38	-0.43**	-0.17	-0.14	-0.13	0.42	0.32	0.32*	0.32*
	(0.23)	(0.26)	(0.27)	(0.29)	(0.13)	(0.12)	(0.12)	(0.11)	(0.22)	(0.16)	(0.16)	(0.15)
$D_3 \times T$		-1.45***	-1.41***	-1.37***		-0.50***	-0.44**	-0.44**		-0.69***	-0.70***	-0.70***
		(0.30)	(0.31)	(0.33)		(0.14)	(0.14)	(0.13)		(0.19)	(0.19)	(0.18)
$D_4 \times T$			-1.38**	-1.3152**			-1.48***	-1.48***			0.01	0.01
			(0.45)	(0.00)			(0.24)	(0.22)			(0.35)	(0.33)
$D_5 \times T$				-1.26**				-1.32***				-0.92**
				(0.39)				(0.17)				(0.30)
$Prec_t$	-1.43	-5.56**	-5.60***	-5.28**								
	(3.92)	(1.91)	(1.55)	(1.68)								
Inc_t	-68.09	-172.98***	-157.45***	-175.145***								
	(37.92)	(38.71)	(38.12)	(38.36)								
$Spread_t$	2.79*	1.78	1.64	1.97								
	(1.28)	(1.39)	(1.44)	(1.51)								
$Temp_t$	-5.31	-4.92	-0.97	-2.35								
	(4.28)	(4.37)	(4.29)	(4.43)								
Exp_t	-0.71	-2.34	-1.45	-1.41	-4.01**	-2.02	-1.18	-1.10				
	(2.27)	(2.63)	(2.65)	(2.70)	(1.42)	(1.35)	(1.30)	(1.18)				
$Field_t$	-12.77*	-10.44	-9.76	-9.83								
	(5.42)	(7.00)	(7.29)	(7.84)								
$Rider_t$	-1.50	-1.20	-1.32	-1.51	-4.03***	-4.11***	-4.14***	-4.14***				
	(1.52)	(1.96)	(2.04)	(2.21)	(1.03)	(1.10)	(1.10)	(1.03)				
$Pole_t$	-0.21	-0.26	-0.15	0.31								
	(0.45)	(0.50)	(0.49)	(0.49)								
Adj. R^2	0.98	0.97	0.97	0.96	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99
End Date	1969	1995	2006	2021	1969	1995	2006	2021	1969	1995	2006	2021
Durbin-Watson	1.67	1.85	2.14	1.97	1.97	1.44	1.47	1.46	0.75	0.75	0.75	0.75
Jarque-Bera	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00

Note: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. All coefficients are multiplied ×100.

Overall, results of the deterministic model regressions offer a mixed view on the central research question motivating this paper. Result #1 indicates that speeds and TFP share a common trend in the long run, but result #4 casts doubt on that conclusion because the speeds and TFP do not appear to clearly and consistently share common trends over shorter subsamples defined by the five predetermined trend breaks. To formalize the latter perspective, we jointly estimated the speed and TFP models using seeming unrelated regression estimate (SURE) and tested the null hypothesis of equality for each speed-TFP pair of trend coefficients in the long run and each subsample.

Table 34 reports the p-values for each pairwise test, with the column headings denoting the two regression models from Table 33 whose coefficients were compared. The first four columns compare the two speeds, the other columns compare the two speeds with TFP. Not surprisingly, equality of the long-run speed and TFP coefficients (on T) clearly is not rejected. In the full-sample models (4,8), equality of race and qualifying speed trends cannot be rejected for any period except the Productivity Slowdown (D_3). However, equality of each speed's trend with the TFP trend is rejected or marginally accepted for most subsample periods. Rejection of equality is more common for average speed (three of five coefficients) than for qualifying speed (only D_4, although two subsamples are rejected at the 12-13 percent level).

Table 34: SURE Tests of Trend Coefficient Equality

Paired Outcomes	s_t^r, s_t^q	s_t^r, tfp_t	s_t^q, tfp_t
Coefficient/Model	(1), (9)	(4), (12)	(8), (12)
T	0.62	0.90	0.53
$D_1 \times T$	0.50	0.26	0.32
$D_2 \times T$	0.29	0.00	0.12
$D_3 \times T$	0.00	0.06	0.32
$D_4 \times T$	0.68	0.00	0.00
$D_5 \times T$	0.87	0.42	0.13

The estimates in Table 33 reflect a distinct temporal pattern in deterministic trend breaks that can be seen clearly in Figure 19. This figure plots the speed and TFP trend estimates (star symbols) and standards errors (horizontal lines) for the full sample and each subsample. The first row depicts the clear equality of the long-run trend coefficients. The remaining rows show the volatility and heterogeneity of deterministic trend estimates over different time periods. The point estimates over time show little consistency, even when accounting for sampling error. Instead, each subsample contains a different relationship between the deterministic trends. Notably, average speeds clearly reflect the 1973 Productivity Slowdown but only TFP reflects the New Economy and Dot-Com productivity booms.

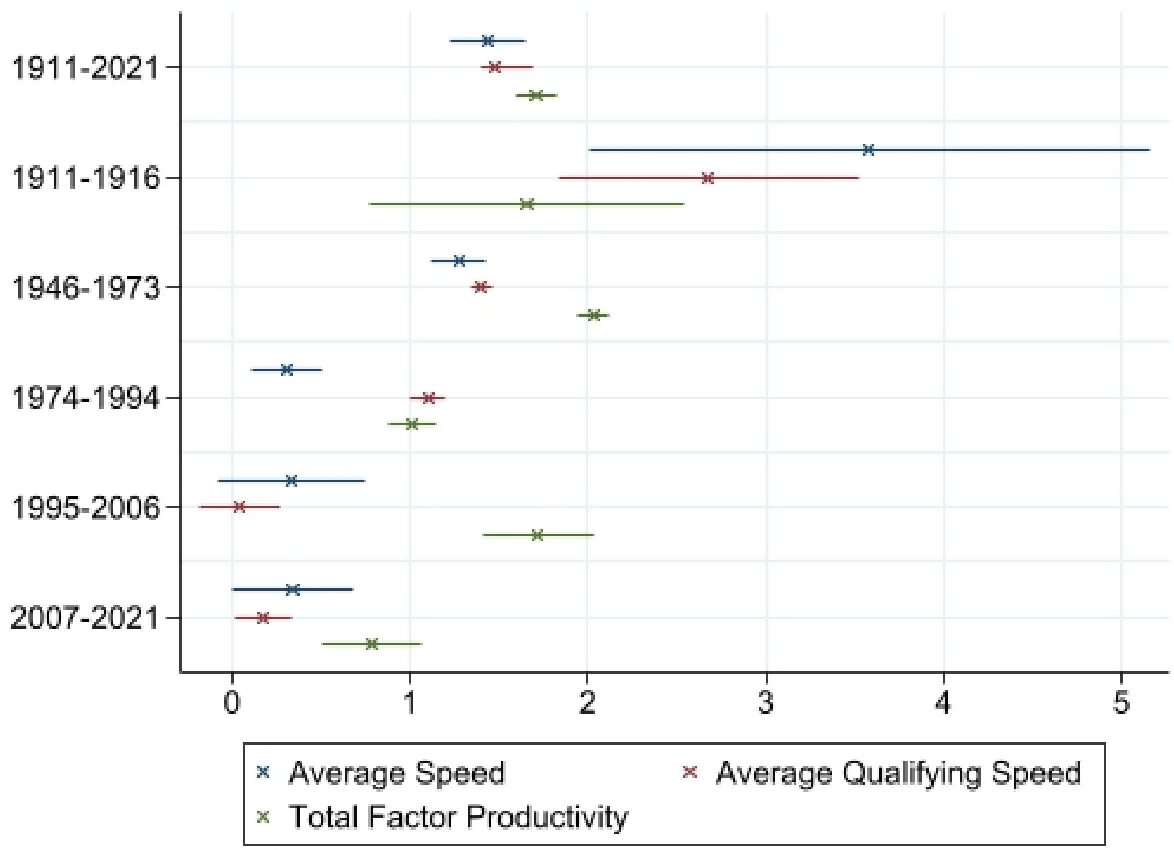


Figure 19: Comparison of Subsample Deterministic Trends

In summary, these results reflect five predetermined, distinct trend periods identified in the literature on labor productivity since World War II. This approach appropriate under two strong assumptions: 1) TFP break points are known *a priori* with certainty; and 2) if auto racing speeds have their own trend breaks, they occur at exactly the same as those in TFP. The second condition could be true for deterministic trends by random chance (highly unlikely) or if TFP and speeds share a common *stochastic* trend. The latter is likely to appear in the data as if TFP and speeds have approximately the same deterministic trends.

2.4.3 Endogenous Break Points

This subsection relaxes the strong assumptions underlying the deterministic trend models in two ways. First, deterministic trends in auto racing speeds are estimated separately from trends in TFP. This changes allows the data to reveal rather than conform to trends, and helps determine whether the breaks in TFP and speeds align independently rather than being forced to be the same. Second, the number of trends and dates of trend breaks are determined endogenously for both TFP and speeds, rather than predetermined exogenously. This change allows the most recent data to reveal the optimal number and location of trend breaks for each variable without imposing any covariance structure.[41]

Figure 20 portrays the results of endogenous breakpoint estimation with unknown number and dates of breaks in the two speeds and TFP for two samples.[42] The first column shows the full sample (1911-2019) and the second column the sample ending before the Dot-Com/New Economy boom that temporarily increased productivity growth. Black symbols denote estimated breakpoints, and color segments illustrate estimated trend ranges. Comparing graphs within a column (across the three rows) reveals the extent to which estimated trend breaks and ranges align across the three variables.

[41] Christiano (1992) first argued for endogenizing break points but acknowledged a tradeoff between precision of estimated break dates and gains in identification of the true number of actual breaks. See also B. E. Hansen (2001) for more recent advances and arguments for endogenous breakpoint estimation, especially the ability to distinguish estimated break dates from exogenous processes and random walks. However, Stock (1994) and D. M. Byrne et al. (2016) stress the limitations to endogenizing both the number and location of breaks, highlighting the uncertainty and imprecision of confidence intervals associated with endogenous breakpoint tests. Trend-break tests also rely on the presence of unit roots (J. P. Byrne et al., 2006), detection of which has low power in small samples. Addressing these limitations, Bai & Perron (1998) and Bai & Perron (2003) acknowledge that "true" simultaneity of estimating the number and location of breaks is unrealistic, so breaks should be tested sequentially.

[42] We use the STATA "xtbreak" command.

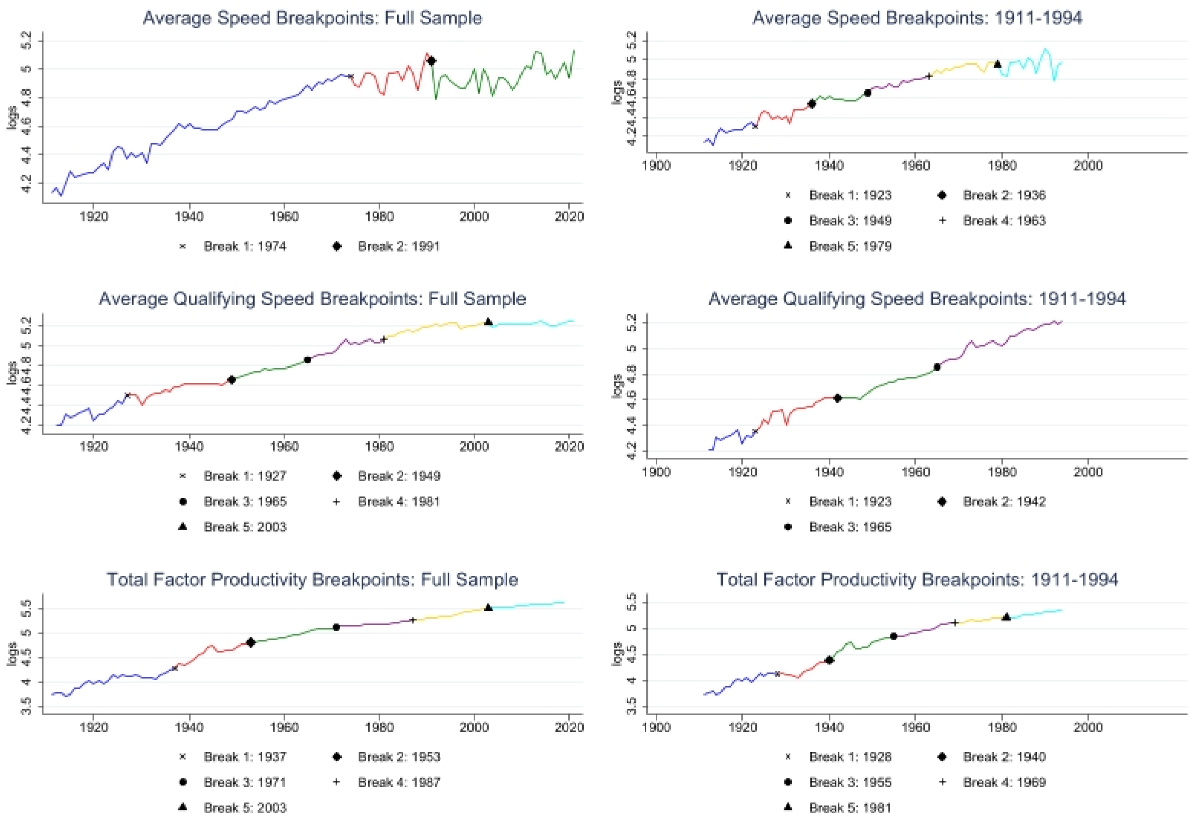


Figure 20: Endogenous Breakpoint Tests: Unknown Number of Breaks

Inspection of Figure 20 reveals that endogenously estimated breakpoints are not consistent across variables or sample periods. Both speeds have a different number of estimated breaks across samples, and wide differences in break years. TFP has five estimated breakpoints in both samples, but the break years vary almost as widely across samples. Most importantly for the research question of this paper, the breaks in speeds do not align consistently with TFP for either sample. In the full sample, qualifying speed also have five breaks that are less than 10 years different from the TFP breaks, but race speed only has three breaks in entirely different years. In the shorter sample, race speeds have five breaks that are roughly aligned with TFP, but qualifying speeds only have four breaks in markedly different years.

Results are more consistent when the number of breaks (but not year) is fixed *a priori*, as shown in Figure 21 for the full sample only. Here columns represent the number of endogenous breaks (2-5). The estimated endogenous breakpoints and subsamples line up across the three variables better as the number of predetermined breakpoints increases. With only two breakpoints, the estimates range almost a half century (1927-1974 for the first and 1963-1991 for the second). With five breakpoints, however, the majority of estimated breakpoints are within a decade of each other across all three variables. Although rarely an exact year match, the five endogenous breakpoints align similarly across variables.

Figure 21: Endogenous Break Point Tests: Fixed Number of Breaks

To summarize this subsection, endogenous breakpoint estimates provide modest—but far from conclusive—support for the view that the speeds may share a common trend with TFP. When allowed to "speak freely," the data show evidence of very roughly similar numbers and dates of trend breaks within the full sample. However, the large magnitude of discrepancies in the number and date of breaks suggests that trends may not be deterministic but rather stochastic.

2.5 Stochastic Trends

In the 1980's, macroeconomists questioned the validity of deterministic trend models for aggregate variables, such as GDP and TFP, and began specifying models with stochastic trends stemming from a unit root(s).[43] Variables that share a common stochastic trend are cointegrated (CI) with dynamic error-correction to the common trend, as in Engle & Granger (1987). This section implements a contemporary vector error-

[43] See Nelson & Plosser (1982), Campbell & Mankiw (1987), and Durlauf & Phillips (1988).

correction model (VECM) that assumes cointegration between auto racing speed and TFP, and is consistent with a benchmark stochastic growth real business cycle (RBC) model of auto racing. While a "deeper" theoretical model may well be needed, it is informative to first assess whether speed and TFP share a common stochastic trend and error correction.[44]

2.5.1 Motivational Model

Consider the following partial-equilibrium stochastic growth real business cycle (RBC) model (Kydland & Prescott, 1982; Long Jr & Plosser, 1983) applied to the auto racing industry (subscript i).[45] The planner chooses consumption of auto racing, c_{it}, and beginning-of-period effective capital stock, $k_{it} = K_t/L_t$, to solve:

$$\max_{c_{it},k_{i,t+1}} \sum_{t=0}^{\infty} \beta^t U(c_{it}) \tag{2.3}$$

Subject to:

$$y_{it} = A_{it} f(k_{it}) \tag{2.4}$$

$$y_{it} = c_{it} + \Delta k_{i,t+1} - \delta k_{it} \tag{2.5}$$

$$A_{it} = \gamma_0 + \gamma A_t \tag{2.6}$$

$$\Delta A_t = \varepsilon_t \tag{2.7}$$

Where β is the discount factor, and $0 < \delta < 1$ is the depreciation rate. The representative auto racing team produces speed as defined earlier, $y_{it} = Y_{it}/L_{it} = P_{it}^{-1}$, where Y_{it} is the distance traveled by the representative race car and $L_{it} = (1/\tau_{it})$ is the inverse of elapsed time of the representative driver (τ_{it}).[46] Output of speed also includes industry TFP, A_{it}. Based on the deterministic trend results linking y_{it} and A_t in the long run, equation (2.6) specifies industry TFP as linear function of aggregate TFP, A_t, with expected $\gamma \approx 1/2$.[47] The unit root in aggregate TFP in equation (2.7) introduces a stochastic trend.

[44] The central hypothesis in the literature is that technological progress somehow is manifest in the ongoing improvements of athletic outcomes measured by speed (or time). Knowing the precise theoretical mechanism by which technological progress diffuses into auto racing speeds is a crucial step in testing this hypothesis. Although basic, the structural model is a conventional macroeconomic rationalization of a common stochastic trend between speed and TFP.

[45] This motivational approach also is guided by the surveys of Hahn & Matthews (1964), Fagerberg (1994), Jorgenson (1991) and Kennedy & Thirlwall (1972).

[46] Y_{it} and L_{it} vary by event: the Indy race is 500 miles and Indy qualifier is 2.5 miles; the NHRA race is .25 miles.

[47] Admittedly, this initial assumption is strong. See Foster et al. (2001), Foerster et al. (2019), Dosi & Nelson (2010) for discussions of the relationship between aggregate versus industry-specific technological change. See Basu et al. (2006) for an alternative approach to jointly modeling aggregate and industry-specific productivity. Note, however, the consumption share of spectator sports was only 0.2 percent in 2019. Thus, the share of auto racing (and especially one Indy 500 race) is a tiny fraction of GDP and A_{it} probably has little or no aggregate implications.

Although the simple benchmark model motivates the VECM estimation, two limitations merit brief discussion. First, the full model requires a more complete specification of demand for consumption of auto racing events. Most likely, consumers get utility from more than just the pure speed of the race cars and value other elements of auto racing entertainment more generally.[48] Second, growth likely involves endogenous diffusion of technological change embodied in capital (race cars), seen as vintage capital differences across cars within and between years. LBD effects likely occur as well, but they are not modeled explicitly given their relatively small contribution to growth inferred in Section 2.3.1. Expanding the model to include these dimensions is a potentially fruitful direction for future research.

2.5.2 VECM Specification

Let $Z_t^k = [s_t^k, a_t]'$ denote the bivariate vector motivated by the RBC model, now with $k = \{r, q, d\}$ where d is NHRA drag race. Lowercase variables continue to denote log levels. The stochastic trend makes $s_t^k, a_t \sim I(1)$. If speed and TFP share a common trend, they are cointegrated:

$$s_t^k = \alpha + \beta a_t + \varepsilon_t^k \ , \tag{2.8}$$

Where the cointegrating residual is $\varepsilon_t \sim I(0)$. The stationary first differences (growth rates) are jointly determined by a VECM:

$$\Delta Z_t^k = K_k + \Pi_k Z_{t-1}^k + \sum_{i=l}^{L} \Gamma_{k,i} \Delta Z_{t-l}^k + \eta_t^k \tag{2.9}$$

For $k = \{r, q, d\}$. Equation (2.8) can be estimated in a separate first-step regression with OLS or as part of the simultaneous system described in equation (2.9) using the more efficient method of Johansen (1988, 1991, 1995).

The reduced-rank (r) matrix $\Pi_k = \alpha\beta'$ contains the cointegrating vector(s) β_k, and adjustment speed(s) α_k and defines the $I(1)$ model $H(r)$. The rank of Π_k is the number of cointegrating vectors, which the structural model predicts is $r = 1$:

$$\alpha_k \beta_k' Z_t^k = \begin{bmatrix} \alpha_{s^k} \\ \alpha_a \end{bmatrix} [s_t^k - \beta_k a_t] = \alpha \varepsilon_t^k$$

Furthermore, the structural model tells us that ε_t^k is the CI error from equation (2.8). If the system exhibits dynamic correction to a common trend, $\alpha_i \leqslant 0$ for $i = \{s, a\}$, at least one adjustment speed must

[48] See H. Hansen & Gauthier (1989) and García & Rodríguez (2002) for exposition on determinants of sports viewership, attendance and demand for competitiveness of sports outcomes.

be strictly negative.

2.5.3 Stochastic Trend Results

Full sample estimation results for the VECM models of speed and TFP are reported herein. Pre-estimation tests for stationarity confirm that all variables are $I(1)$ in log levels, and $I(0)$ in growth rates at conventional levels of significance, as required (see appendix for details). Information criteria tests (unreported) for the Johansen system with annual data in log levels generally suggest an optimal lag length of one (hence $L = 0$ lagged growth rates), but no more than four lagged levels ($L = 3$ lagged growth rates).

On balance, the data generally support the hypothesis of a single cointegrating vector. The Johansen rank tests in Table 35 reject the null hypothesis of no cointegrating vector ($r = 0$) at the 5 percent level or better for the Z_t^r and Z_t^d models.[49] The absence of a cointegrating vector ($r = 0$) for the Z_t^q model cannot be rejected with the same high confidence, but is close to rejection at the 10 percent level (not reported). Because the sequential test for $r \geqslant 1$ can be rejected at the 5 percent level, cointegration is less likely for the Z_t^q model by this measure. However, residuals from two-step estimation of equation (2.8) test $I(0)$ at the 10 percent level (or better) for all three models, and the ADF statistics are nearly identical to those from the VECM estimated residuals.

Table 35: Cointegration Test Results

Johansen Rank					ADF		
			Critical Values			t-stat. (p-value)	
Vector	H_0	Trace stat.	1%	5%	Error	VECM	Two-Step
Z_t^r	$r = 0$	29.31**	20.04	15.41	$\widehat{\varepsilon}_t^{r,a}$	−5.18	−5.18
	$r \geqslant 1$	4.40*	6.65	3.76		(< .000)	(< .000)
Z_t^q	$r = 0$	12.02	20.04	15.41	$\widehat{\varepsilon}_t^{q,a}$	−2.62	−2.65
	$r \geqslant 1$	5.02*	6.65	3.76		(.089)	(.083)
Z_t^d	$r = 0$	18.56*	20.04	15.41	$\widehat{\varepsilon}_t^{d,a}$	−3.73	−3.79
	$r \geqslant 1$	4.43*	6.65	3.76		(.004)	(.003)

Note: * and ** indicate rejection of H_0 at the 5% and 1% levels, respectively.

VECM coefficient estimates for all three models are correctly signed and mostly statistically significant, as shown in Table 36. As expected, the cointegrating coefficients ($\widehat{\beta^k}$) are highly significant and range from $0.38 - 0.53$, consistent with the relative long-run growth rates of speed and TFP.

The estimated adjustment speeds ($\widehat{\alpha^k}$) reveal an interesting asymmetry. Adjustment speeds for the auto races are negative (as expected) and statistically significant, but the adjustment speeds for TFP are close to zero and statistically insignificant. Thus, the data indicate that the levels of auto racing speeds adjust to the level of TFP but not vice versa. The absolute values of the adjustment speeds coefficients are economically

[49] We use the "vecrank" command in STATA. The results are qualitatively the same when estimating the model with two and four lagged levels.

large as well. The Indy 500 race and NHRA drag race adjustment speeds are about 0.4 per year, but the Indy qualifying adjustment speed is 0.1, only about one-fourth as large.

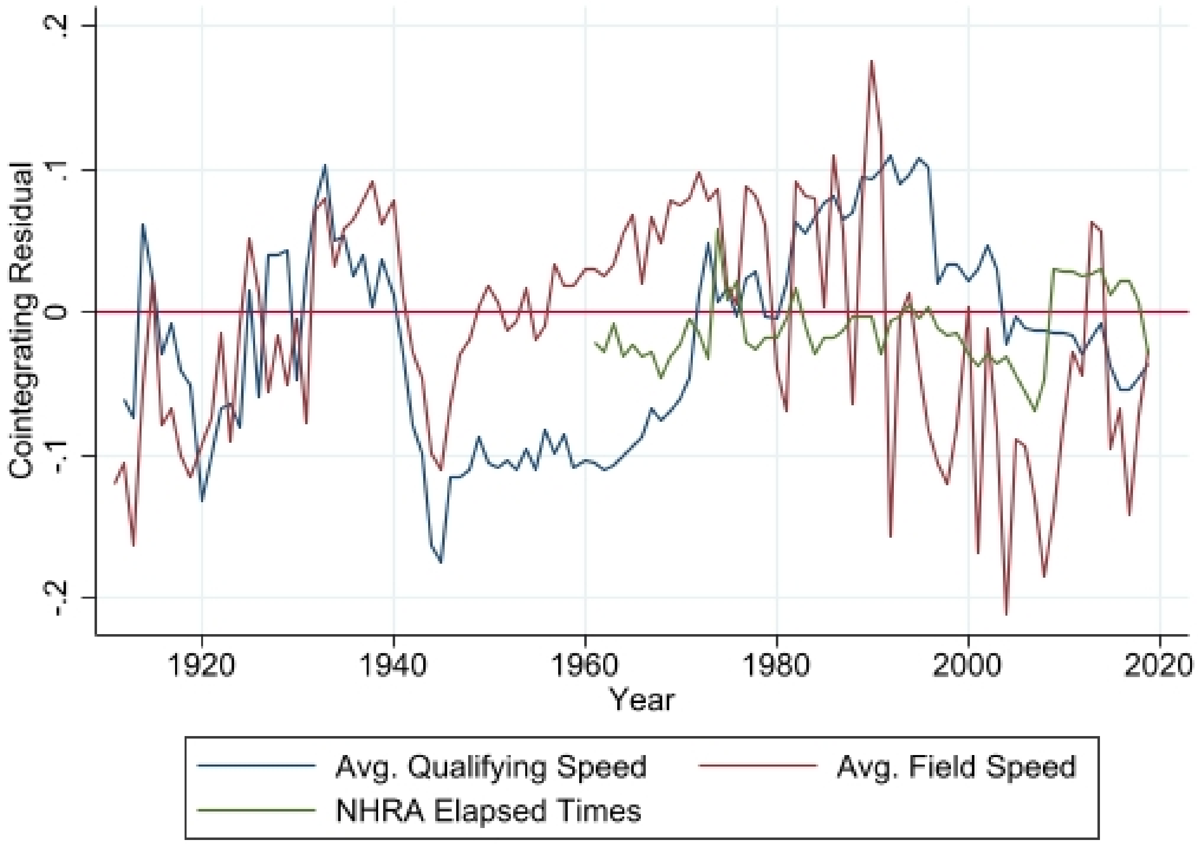
Table 36: VECM Model Coefficient Estimates ($L = 0$)

	Δs_t^r	Δa_t	Δs_t^q	Δa_t	Δs_t^d	Δa_t
K	0.004	0.019***	0.008**	0.018***	0.000	0.012***
	(0.006)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)
α	−0.360***	0.082	−0.099*	0.042	−0.398***	0.008
	(0.075)	(0.044)	(0.041)	(0.050)	(0.102)	(0.061)
β_{VECM}	0.428***		0.534***		0.380***	
	(0.000)		(0.000)		(0.000)	
β_{OLS}	0.427***		0.556***		0.404***	
	(0.000)		(0.000)		(0.000)	
R^2	0.193	0.219	0.140	0.196	0.249	0.545
$D.W.$	2.12		2.43		2.13	
N	108		107		58	

Note: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Figure 22 plots the estimated cointegration residuals from the VEC models. The two race speeds (Indy 500 and NHRA drag) fluctuate around zero relatively frequently with little occurrence of persistent deviations. Although Indy 500 qualifying speeds may be marginally cointegrated with TFP, they exhibit two lengthy trend deviations in the 1940's-1960's and 1980's-1990's. Indy 500 race speeds exhibit similar but less persistent deviations, such as the 1960's-1970's and 1990's-2010's. Persistent deviations likely reflect periods when the bivariate VEC models are not capturing important factors affecting either the trends or adjustment to them. For examples, rule changes affecting adoption of embodied technology or governing the conduct of actual races and qualifying events may help explain these periods, as argued in Mantel Jr et al. (1995).

Figure 22: Estimated Cointegration Residuals from VECM Models



The results in this section provide thought-provoking support for the stochastic trend model of auto racing speeds and TFP. Consistent with a basic stochastic RBC growth model, there is one common trend driven by a unit root in TFP. The asymmetric adjustment patterns evident in the estimated VEC models (speeds adjust to TFP but not vice versa) also support the simplifying model assumption that industry TFP (speeds) are linearly proportional to aggregate TFP in the long run. We also take time to acknowledge that bivariate stochastic trend models can suffer omitted variable bias. In this setting in particular, it is not unreasonable to also think that wages or some proxy for the marginal product of labor of the drivers could also be cointegrated with both TFP and race speeds.

We have considered this possibility by testing for cointegration across trivariate systems that also included the average level of prize support the Indy 500 contest offers in year t, or the prize spread between first and second place (consistent with tournament theory). While not reported, Johansen tests on trivariate systems with prize earnings do not produce evidence of a rank $\geqslant 1$. Furthermore, the prize level as a proxy for wages tests as I(0) making its inclusion in VECM models precarious, and in violation of the necessary model assumptions.[50] Reasons for these results could stem from the possibility that driver wages are not dependent on the potential prize support for the race itself. Today, drivers negotiate compensation with their race team, and also get considerable compensation from sponsorships. Given limited historical data availability on these driver-specific wages, it becomes difficult to build a fair wage that could otherwise be included in a higher-dimension VECM system.

2.6 Conclusion

Based on a long time series of annual data, econometric results from deterministic and stochastic trend models show that Indy 500 speeds and US TFP share a common long-run trend since 1911. The results represent important evidence of a link between a micro-level factor driving productivity growth and macro-level outcomes. We empirically identify a link between micro and macro level outcomes linked to productivity growth. This confirms results in a large body of theoretical research on LBD and TFP. The magnitudes of trends in auto racing speeds and technological process vary over time but do not align well with subsamples of the deterministic trend model. Thus, the stochastic trend model appears to fit the data better. This finding aligns with standard RBC model specifications of the dissemination of technical progress and trends in long-run output growth.

Because this paper takes a relatively simple, aggregate (macroeconomic) approach, future research may yield even more compelling and enlightening evidence on the link between technical progress and auto racing speeds. Three extensions merit consideration. First, developing a more microfounded, and general equilibrium theoretical framework is likely to produce better understanding of the precise transmission mechanism(s) from technology and productivity to speed. Allowing for industry-specific technical progress and introducing explicit demand for speed and auto racing entertainment are important in this regard. Second, using disaggregated data on individual drivers, cars, and racing teams would provide more precision in the estimation of model parameters and an opportunity to leverage heterogeneity in model identification. Third, including more explicit measures of changes in rules governing the allowable embodiment of technology in race cars and in the conduct of races (including qualification) is necessary to distinguish them from the diffusion of frontier technology to speed.

[50] Results available upon request.

2.7 Appendix A: Notes on Data

Indianapolis 500 Racing The Indianapolis 500 operator, IndyCar, is the latest in a series of open wheel automobile race sanctioning bodies. The automobiles competing in races sanctioned by IndyCar are, not surprisingly, called Indy cars. These automobiles have open wheels and a single seat. The vehicles must conform to certain technical specifications in terms of drive train and body. Indy car races take place on both oval tracks and road-street courses. The Indianapolis 500 is the premier Indy car race in the world and attracts top teams and drivers from around the world. The race takes place a the Indianapolis Motor Speedway (IMS) in Speedway, Indiana, a suburb of Indianapolis. The race traditionally occurs held over Memorial Day weekend. The track is a 2.5 mile oval with four turns. Prior research analyzed outcomes in the Indianapolis 500 in the context of productivity growth. The race occurs annually and race–day data are available from the Indianapolis Motor Speedway's historical data archives from the early 20^{th} century to the present. These data include the average speed of the winning car, qualifying results by driver by race for all drivers who qualified for the Indy 500, as well as tertiary data on race conditions such as on–track collisions, caution flags, and driver–specific statuses during each race. Finally, the Indianapolis Motor Speedway also publishes data on the prize spread in dollars for all races. We collected the full series for both the Indy 500 race (1911–2021) and its qualifying events (1912–2021).

Indianapolis Motor Speedway does not collect race-day whether conditions. We augment the race outcome data with data from the National Oceanic and Atmospheric Administration (NOAA). We collect data on temperature and precipitation from the Anderson Sewage Plant monitoring station, which contains weather data back to the first Indy 500 race in 1911.

NHRA Drag Racing We also collect data from another automobile racing event, the National Hot Rod Association (NHRA) Winternationals contest, held annually in February in California since the 1960s. Drag racing differs dramatically from open wheel IndyCar racing. Drag racing involves highly specialized automobiles competing on a short (0.25 mile) straight track. We the automobile technology affecting drag racing outcomes could differ substantially and should showcase all technological progress in drag racing made from the previous year. Its significance in drag racing is comparable to the Indy 500 for Indy Car racing. The National Hot Rod Association (NHRA) publishes Winternationals results for a little less than two decades worth of observations, thus, to get the full time series for the entirety of the event, we turn towards identifying news excerpts and articles via ProQuest that document the winner's top speed and elapsed time. Despite this, we have three years in the sample that require linear interpolation. The full sample extends from 1961–2021.

Aggregate Productivity To find measures of productivity that date back to 1911, we use historical productivity data from the Long Term Productivity Database (LTPD). While this data is not the traditional productivity measurements that would be reported by the Penn World Tables (PWT) or the Bureau of Labor Statistics (BLS), a simple correlation of the annual growth rate of total factor productivity from both the LTPD and PWT uncover a correlation of roughly 0.91, which is strong enough for us to justify its usage in our study.

Other Data For race day controls such as precipitation, we find leverage data from the National Oceanic and Atmospheric Administration (NOAA). Precipitation levels are coded for the specific race day and measured in terms of inches. "Trace" levels of precipitation are interpolated as 0.05 inches, which is the midpoint between zero and the lowest reported levels of 0.10 inches.

2.8 Appendix B: Measuring Race Outcomes

The literature has focused on two measures of race outcomes (time or speed): 1) the winning outcome (specific year or event); or 2) the new world record (best outcome of all time in an event). Both measures call for econometric specifications designed to fit data generated by extreme value distributions rather than the normal distribution used in OLS estimation of basic linear econometric models. However, aggregate data like TFP represent *average* values (or sums) across all agents rather than an extreme value. For this reason, we measure Indy 500 speeds (race and qualifying) as the average across all participating drivers that year to better match TFP.[51] However, the NHRA drag race data are harder to collect and thus represent only the winning times for the Winternationals event.

Central-moment and extreme-value measures of measures of race outcomes do not exhibit quantitatively large differences in long-run trends, as shown in Figure 23. The top row shows that the levels of extreme (winning car) and central-moment (field average) speeds are very similar. Most importantly, the long-run trends in the winning car and average speeds show little economically significant differences, particularly in the second half of the sample. The bottom row shows there are non-trivial differences between the speed measures early in the sample, but the differences decline over time. The difference eventually become negligible for race speeds, and consistently trends downward for qualifying speeds.

[51] As noted earlier, measures of Indy 500 race and qualifying speeds are not exactly comparable. Data for race speeds are available only for the drivers still running when the winner crosses the finish line, which is often less than half the field and thus a truncated mean. In contrast, qualifying speeds are available for essentially all drivers and thus represent a true central moment.

Figure 23: Extreme and Central Moment Measures of Indy 500 Speeds

2.9 Appendix C: Learning-By-Doing Models

Both Barzel (1972) and Fellner (1969) motivate their studies under the pretense that learning-by-doing (LBD) drives aggregate technical change. Their model specifications advocate that the passage of time alone is sufficient for capturing and quantifying LBD over time. However, more modern approaches of estimating LBD can be leveraged that allow us to capitalize on the increased availability of Indy 500 race results data, including the names of the drivers participating each year. LBD would tell us that increasing race speeds over time can partially be explained by the stock of driver experience. Looking across the field of all drivers i in years $t = 1911$ through $t = 2021$, the distribution of experience can be described by Figure 24.

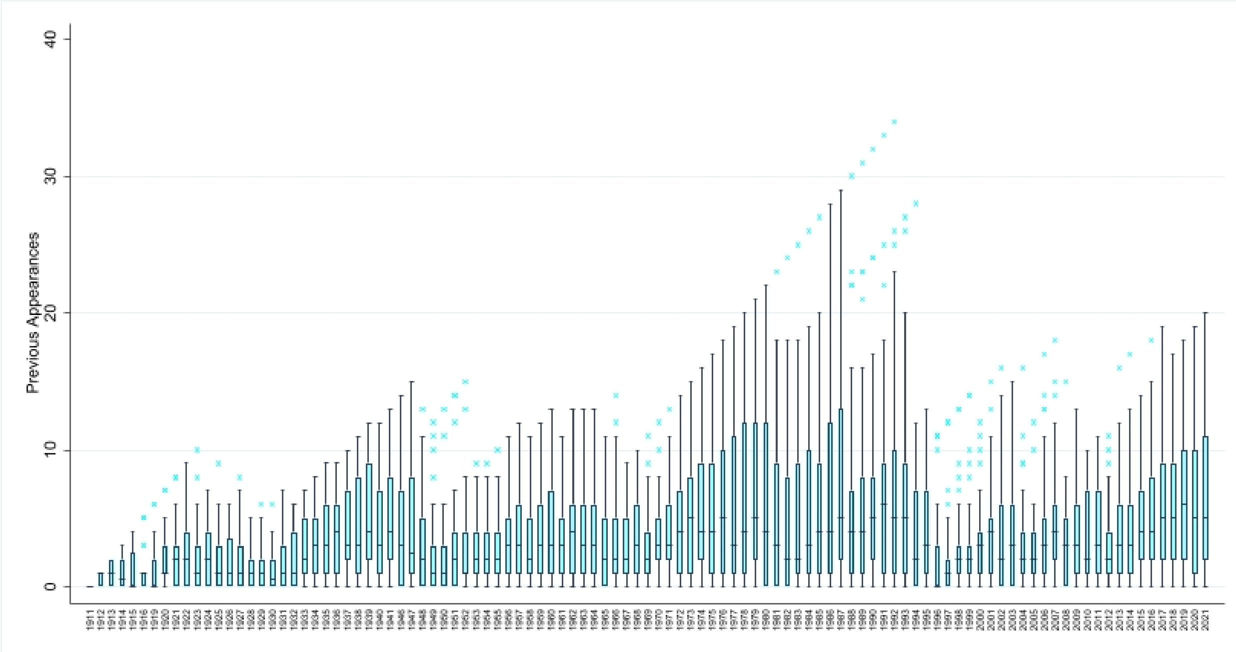


Figure 24: Distribution of Accumulated Driver Experience Over Time

At a microeconomic level, we observe that the upper quartile and outlier drivers' experience *appear* to be increasing over time; however, the median experience of the field in year t seems to hover between zero and five years. In order to estimate the effects of LBD using driver experience in an aggregate setting, we need a model that links experience to macroeconomic productivity (or an appropriate proxy such as race speeds) as well as a strategy for aggregating experience of all i drivers in year t to adequately express aggregate experience.

Following the form of Levitt et al. (2013), we assume that learning at a microeconomic level is akin to an inventory that is stored and accumulated over time. Thus, in the context of the Indy 500, E_{it} captures the experience of driver i in year t as the count of all previous appearances in the Indy 500 that driver i has achieved up to year t. The average of E_i across the field of qualified racers gives us E_t, which captures the average level of accumulated experience of the field in year t. The underlying structure to capture LBD is expressed by equation (2.10).

$$A_t = BE_t^{\gamma} \tag{2.10}$$

Where A_t is the level of productivity at time t, E_t is the average level of accumulated experience, and γ is effectively a learning rate. In the study undertaken by Levitt et al. (2013), γ is implied to be negative as productivity in the manufacturing setting of their study is captured by the number of defects identified in the output good observed. Their hypothesis is that as E_t increases, defects should fall. In the context of our model, productivity is reflected in the outcome variables of the racing events (speeds or times). If the speed of the field is our outcome variable of interest, we would hypothesize that $E_t > 0$, implying that cars get faster as time passes and drivers accumulate experience. By taking the natural logs of both sides of equation (2.10), we arrive at equation (2.11).

$$\log(A_t) = \log(B) + \gamma E_t \tag{2.11}$$

Adopting our previous notation, we estimate the following LBD-type models described by equation (2.12).

$$s_t^k = a_{s0} + \gamma Exp_t[+\tau s_m(T)] \tag{2.12}$$

s_t^k has the same meaning described in our main results along with $Exp_t \equiv \log(E_t)$. Furthermore, we estimate our models with and without a time trend, T. Our time trend is generalized as an unknown function $s_m(\cdot)$. For the purpose of this study, we allow $s_m(\cdot)$ two take on two explicit function forms: $s_1(\cdot) = T$ and $s_2(\cdot) = \log(T)$.[52] Table 37 describes our estimation results.

Table 37: LBD Models of Race Performance

	Race Speeds			Qual Speeds		
	s_t^r	s_t^r	s_t^r	s_t^q	s_t^q	s_t^q
a_{s0}	4.278***	4.237***	3.717***	4.310***	4.243***	3.503***
	(0.039)	(0.021)	(0.038)	(0.056)	(0.015)	(0.053)
Exp_t	0.371***	0.150***	0.046	0.434***	0.072***	-0.033
	(0.029)	(0.021)	(0.024)	(0.041)	(0.015)	(0.033)
T		0.006***			0.009***	
		(0.000)			(0.000)	
$\log(T)$			0.257***			0.369***
			(0.015)			(0.021)
Adj. R^2	0.600	0.884	0.897	0.500	0.963	0.874

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Consistent with our predictions, Exp_t is positive and statistically significant in almost all specifications of equation (2.12) for both race speeds and qualifying speeds, however, the magnitude of experience and its effect on race outcomes varies considerably when accounting for the passage of time. In the case of including a linear time trend, T, the magnitudes of Exp_t diminish substantially. In the case of our log-log models, experience becomes indistinguishable from zero in terms of overall significance.

We also observe that failing to account for the passage of time considerably harms the overall fit of our estimated equations. We note that the fit of both log-log and semi-log time trends for race speeds are

[52] This allows for the possibility of log-log or semi-log passage of time to be flexibly accounted for.

an improvement in fit over models with their absence. Furthermore, in most cases, the difference between semi-log and log-log fits are negligible with the exceptions of models where qualifying speeds are our outcome variable of interest, wherein the semi-log specification produces a stronger fit by a margin of roughly 0.10, once more echoing the findings of both Barzel (1972) and Fellner (1969).

Overall, while our simple LBD exercises provide some evidence that accounting for experience is important for understanding the evolution of productivity and dissemination of technical progress in race outcomes, it is far from only significant determinant. As stressed in Mantel Jr et al. (1995), the labor input to racing is usually equal to unity (with a few exceptions in the early years) while most true technical change accounting for improvements in speeds are likely embodied in the capital stock of the vehicle itself. This inherently limits the degree to which human experience in the aggregate can affect race outcomes. Thus, it is unsurprising that our time trends (regardless of functional form) are capturing unobserved embodied technical change in the vehicles themselves, thereby deflating the coefficient estimates of Exp_t in models that do not account for the passage time.

2.10 Appendix D: Labor Productivity Spline Models

Table 38: Labor Productivity Deterministic Trend Models

	Labor Productivity			
	(A.1)	(A.2)	(A.3)	(A.4)
T	0.024***	0.024***	0.024***	0.024***
	(0.001)	(0.001)	(0.001)	(0.001)
$D_1 \times T$	0.002	0.002	0.002	0.002
	(0.010)	(0.008)	(0.008)	(0.007)
$D_2 \times T$	-0.003	0.002	0.002	0.002
	(0.002)	(0.001)	(0.001)	(0.001)
$D_3 \times T$		-0.011***	-0.011***	-0.011***
		(0.002)	(0.002)	(0.001)
$D_4 \times T$			-0.001	-0.001
			(0.003)	(0.003)
$D_5 \times T$				-0.014***
				(0.002)
Adj. R^2	0.99	0.99	0.99	0.96
End Date	1969	1995	2006	2021
Durbin-Watson	0.75	1.49	1.49	1.48
Jarque-Bera	0.00	0.09	0.00	0.00

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

2.11 Appendix E: Time Series Diagnostics

Table 39 reports results of augmented Dickey-Fuller (ADF) tests for stationarity in speeds and TFP.

Table 39: ADF Tests for Stationarity

	Levels		Growth Rates	
Variable	T-Statistic	p-Value	T-Statistic	p-Value
s^r	-1.96	0.30	-13.79	< 0.001
s^q	-1.81	0.37	-13.56	< 0.001
s^d	-1.63	0.47	-10.94	< 0.001
a	-1.92	0.32	-10.76	< 0.001

Note: 5% critical values are -2.922 for s^d and -2.889 for s^r, s^q, a

Chapter 3 How Does Moral Hazard Impact Critical Market Banking Performance?

Corey J.M. Williams

When Alexander saw the breadth of his domain, he wept for there were no more worlds to conquer...

Hans Gruber (portrayed by Alan Rickman)
From *Die Hard* directed by John McTiernan (1988)

3.1 Introduction

The commercial banking industry is one of vital importance and interest for academics and policymakers alike. While characterizing the banking sector is difficult, one such philosophy is that banks act as intermediaries or service providers for firms and households through the generation of loans, securitization of debt, financial management services, and the provision of interest-bearing deposit accounts (Berger & Mester, 1997). The reliance of both firms and households on the health and performance of the banking sector cannot be understated, however, despite being of paramount importance, banking performance indicators are often arduous to measure and not well understood. Furthermore, identifying and investigating performance of too-big-to-fail banks, in particular, is also challenging.

In the economics literature, banking performance is typically characterized one of three ways: efficiency, productivity or returns-to-scale. Efficiency makes up the lion's share of the existing literature, particularly when attempting to project bank failure or risk exposure in light of the Global Financial Crisis (Barr et al., 1994; Berger & DeYoung, 1997; Nkusu, 2011; Berger & Bouwman, 2013; Makri et al., 2014). Productivity, while distinct, is not as widely researched as efficiency in part due to difficulties in defining production technology in the banking industry. Returns-to-scale, however, are of growing interest due to the ever-increasing size of banks concurrent with the shrinking competitive landscape. Economies of scale by definition describe some degree of cost savings achieved by increased levels of production. While researchers postulate that the growth in the average bank size is, in part, attributable to the desire for banks to capitalize on increasing returns-to-scale (Wheelock & Wilson, 2009, 2012, 2018), others, in particular policymakers, would argue that banks are growing *too large* and engage in unnecessary risks in part due to their size and scale.

Growing bank size has created worry among policymakers that some banks might become *too-big-to-fail* (TBTF) and expect policy intervention on behalf of governmental bodies for taking excess risk (Stern & Feldman, 2004; Stern, 2012). In a similar vein, policymakers and households may also worry about decreasing competition in the banking industry arising from consolidation, mergers, and acquisitions.

While the idea of banks becoming TBTF is not new, it has seen significantly renewed interest since the Global Financial Crisis, particularly in light of the recent failures of Silicon Valley Bank, First Republic Bank, Silvergate Bank, and Signature Bank—all of which were small-to-mid-size banks. Despite this, however, identifying moral hazard and risk appetite in the banking sector is no easy task. Beyond identification of moral hazard, investigating the performance of TBTF banks is also challenging. Furthermore, most studies that identify banks who may be TBTF do so via a bank's size, as measured by their total assets (Henderson et al., 2015; Wheelock & Wilson, 2018). However, works like Stern & Feldman (2004) stress that size, while important, is not the only factor that determines if a bank is potentially TBTF.

Despite competing definitions on TBTF, several questions remain largely unanswered. Firstly, *how does moral hazard impact banking performance*? Secondly, *what banks are capable of engaging in moral hazard set by bailout expectations*? Thirdly, *do banks who are TBTF stand to gain much from engaging in such behavior*? These questions are all interconnected and can be partially addressed simultaneously through recent advances in nonparametric econometric techniques. The results of this study suggest that the median critical market bank tends to enjoy slight increasing returns-to-scale, but over time, most critical market banks have exhausted economies of scale concurrent with the shrinking competitive landscape. We posit that banks capable of engaging in moral hazard are not necessarily just *large* banks, rather they are banks heavily involved in critical markets such as the markets for mortgage-backed securities, government securities, Treasury securities, or agency debt. These critical market banks are mostly large, but can include mid-size banks that also do considerable business in these riskier and highly integrated markets.

We attempt to partially answer these questions by leveraging a particularly appealing innovation in nonparametric econometrics from Henderson et al. (2015) who propose a novel estimator known as the semiparametric smooth coefficient seemingly unrelated regressions (SPSC SUR) model. This estimator extends the original Zellner (1962) model that is commonplace across various economics literatures, including the literature on banking performance by allowing model coefficients to vary smoothly by some environmental factor, z. Furthermore, the identification strategy described by Zellner (1962) has been leveraged heavily by production economists when estimating dual-cost (or production) functions in the absence of output prices wherein efficiency and theoretical consistency can be improved upon by estimating a set related cost-share equations resulting from the derivative conditions (a la Shephard's Lemma) of the same generalized cost (or production) function (Chambers et al., 1988; Kumbhakar, 1997; Geng & Sun, 2022).

The necessity of a dual-cost formulation arises due to idiosyncracies in the banking industry itself. Most notably, banking outputs like loans are in many ways services that accrue interest income over time, thus, they cannot be stored in the same sense that a production good or inventory can be (Henderson et al., 2015; Wheelock & Wilson, 2018). As a result, the specification of a dual-cost function is a standard approach in the literature. Such a specification necessitates the use of a system-wise estimation procedure to capture not only a firm's total cost minimization efforts, but also the conditional factor demand of their input quantities treating outputs and input prices as exogenous.

Contributions from Henderson et al. (2015) notwithstanding, the Zellner (1962) SUR model is, in general, more efficient for estimating simultaneous equations in settings where cross-equation correlation could exist. The Henderson et al. (2015) estimator provides a framework from which one can both estimate a firm's cost minimization efforts and flexibly capture moral hazard's impact on the production environment, meaning that any elasticity-based economies of scale estimates are in-and-of themselves functions of risk appetite. The flexibility that the SPSC SUR estimator provides makes it a superior choice for our analysis while retaining the traditional features and benefits that the Zellner (1962) SUR model also provides.

Finally, a key feature of system-wise estimation in the SUR model proposed by Zellner (1962) and expanded upon by Henderson et al. (2015) is the ability to impose cross-equation restrictions that emulate economic theory. A common metric used to assess performance of TBTF-type banks is economies of scale which is derived from a linear combination of reduced form coefficients typically estimated in a SUR system. Thus, the ability to implement cross-equation restrictions consistent with theory should improve the precision of returns-to-scale estimates.

Using a restricted translog SPSC SUR cost model, we estimate nonperformance-adjusted returns-to-scale for critical market banks from 2001 through 2021, where critical market banks are characterized by their relative involvement in markets for mortgage-backed securities, treasury securities, and agency and corporation securities. Over our full sample, elasticity-based returns-to-scale estimates suggest that the median critical market bank operates under modest increasing returns-to-scale, however, the majority of critical market banks operate under decreasing or constant returns-to-scale. The number of critical market banks operating under decreasing returns-to-scale, in particular, was exacerbated during the Global Financial Crisis and during Covid-19 pandemic suggesting that the shrinking competitive landscape over the past two decades has exhausted scale economies for most critical market banks. In a sense, with opportunities for mergers, acquisitions and consolidation waning as the competitive environment shrinks, fewer TBTF banks stand to capitalize on economies of scale, particularly under the gravity of large holdings of nonperforming assets during periods of economic contractions.

The remainder of this paper is as follows: we first discuss the varying perspectives on too-big-to-fail in US banking. Secondly, we document the existing literature concerned with both banking performance and environmental factors in US banking. Thirdly, we present a brief exposition of production theory and the formulation of economies of scale from the representative firm's cost minimization efforts. Fourthly, we discuss our primary empirical strategy paying particular attention to how to best quantify TBTF moral hazard and define the spectrum of banks most intensely operating in critical markets. With our theoretical framework in mind, we showcase the primary data we use for our analysis and, finally, specify our econometric model to derive elasticity-based nonperformance-adjusted returns-to-scale estimates over time for banks with the strongest involvement in critical markets.

3.2 Perspectives on Too-Big-To-Fail

The tension between gains in scale economies and growing TBTF is largely unresolved. The literature presents mixed evidence of increasing and decreasing returns-to-scale depending on the year and depending on the size of the banks analyzed in a given sample, thus there is a need for much more extensive analysis of scale economies across banks of all sizes and across all time that data may be available (J. A. Clark, 1996; Zhou, 2009; Barth & Wihlborg, 2016; Wheelock & Wilson, 2018). Given that mergers and acquisitions are costly for firms, regulators and bank users, it is important that measures of scale economies are accurate for informing the appropriateness of any such consolidation effort. More importantly, the traditional definition of TBTF is somewhat restrictive in its scope—while TBTF is traditionally concerned with the *size* of banks, it is not *exclusively* related to size alone, thus scale economies need to be considerate of factors in the banking environment beyond simply bank size.

With these nuances in mind, we take time highlight that fact that TBTF has a few key competing definitions. Papers like Henderson et al. (2015) and Wheelock & Wilson (2018) generalize TBTF is a push-pull relationship between regulators and policymakers. In particular, this push-pull dynamic seeks to balance regulation of bank size and the desire of individual banks to capitalize on scale economies. Policies like the Dodd-Frank Act and Consumer Protection Act of 2010 were intended as an overhaul to the regulation of the financial sector with particular attention given to the regulation of a bank's size among other characteristics. The concern with this balance grows out of the increasing number of mergers, consolidations and acquisitions particularly in the past two decades.

On the one hand, a bank's desire to merge or be acquired could very well be to capitalize on their increasing returns-to-scale, but expectedly, as banks merge or consolidate, the competitive landscape also shrinks. Declining competition coupled with growing bank size represents one perspective on the sources of the TBTF phenomenon. Figure 25 reiterates the key point of Wheelock & Wilson (2009) by describing the overall growth of assets (a reputable proxy for size) across banks and the number of unique reporting institutions insured by the FDIC (capturing the extent of the competitive landscape).

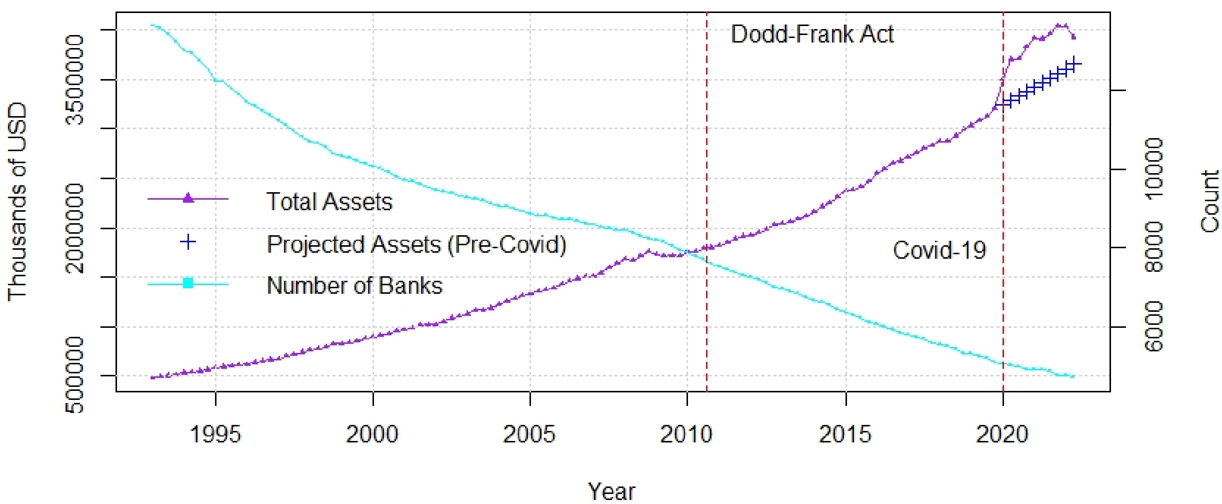


Figure 25: Total Bank Assets & Operating Banks

Visual evidence from Figure 25 is the driving force behind the perspective that TBTF is a problem of bank size and the competitive landscape of the banking industry. This definition notwithstanding, the alternative generalization of TBTF can be found in the seminal work of Stern & Feldman (2004), which has a very straightforward explanation of TBTF in banking:

> *The failure of a large banking organization is seen as posing significant risks to other financial institutions, to the financial system as a whole, and possibly to the economic and social order. Because of such fears, policymakers in many countries—developed and less developed, democratic and autocratic—respond by protecting uninsured creditors of banks from all or some of the losses they otherwise would face. These banks have assumed the title of "too big to fail" (TBTF), a term describing the receipt of discretionary government support by a bank's uninsured creditors who are not automatically entitled to government support.*

Other studies such as Hughes & Mester (1993) and Zhou (2009) also take a keen interest in empirically defining how systemically important large banks have become. Zhou (2009) specifically introduces an indicator for the systematic importance of a bank and finds that size does not always matter. In other words, there is evidence that size is not *always* a valid proxy for a bank's relative importance and macroeconomic interconnectedness.

Alternative perspectives reinforcing the key claims of Stern & Feldman (2004) characterize TBTF as the idea that banks have grown so large that they cannot be allowed to fail. By reaching a certain size and scale, TBTF implies that regulators and policymakers have no choice but to bailout banks despite risky portfolio decisions. In a sense, TBTF is as much an issue of moral hazard and bailout expectations as it is a balance of industry competitiveness with banking performance (Mishkin, 2006; Stern, 2012; Afonso et al., 2014).

These findings reinforce the claims put forth in Stern & Feldman (2004), but also call into question the policy effectiveness of the Dodd-Frank Act and Consumer Protection Act of 2010. In theory, there's an underling assumption that the aforementioned policies force banks to take fewer risks thereby reducing the expectations of future bailouts. However, as evident by Wheelock & Wilson (2018) and Wheelock & Wilson (2012) (among others), large and highly integrated banks face near-zero marginal costs and increasing returns-to-scale, thus, banking consolidation, mergers and acquisitions still continue to rise at an accelerating rate in spite of policy intervention or similar regulatory efforts. This makes disentangling the desire for achieving scale economies from risk appetite somewhat difficult.

To begin disentanglement effort, we must first address what types of banks might engage in risky behavior due to expectations of policy intervention. Works Stern & Feldman (2004) and the Federal Reserve Board of Governors (FRB) provide a handy definition of banks that operate in so-called *critical markets* and provide us with a starting point for developing our own TBTF criterion:

> *The agencies consider a firm significant in a particular critical market if it consistently clears or settles at least 5 percent of the value of transactions in the critical market.*

Where critical markets include markets for federal funds, foreign exchange markets, commercial paper, government securities, corporate securities and mortgaged-backed securities (MBS) as noted in Stern & Feldman (2004).[53] In theory, because these banks are identifiable from characteristics of the above criteria, we can separate them from banks that do not substantially operate in the aforementioned markets.

However, this definition is not necessarily compatible with the traditional publicly available data used in most banking studies.[54] Furthermore, it is limited in scope. For example, a bank may conduct 5%

[53]These banks are also known as *large complex banking organizations* (LCBO), although "large" is misleading in this definition. While "large" and "complex" are not mutually exclusive, the degree of complexity that a bank's operations fall under tends to be more predictive of bailout expectations as noted in Grammatikos & Papanikolaou (2018) than its size alone. Grammatikos & Papanikolaou (2018) also show empirically that less-complex banks tend to be less likely to be bailed out regardless of size. In a sense, large banks tend to be complex, but complex banks are not necessarily large.

[54]For example, data on a bank's holdings of commercial paper is not possible to discern from most high-frequency and long

of transactions in the market for mortgage-backed securities, but significantly less than 5% in all other critical markets, yet still be defined as a critical market bank. Alternatively, a bank may conduct 4% of its transactions each in multiple markets, and fail to be defined as a critical market bank despite doing sizeable business across all critical markets in the aggregate. Additionally, the FRB no longer uses the term "complex" explicitly in its lexicon when describing potential TBTF banks, rather the FRB provides a definition for *Large Financial Institutions*, which is, expectedly, more limited in scope:

> *Large financial institutions include U.S. firms with assets of $100 billion or more and foreign banking organizations with combined U.S. assets of $100 billion or more.*

The above definition for Large Financial Institutions, while useful for identifying the most extreme outliers in the commercial banking industry, is not valuable and suffers from heavy omission other highly-integrated institutions that fall short of the $100 billion threshold. The reductive definitions of TBTF banks between both Stern & Feldman (2004) and the FRB in the present day reflect the inconsistencies and difficulty associated with identification of TBTF banks.

Finally, a rather strong assumption nested in the description of TBTF banks described in Stern & Feldman (2004) is that banks who are **not** operating in critical markets are not expecting policy to intervene should they take misguided risks or engage in other forms of moral hazard. In reality, the degree to which banks are involved in critical markets is in and of itself a sliding scale. Later, we propose a means to more flexibly differentiate between banks highly involved in so-called critical markets from banks who are less-involved in the same markets. It is worth noting that a limitation of any TBTF study is that we can only *infer* TBTF-type behavior using the data at our disposal. Banks do not openly signal whether or not they are intentionally taking excess risk and expect policy intervention should they become insolvent.

Beyond defining our separation criteria, we must also think about the cost-benefit analysis of engaging in TBTF moral hazard. Fundamentally, we must ask ourselves if these critical market banks stand to gain from engagement in TBTF moral hazard. An interesting twist on TBTF in this context is that banks need not be necessarily "too big" to be TBTF, rather, as highlighted in Stern & Feldman (2004) as well as works like Barth & Wihlborg (2016) and Kaufman (2014), said banks might be too-*complex*-to-fail, which is to say they are so integrated in critical markets that to allow them to fail would result in spillovers outside of the banking sector into the general macroeconomy.[55]

In order to answer our motivating questions, we need to quantify moral hazard or capture risk appetite.

banking panel data sets. Additionally, quantifying a bank's involvement in foreign exchange markets is exceedingly difficult and is often ignored for most banking performance studies, which tend to lean more towards domestic accounts and transactions when specifying inputs and outputs of a given bank.

[55] In this sense, bailouts are performed based on the assumption that mitigating potential spillovers is more critical to policymakers than the expectations of future bailouts created by said act.

Thankfully, the literature has some guidance on this. Papers like Berger & DeYoung (1997) and Hughes et al. (2001) would stress the importance of the role nonperforming assets have in predicting bank failure. Nonperforming assets, while imperfect, partially captures the excess risks banks may take and can be thought of as a factor that influences the production environment. If this is the case, then elasticity-based estimates of scale economies should be in-turn affected by nonperforming assets as well.[56]

If nonperforming assets are critical to the banking environment as a form of moral hazard as the literature suggests, then we can directly measure its impact on scale economies across a distribution of banks operating in critical markets. However, in order to achieve this, we must first form a more flexible criteria for identifying banks operating in critical markets. This is an achievable task discussed in greater detail in our empirical methodology. Secondly, we must map moral hazard via nonperforming assets into the production environment through an approach that is less restrictive than a traditional least squares specification.

To map moral hazard into the production environment appropriately, we turn to best practices from the production economics field. As is the norm in most works in the sphere of production economies, seemingly unrelated regressions (SUR) models are commonly used to jointly estimate both a representative firm's profit [cost] function as well as their factor demand equations. A typical SUR model takes the general form of $y_{si} = x_{si}^T\beta_s + u_{si}$, where y_{si} describes a stacked matrix of dependent variables for s equations, x_{si} is a matrix of our independent variables for all s equations, and β_s captures the coefficient estimates for each equation. Technically speaking, the SUR model originally described by Zellner (1962) is more efficient than OLS due to its ability to account for cross-equation correlation as well as its flexibility to handle cross-equation restrictions.

Despite the popularity of SUR model, it faces major limitations in its ability to estimate the impact of environmental factors on firm returns-to-scale, which are oftentimes elasticity-based. We posit that major innovations in nonparametric and semiparametric econometrics makes overcoming this limitation possible. In particular, the Henderson et al. (2015) SPSC SUR estimator allows for model coefficients to be in-and-of themselves functions of environmental factors (denoted as z) such that the Zellner (1962) SUR model can be defined more flexibly as $y_{si} = x_{si}^T\beta_s(z_{si}) + u_{si}$, wherein our model coefficients, β_s, now vary in accordance to some nonlinear function of our environmental factors, z_{si}. In essence, if our coefficients are now functions of z, then so too will be our returns-to-scale estimates, making the SPSC SUR more appropriate than a traditional linear SUR for returns-to-scale studies wherein environmental factors play an important role in the representative firm's production setting.[57]

[56] Note that this is fundamentally different in interpretation than works in the literature that control for nonperforming assets. Controlling for nonperformance is a marginal affect on total costs, but has no influence on other cost elasticities from which returns-to-scale are derived.

[57] We outline the SPSC SUR in more detail in econometric strategy as well as in our appendix.

3.3 Existing Literature

To be specific, there are two banking literatures we are concerned with. The first is the literature associated with estimating banking performance and performance indicators. Alongside banking performance is a literature concerned with so-called "environmental factors" and their entry into the banking production environment. Abstracting from these two specific literatures, there are two seminal literature reviews that cross over elements of both literature bodies of interest: Colwell & Davis (1992) and Berger & Mester (1997).

Colwell & Davis (1992) provides a survey of the literature associated with banking output, including output concepts, empirical methodologies and definitions of banking productivity. The author provides a clear definition of the "production approach" of conceptualizing banks, which is to say that they are firms who strictly utilize labor and capital to produce different categories of deposits and generate loans. The author highlights how ordinary least squares (OLS) cannot disentangle technical efficiency from technical change and concludes by reiterating that banking as a industry is difficult to both theoretically conceptualize and empirically measure.

Similarly, Berger & Mester (1997) acts as an excellent primer and literature survey of sources of variation in measured banking efficiency, differences in the efficiency concept, and of efficiency measurement methods. In terms of efficiency as a concept, the authors identify standard profit, alternative profit, and cost efficiency as the three main efficiency concepts across the literature. Related to our study, the authors highlight the differences in parametric versus nonparametric approaches in that nonparametric methods generally ignore input prices and do well to account for a firm's "technical optimization" while a parametric approach does well to capture a firm's "economic optimization."

One final literature review would be that of Sharma et al. (2013) who provide an extensive literature review of 106 banking efficiency and productivity studies. The authors emphasize that the largest gap in the literature is the lack of an underlying "conceptual" or theoretical modeling for the banking sector. Descriptive statistics of the literature at the time of publication highlight that the majority of studies focus on "efficiency and productivity estimation" and "empirical evidences on the determinants of efficiency and productivity." Most studies focus on US banking and apply data envelopment analysis (DEA) or stochastic frontier analysis (SFA).

3.3.1 Banking Performance

Broadly speaking, banking performance can be characterized in one of three ways: efficiency, productivity and scale economies. The lion's share of the existing literature focuses on banking efficiency, specifically cost efficiency. The cost efficiency of banks has especially been under a microscope since the Global Financial

Crisis (GFC).

The bulk of the literature focuses on cost frontier analysis through DEA methods. These methods are motivated by production theory and seek to evaluate relative technical inefficiency in US banking. Papers like Feng & Serletis (2010), Berger et al. (1993), Berger & Humphrey (1997) and Berger (1993) are all prime examples of such a body of literature. Feng & Serletis (2010) focuses only on the largest US banks and uses a Bayesian framework to estimate efficiency, economies of scale, and productivity between 2000 and 2005 (pre-GFC).

Berger et al. (1993) and Berger & Humphrey (1997) are related in scope, but seek to understand the determinants of technical inefficiency in banking. Specifically, Berger et al. (1993) finds that most technical inefficiency in US banking is not due to cost inefficiency on the side of their inputs, but rather on the output side from deficient output revenues while Berger & Humphrey (1997) finds that "problem loans" or nonperforming loans tend to be the strongest source of cost inefficiency—a detail that has become even more relevant during the GFC. However, to truly understand banking efficiency, we must differentiate between banks in "normal" times from banks during economic crises, although, the two regimes are not mutually exclusive from one another as noted in Assaf et al. (2019).

In more explicit detail, Assaf et al. (2019) examines how bank efficiency during "normal" times can be used to predict bank survivability during economic crises. In particular, the authors stress that *cost* efficiency during normal times enhances survivability during crises and can even enhance profitability relative to less-efficient institutions. The authors highlight that *profit* efficiency has limited benefits and that cost efficiency is a superior efficiency concept for assessing the probability of banking failure during crises. A policy implication of these findings are that bank regulators, managers and other stakeholders in the financial sector should target or monitor cost efficiency during normal times to mitigate risks prior to the onset of economic downturns.

Related to the efficiency concept of banking performance, Wheelock & Wilson (1999) motivate their study of banking efficiency and productivity by the preceding empirical evidence that US banks are generally cost-inefficient. Despite this, however, the authors find that technical efficiency and productivity have grown considerably between 1984 and 1993. The authors show that even small gains in technical efficiency can push out the production frontier for the industry, thus cost-inefficiency is not necessarily reflective of the industry's long-run behavior as a whole; rather, it reflects the growing technical change diffusing across the banking sector.

Related to projecting bank failure, Barr et al. (1994) estimate a variety of multiple-input, multiple-output DEA models to quantify bank management quality with the goal of contributing to the literature's understanding of bank failures and the determinants of bank failures. The authors find that bank management is an important determinant of bank failure. When forecasting bank failure using existing models from the literature, the authors find the exclusion of managerial quality significantly lowers forecast precision.

While both Assaf et al. (2019) and Barr et al. (1994) are interested in projecting bank failure, there is not much discussion beyond management quality with regards to other determinants of banking failure. However, Berger & DeYoung (1997) investigates the intersection between "problem loans" and banking efficiency. Granger causality tests are used to test various hypotheses concerning loan quality and cost efficiency. The authors' core conclusion is that estimated cost efficiency is a strong predictor for future problem loans.

Looking at bank survival in light of the events of the GFC, Berger & Bouwman (2013) examines variation in banking performance (survival, market share) over different eras of financial strain in US history. The main results highlight the importance of bank capital for small bank survival and stresses that capital does well to increase the performance of larger banks during financial crises. Simply put, there are asymmetric effects of capital on a bank's performance over economic downturns depending on whether the bank is large or small.

Turning towards the importance of scale economies in US banking, Henderson et al. (2015) and Wheelock & Wilson (2018) both have looked at scale economies from the lens of the TBTF perspective, however, there are many other papers providing a focus on scale economies in other contexts. Take for instance Feng & Zhang (2014) who examine returns-to-scale for large US banks from 1997-2010 using a random coefficient stochastic distance frontier model. This allows for coefficient variation across individual banks. Within this framework, the authors discern some level of unobserved technical heterogeneity in large banks and find that large banks exhibit constant returns-to-scale. The authors claim that ignoring this heterogeneity across banks can mislead how banks rank in scale economies.

In a similar vein, Wang (2003) takes the position that banks operate a service provider who add value to households and firms through unified banking services. This value-add position is different than the output-oriented or financial intermediation approach. The author then estimates a series of production and cost functions spanning from 1986 through 1999. The simple conclusion is that banking service providers likely exhibit small, increasing returns-to-scale (IRS), as opposed to the popularly held belief that banks experience constant returns-to-scale (CRS).

A final angle in the banking performance literature is that of regional heterogeneity in banking services and performance, in particular with a focus on community banks (CB) versus non-community banks. Papers like Feng & Zhang (2012) and Minuci (2021) are both excellent works in this smaller body of literature. In particular, Feng & Zhang (2012) compare productivity and efficiency of large and community banks between 1997 and 2006. The authors use a Bayesian approach and a true rand effects stochastic distance frontier model to parse out heterogeneity in efficiency between large and community banks. The authors then estimate a Divisia-like productivity index and find that failure to account for observed efficiency heterogeneity results in improper rankings of banks and mismeasured estimates of technical efficiency and productivity growth.

Unlike, Feng & Zhang (2012), but still somewhat related, Minuci (2021) looks at state-level CB cost efficiency in low-interest macroeconomic environments and finds that over the past two decades, community banks' relative cost efficiency has fallen by around 30%. Additional findings indicate that customer characteristics in a given state or region shares a strong correlation with bank performance.

3.3.2 Environmental Factors in Banking

Alongside the literature concerning banking performance is the literature related to environmental factors and their impacts on a representative firm's optimizing behavior. Similar to the estimation of banking performance, there is neither consensus, nor clarity on what the most important and relevant environmental factors are, however, there is strong evidence that omitting environmental factors can heavily bias coefficient estimates that would be used to estimate productivity and scale economies, in particular.

Specifically for the banking industry, some popular environmental factors frequently leveraged for econometric analysis are nonperforming loans, liquidity, and leverage. In most studies, environmental factors are treated as controls, and, as such, provide a marginal effect that one can interpret straightforwardly. A critique to this approach, however, is that it is unclear *how* environmental factors should enter into the production environment. If environmental factors are neither an input to production, nor a firm output, then a linear expression of their marginal effects on the representative firm's cost minimization or profit maximization efforts may be inappropriate, despite an otherwise well-specified, and theoretically consistent modelling framework such as a Cobb-Douglass production function or a translog cost function.

When thinking about environmental factors in the context of banks, we must first delineate between *internal* versus *external* environmental factors. Broadly speaking, environmental factors are measurable dimensions of the production environment that impact a representative firm's objective function (be it profit maximization or cost minimization), but are not strictly speaking inputs or outputs of the production process itself.

Internal environmental factors are firm-specific characteristics that impact the production environment. For example, nonperforming assets, liquidity, managerial quality, leverage or other such measures of a bank's solvency are not necessarily inputs to production, nor necessarily related to production technology, but differ firm-to-firm. External environmental factors on the other hand are factors that are specific to the macroeconomic environment that all firms face. Such factors would be things like per capita income, quantifiable regulations, tax rates, real interest rates and so on. In a sense, internal environmental factors are more microeconomic in nature while external environmental factors lean more macroeconomic in nature.

A key example of the role of environmental factors impacts banking would be Dietsch & Lozano-Vivas (2000). Looking at banks in France and Spain, the authors investigate the role that environmental factors have on estimated cost efficiency. The authors seek to identify whether there are asymmetries in environmental factors in a cross-country setting. Without environmental factors, the authors find a large gap in cost efficiency between France and Spain, however, the gap is negligible when environmental factors are accounted for.

Similarly, Lozano-Vivas et al. (2002) conducts a cross-country study of banking efficiency using banking data from various European nations. The authors argue that idiosyncratic factors specific to each individual nation has a significant impact on that nation's measured cost efficiency. The authors specifically leverage environmental factors like income per capita, population density, deposit density, deposits per branch, the equity-to-assets ratio and return-on-equit (among other facgtors). When comparing their "complete" DEA model to a "basic" DEA model that omits the aforementioned the environmental factors, the authors find that such omissions can lead to misleading measurements of efficiency.

A complementary piece to Lozano-Vivas et al. (2002) would be that of Chaffai et al. (2001) who also conduct a cross-country study of banking productivity by constructing a Malmquist-type index built off the back of an output-distance function. The authors find that the largest gaps in productivity across large European countries can be explained by differences in production technologies as well as country-specific environmental factors.

Abstracting from cross-country comparison studies, pieces like Drake et al. (2006) focus solely on regulatory factors that impact a single region's banking sector, specifically its technical efficiency—in this case, the region of Hong Kong. Using a slack-based nonparametric framework, the authors quantify the impacts of changing regulatory regimes in the Hong Kong banking sector and find that failure to incorporate regulatory "slack" into models of technical efficiency produce heavily biased measures of technical efficiency. The authors key takeaways echo previous pieces by stressing that the omission of environmental factors can heavily impact efficiency scores both cross-sectionally and temporally.

Slightly outside of the field of economics, but related to analysis of banking and finance, pieces like DeVaughn & Leary (2010) evaluate ownership quality of a small sample of US banks from the state of Florida. By quantifying ownership characteristics like concentration, industry experience, the authors find that prior industry experience and shared industry experience amongst ownership entities generally reduces organization distress, thereby minimizing (in theory) probabilistic failure.

Ultimately, while the literature associated with the variation in banking production environments is smaller and, in some ways, a subset of the larger literature associated with banking performance, its relevance continues to grow especially with changes to exogenous factors such as regulatory regimes, but also with the discoveries of import internal factors, particularly that asset nonperformance. Additionally, given that increasing adoption of semiparametric and nonparametric methods in the analyses of banking performance, it stands to reason that environmental factors will, too, become more commonplace for the future studies in the banking performance literature overall.

3.4 Theory

Like the majority of papers in the banking literature, we draw heavily from production theory to motivate our econometric specification. Furthermore, given that scale economies are of interest to us post-estimation, it is essential that the formulation of our cost function along with our derivation of economies of scale echo strong theoretical foundations.

3.4.1 Cost Function Formulation

Following Wheelock & Wilson (2018), we assume that $\mathcal{X} \in \mathbb{R}^J_{\geqslant 0}$ is a vector of dimension J input quantities and $\mathcal{Y} \in \mathbb{R}^Q_{\geqslant 0}$ is a vector of dimension Q output quantities. Corresponding to our set of input quantities, $\mathcal{X}$, let $\mathcal{W} \in \mathbb{R}^J_{\geqslant 0}$ be a set of input prices also of dimension J.[58]

With this notation in mind, we can define our total costs as $\mathcal{C} = \mathcal{W}^T\mathcal{X}$. The cost-minimizing firm seeks to minimize $\mathcal{C}$ with respect $\mathcal{X}$ subject to the production-transformation function $f(\mathcal{X}, \mathcal{Y})$. The solution to this minimization problem yields a representation of input quantities as a function of output quantities and input prices such that $\mathcal{X} = g(\mathcal{Y}, \mathcal{W})$. Via substitution, we arrive at the following cost function described by equation (3.1).

$$\mathcal{C} = \mathcal{W}^T\mathcal{X} = \mathcal{W}^T g(\mathcal{Y}, \mathcal{W}) = \mathcal{C}(\mathcal{Y}, \mathcal{W}) \tag{3.1}$$

In logs, it can be shown that the cost-minimizing decision for any one input price $w_j \in \mathcal{W}$ results in equation (3.2).

$$\frac{\partial \log(\mathcal{C})}{\partial \log(w_j)} = \frac{\partial \mathcal{C}}{\partial w_j} \times \frac{w_j}{\mathcal{C}} \tag{3.2}$$

Via the envelope condition, we can express the representative firm's conditional factor demand for input $x_j \in \mathcal{X}$ as in equation (3.3).

[58] It is worth pointing out our assumptions are slightly different than Wheelock & Wilson (2018) who allow for their vector of input prices, $\mathcal{W}$, to be negative. Furthermore, we allow for both $\mathcal{X}$ and $\mathcal{Y}$ to contain values possibly equal to zero, rather than strictly positive.

$$x_j = \frac{\partial \mathcal{C}}{\partial w_j} \tag{3.3}$$

Via substitution, we can rewrite equation (3.2) as equation (3.4).

$$\frac{\partial \log(\mathcal{C})}{\partial \log(w_j)} = \frac{x_j \times w_j}{\mathcal{C}} = S_j \tag{3.4}$$

Where $S_j \in \mathcal{S}$ is the cost-share corresponding to input j where $\mathcal{S} \in \mathbb{R}^J_{\geqslant 0}$ and $\sum_j S_j = 1$.

3.4.2 Derivation of Scale Economies

Building off of Sandmo (1970) in conjunction with the definitions described by Wheelock & Wilson (2018), we formulate and derive scale economies from the representative firm's cost function. Given that we can decompose $\mathcal{W}^T\mathcal{X}$ to simply be the sum of the product of their individual components, we can begin to formulate scale economies in the following manner using equation (3.5) as our starting point.

$$\mathcal{C} = \sum_{j=1}^{J} x_j w_j \tag{3.5}$$

Note that we can allow the inclusion of a production constraint, $\mathcal{Y}$, to be equivalently written as a function of $\mathcal{X}$ such that $\mathcal{Y} = h(\mathcal{X})$. This allows us to form the Lagrangian described by equation (3.6).

$$\mathcal{L} = \sum_{j=1}^{J} x_j w_j - \lambda\{h(\mathcal{X}) - \mathcal{Y}\} \tag{3.6}$$

Where the λ term is our Lagrangian multiplier. Accordingly, equation (3.7) captures the necessary conditions for cost minimization in this setting.

$$w_j - \lambda h_j(\cdot) = 0, \text{ for } j = 1, ..., J, \text{ where } h_j(\cdot) = \frac{\partial h(\cdot)}{\partial x_j} \tag{3.7}$$

The result of equation (3.7) taken with our production constraint, $\mathcal{Y} = h(\mathcal{X})$, provides $J + 1$ equations that describe the factor inputs and shadow price, λ, as functions of our outputs, $\mathcal{Y}$, and each w_j input price. As a result, we can define total costs indirectly as a function our the level of output, $\mathcal{Y}$ such that $\mathcal{C} = \mathcal{C}(\mathcal{Y}, \mathcal{W})$. This indirect cost function expresses $\mathcal{C}$ based on the level of output the representative firm faces conditioned on minimizing costs associated with each input, x_j, at every dimension of $\mathcal{Y}$.

Furthermore, this result also contextualizes λ as the representative firm's marginal cost. Via differentiation of equation (3.5), we arrive at equation (3.8).

$$\mathrm{d}\mathcal{C} = \sum_{j=1}^{J} w_j \mathrm{d}x_j \tag{3.8}$$

Similarly, by differentiating our production constraint, and via substitution and some manipulation from equation (3.7), we have equation (3.9).

$$\mathrm{d}\mathcal{Y} = \sum_{j=1}^{J} h_j(\cdot)\mathrm{d}x_j \tag{3.9}$$

A direct implication of equation (3.9) is that $\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}\mathcal{Y}} = \lambda$. With this in mind, we can define the average cost as a function of our vector of outputs $\mathcal{Y}$. This is described by equation (3.10).

$$m(\mathcal{Y}) = \frac{\mathcal{C}(\mathcal{Y})}{\mathcal{Y}} \tag{3.10}$$

If one were to take the derivative of equation (3.10), then we arrive at equation (3.11).

$$\frac{\partial m(\mathcal{Y})}{\partial \mathcal{Y}} = \frac{1}{\mathcal{Y}^2}\left\{\frac{\partial \mathcal{C}}{\partial \mathcal{Y}}\mathcal{Y} - \mathcal{C}(\mathcal{Y})\right\} \tag{3.11}$$

Subsequently, it can be shown from equation (3.11) that $\frac{\mathrm{d}m(\mathcal{Y})}{\mathrm{d}\mathcal{Y}} \gtreqless 0 \iff \frac{\mathrm{d}\mathcal{C}}{\mathrm{d}\mathcal{Y}} \gtreqless \frac{\mathcal{C}(\mathcal{Y})}{\mathcal{Y}}$. Furthermore, from equation (3.8) and our definition of λ, it can be shown that $\lambda \gtreqless \frac{\sum_{j=1}^{J} w_j x_j}{\mathcal{Y}}$. With our definition of production, $\mathcal{Y} = h(\mathcal{X})$, as well as our necessary condition described by equation (3.7), we arrive at equation (3.12).

$$\lambda \gtreqless \lambda \sum_{j=1}^{J} \frac{h_j(\cdot)x_j}{h(\cdot)} \tag{3.12}$$

From our production constraint, $\mathcal{Y} = h(\mathcal{X})$, economies of scale can typically be defined as $k^{\varepsilon^y}\mathcal{Y} = h(kx_1, ..., kx_J)$ where k is some constant describing the magnitudinous increase in output for an increase in our factor inputs.[59] Via differentiation, it can be shown that $\varepsilon^y k^{\varepsilon^y - 1}\mathcal{Y} = \sum_{j=1}^{J} h_j x_j$. Given that $k^{\varepsilon^y - 1}\mathcal{Y} = h(\cdot)/k$, we can precisely define our scale economies by equation (3.13).

$$\varepsilon^y = \sum_{j=1}^{J} \frac{h_j(\cdot)x_j}{h(\cdot)}, \text{ where } k = 1 \tag{3.13}$$

With the results of equations (3.12) and (3.13) taken together, it can be shown that $\mathrm{d}m(\mathcal{Y})/\mathrm{d}\mathcal{Y} \lesseqgtr 0$, which implies that $\varepsilon^y \lesseqgtr 1$, where $0 < \varepsilon^y < 1$ implies decreasing returns-to-scale, $\varepsilon^y = 1$ implies constant returns-to-scale and $\varepsilon^y > 1$ implies increasing returns-to-scale. Alternatively, we can also express scale economies with our formulation.

[59] An important assumption in our definition of economies of scale is homogeneity of degree ε^y in $\mathcal{Y}$.

$$\xi^y = 1 - \varepsilon^y \tag{3.14}$$

A negative value of $\xi^y < 0$ implies increasing returns-to-scale and a positive value of $\xi^y > 0$ implies decreasing returns-to-scale. Of course, a value of $\xi^y = 0$ would be constant returns-to-scale. Depending on the environment, equation (3.14) and its formulation of performance can be preferable to equation (3.13).

Returning to equation (3.1), so long as we can express our production set, $\mathcal{Y}$, as $\mathcal{Y} = h(\mathcal{X})$, then we can express our cost function as in (3.15).

$$\mathcal{C} = \mathcal{C}(\mathcal{Y}, \mathcal{W}) = \mathcal{C}(h(\mathcal{X}), \mathcal{W}) \tag{3.15}$$

The mapping of both (3.15) to (3.4) and (3.13) provides a tractable relationship not only to our cost-share equations, but also our returns-to-scale formulation.

3.4.3 Practical Considerations for Applications in Banking

We show using basic production theory in conjunction with key literature pieces the construction of a representative firm's cost function as well as their subsequent cost-share equations implied by the derivative conditions presented in equation (3.4). Furthermore, using the same cost function and a production constraint, $\mathcal{Y} = h(\mathcal{X})$, we arrive at our derivation of economies of scale described by equation (3.13).

While some practitioners may prefer profit or alternative profit specifications over a cost function, we argue that the choice of a cost function is appropriate for three main reasons over its alternatives:

1. In the banking environment, output prices are often exceedingly difficult to measure and, more often than not, are treated as exogenous. Given that balance sheet data is widely available, it is quite straightforward to approximate both input prices, $\mathcal{W}$, and input quantities, $\mathcal{X}$, using information from the balance sheet or income statement making cost functions as described implicitly by equation (3.4) more attractive econometric specifications

2. Compared to single-equation models, there are considerable gains in efficiency when estimating the cost function as a system of equations in a form such as a transcendental logarithmic (translog) cost function

3. Cost functions provide much more flexibility, particularly translog cost functions, where cross-equation restrictions can be readily imposed with little cost to practitioners while simultaneously satisfying conditions set forth in production theory. Given that banks are often viewed as multi-output firms, the flexibility of a cost function specification is paramount when $\mathcal{Y}$ is of a high dimension

Conceptually, returns-to-scale derived from a reduced form cost function can be thought of in two ways: firstly, it can tell us the relative increase in outputs a banking firm can achieve for a unit increase in total costs; secondly, as noted in Wheelock & Wilson (2018), it can tell us how efficiently resources are employed for the provision of banking services. This is an important distinction from interpretations of returns-to-scale from a profit function, which in a sense capture shareholder value created by expansionary efforts of a given bank. Depending on the context, profit functions may be seen as more desirable of a setup compared to cost functions, although the literature heavily leans in favor of cost functions.

In a sense, scale economies, particularly among bank holding companies, derived from a cost function provide insights about both firm *efficiency* and *performance*. It is because of these features, that we prefer economies of scale as our performance indicator of interest over estimates of productivity or frontier efficiency.[60]

3.5 Empirical Strategy

Our empirical methodology is threefold. First and foremost, it is essential to quantify moral hazard, a feat easier said than done. Secondly, we must define a sensible criterion to differentiate banks highly involved in critical markets from banks with minimal involvement in the same markets. Finally, we must be able to map such a quantification of moral hazard to banking performance from the same econometric specification while maintaining theoretical consistency.

In abbreviated detail, we opt for nonperforming assets to function as our environmental variable, z_i, that our returns-to-scale estimates will necessarily be functions of. Furthermore, we separate TBTF banks from non-TBTF banks by their relative involvements in so-called critical markets guided by separation criteria originally outlined in Stern & Feldman (2004). We stress that our sample separation criteria is a first attempt to disentangle TBTF banks from non-TBTF banks by using separation criteria that is not wholly dependent on the absolute size of the banks themselves. Aside from the stylized facts on critical banking markets outlined in Stern & Feldman (2004), the separation criteria that follows in Section 3.5.2 has never been considered before, and as such should be taken as a "first step" or an appealing alternative to separation criteria that are entirely size-dependent. Finally, we quantify the impact of moral hazard on the returns-to-scale estimates expressed by critical market (TBTF) banks in our sample through the use of an SPSC SUR model treating nonperforming assets as our environmental variable, z. For some evidence on the relationship between nonperforming assets and critical market involvement, consider Figure 26.

[60]While outside the scope of analysis, it is possible to extend the reduced form framework we discuss herein for applications such as stochastic frontier analysis or similar DEA methods.

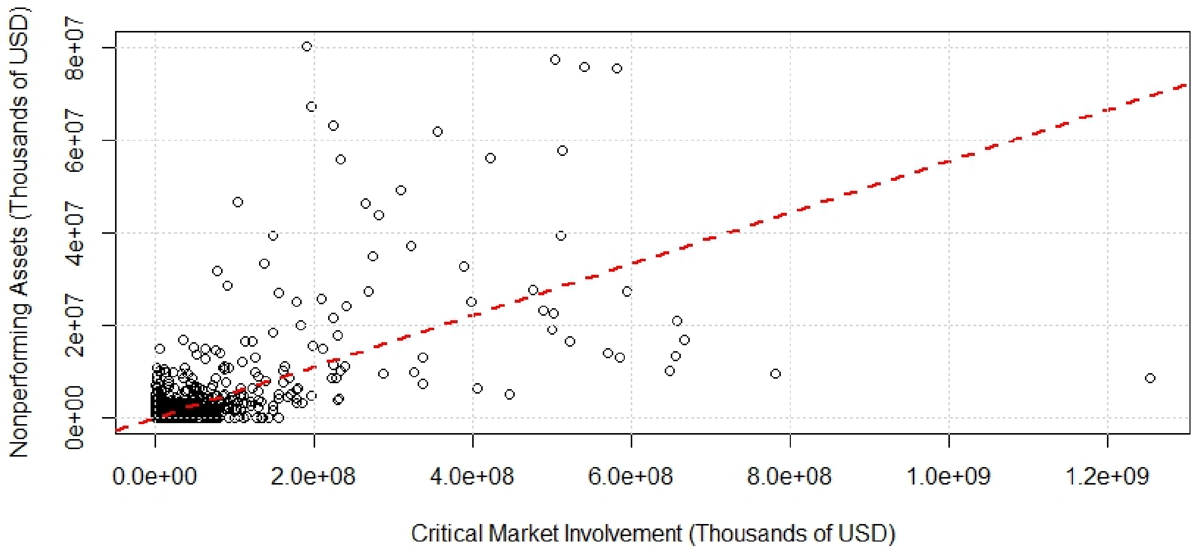


Figure 26: Relationship Between Nonperforming Assets & Critical Market Involvement

In particular, Figure 26 reaffirms that excess risk (captured by nonperforming assets) shares a moderate relationship with the magnitude of transactions a bank processes in critical banking markets such as the markets for mortgage-backed securities, treasury securities, and agency debt. Thus, our separation criteria that follows in Section 3.5.2 should be an adequate first-pass at identifying potential TBTF banks that is not exclusively reliant on size, but still consistent with the literature's view on the risk that nonperforming loans and assets have on the representative bank's survivability and overall performance.

3.5.1 Measuring Too-Big-to-Fail Moral Hazard

While some general parameters exist for identifying banks who might qualify as TBTF, measuring the presence of such a phenomena can be difficult. As argued, being a "large" bank does not necessarily indicate that a bank is TBTF. The literature offers some guidance on alternative measures of TBTF. Zhou (2009) leverages methodologies in extreme value theory (EVT) to identify likely candidates for TBTF banks while papers like Hughes & Mester (1993) look at a myriad of financial indicators like uninsured loans, output quality, and financial capital to evaluate how economies of scale and scope are impacted by indicators of TBTF. Returning to Stern & Feldman (2004), the authors posit that there are three key TBTF indicators:

1. Lower prices of bank funding

2. Higher shareholder wealth

3. Market prices

While, in theory, these are all excellent indicators of *symptoms* associated with TBTF, price data is incredibly difficult to measure, let alone generalize—both for market and bank funding prices. Shareholder wealth is also an unclear indicator and difficult to measure with existing data. As noted in works like J. A. Clark (1996), it is tempting to envision banks engaging in moral hazard if they have potential to maximize shareholder wealth, however, there are many wealth-maximizing decisions that banks can pursue outside of caustic TBTF-like financial engagements. Studies like Hughes et al. (2001) alongside the appendix of Wheelock & Wilson (2018) provide some guidance for the identification of scale economies while taking into consideration a given bank's risk appetite. We use these pieces as a starting point for constructing our own measure of TBTF through the lens of firm nonperformance.

For the purpose of our study, we look at *nonperforming assets*. We define nonperforming assets as the sum of total loans and leases both > 30 and > 90 days past due, total nonaccrual loans and leases financing receivables, and other real estate-owned assets.[61] Formally, our definition of nonperforming assets is linked by the above data series such that $NPA_{it} = P3LNLS_{it} + P9LNLS_{it} + NALNLS_{it} + ORE_{it}$.[62] One caveat with our definition of NPA_{it} is that $P3LNLS_{it}$ was considered confidential by the FDIC until the end of the 2000 calendar year, thus for consistency, we limit our empirical analysis to look from 2001 through 2021.

For illustrative purposes, Figure 27 shows the intensity of nonperforming assets in the aggregate held by banks of varing sizes.[63]

[61] While some studies define nonperforming loans as leases as those that are strictly > 90 past due, we prefer the flexibility of both past due buckets for the sake of maximizing data availability per observation and for ease of kernel weighting in our nonparametric procedure. Furthermore, given that we are opting for nonperforming assets, which is a composite of more than just nonperforming loans, we believe that we will still be able to confidently separate TBTF banks from non-TBTF banks. Also, while not immediately obvious, past due greater than 30 days does is not double counted in loans past due 90 days. In a sense, greater than 30 days past due as a measure includes 30-89 days past due.

[62] Definitions of FDIC BankSuite pneumonics can be found in our appendix.

[63] Contour graphs herein are produced using the "plotly" package in R. While we are using discrete data, interpolation is used between the discrete observations to produce a smooth contour graph.

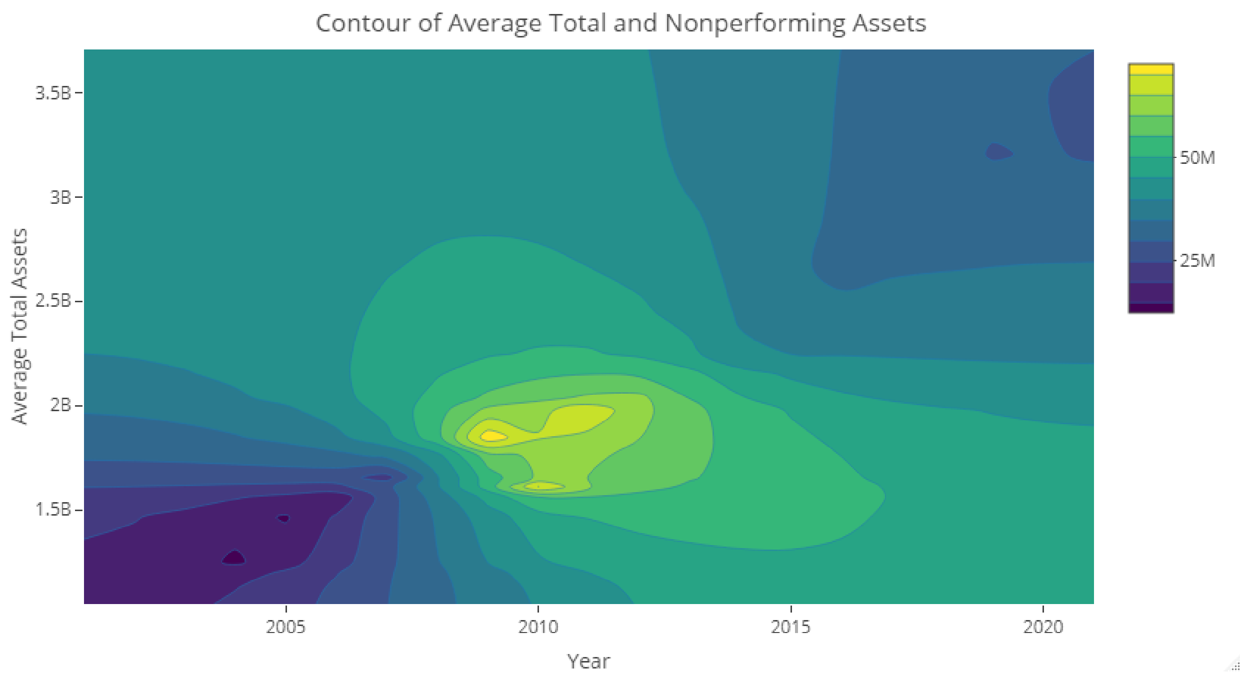


Figure 27: Contour of Avg. Total Assets & Nonperforming Assets Over Time

We see that, in the aggregate, the largest banks in any given year are not the ones holding the highest intensity of nonperforming assets. Rather, it appears to be mostly mid-sized banks in the periods of time surrounding the Global Financial Crisis. This confirms our initial suspicion that banks engaging in moral hazard *can* be large, but are not necessarily *always* large.

With our definition of moral hazard defined and quantified, we turn towards defining the degree to which a given bank operates in so-called critical markets.

3.5.2 Defining Critical Market Involvement

As argued, the Stern & Feldman (2004) description and former FRB definition of critical market banks or LCBOs are both somewhat dated and limited in scope. Expectations of policy intervention and engagement in moral hazard are sliding scales and, as such, warrant a more flexible separation criteria. Critiques notwithstanding, Stern & Feldman (2004) provides excellent guidance for thinking of TBTF banks and the types of markets they have a greater tendency to conduct business in. We propose a more flexible separation criteria, one that considers the relative involvement of banks in said critical markets relative to their size, but allows for variation in their market involvement year-to-year. Specifically, we adopt the following methodology for defining critical market involvement:

1. For each bank i in fiscal year t, we take the summation of said bank's holdings of U.S. Treasury securities, U.S. agencies and corporation securities, and mortgaged-backed securities and divide it by

a bank's total assets.[64] We denote this variable as $CMR_{it} = \frac{(SCUST_{it}+SCAGE_{it}+SCMTGBK_{it})}{ASSET_{it}}$, where $SCUST_{it}$, $SCAGE_{it}$, and $SCMTGBK_{it}$ are the FDIC SDI pneumonics for treasury securities, agency and corporation securities and mortgage-backed securities, respectively

2. For each bank i in year t, we assign a decile based on where said bank's CMR_{it} indicator falls relative to its peers in the same fiscal year. This decile indicator is our separation criteria wherein a bank that falls in the lowest decile (10^{th} percentile) is the least-involved in critical market activities while banks in the highest decile (90^{th} percentile) are the most-involved in critical market activities. For the purpose of our study and motivating questions, we focus on banks most-involved in critical markets

Given that this is one of the first studies to explicitly examine the impact of moral hazard on returns-to-scale for banks highly involved in critical markets, we believe that our separation criteria, while imperfect, is a reasonable starting point that offers a considerable improvement in categorizing bank complexity, regardless of size, over the preexisting criteria.[65] In fact, in the aggregate and over time, our criteria tends to express that participation in critical markets is largely irrespective of bank size in most years. Figure 28 provides illustrative evidence of this stylized fact.

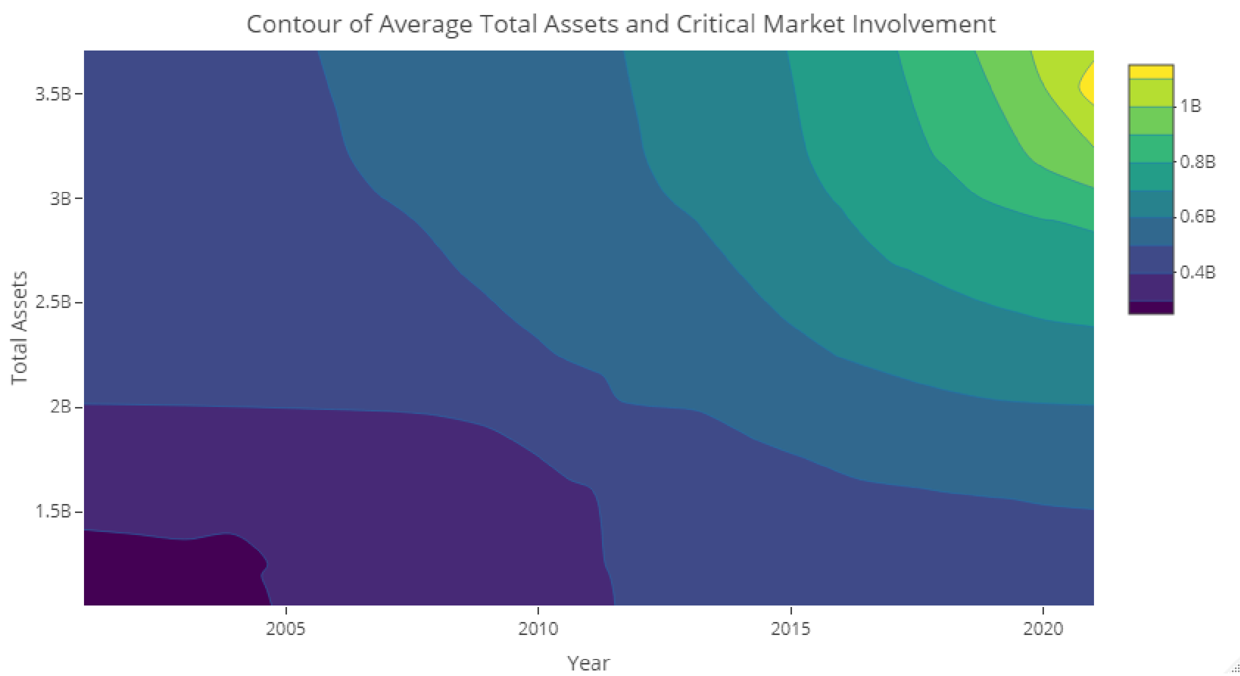


Figure 28: Contour of Avg. Total Assets & Critical Market Involvement Over Time

Figure 28 tends to illustrate that over most years in our sample (particularly from 2005 through 2017),

[64]In effect, we are normalizing the dollar value of given bank's involvement in a critical market by their relative size. This allows us to consider the possibility that banks do not **need** to be large to be TBTF.

[65]The choice of focusing only on the 90^{th} percentile, while somewhat arbitrary, serves to capture the most extreme-end of the right-tail of critical market involvement in any given year while preserving the flexibility of our definition year-to-year.

in the aggregate, the average bank engaged in critical market activities is not necessarily the largest bank in any given year. It should also be noted that the sizes of banks and their respective involvement in critical markets tends to increase monotonically over time. In fact, it would appear that *only* the largest average banks are engaging in critical markets during 2020 and 2021, thus, our separation criteria will pick up TBTF banks that are both large and small in a lot of years, but only large in others. This flexibility offers a strict improvement over more traditional TBTF criteria that focus solely on size.

3.6 Data & Descriptive Statistics

Our principle data source is the FDIC BankFind Suite: API for Data Miners & Developers. This is a relatively new provision on behalf of the FDIC. Previously, FDIC data used for most banking studies came from Call Report data or the statistics on depository institutions (SDI). The data provided from the FDIC BankFind Suite is closer to the SDI than Call Reports in terms of construction and nomenclature, however, in practice, the SDI data is effectively a lower-dimension and cleaner version of the Call Report data. The BankFind Suite data when compiled is functionally an unbalanced panel with each unique indicator being tied to an identifiable FDIC-insured depository institution in a given fiscal quarter.

The BankFind Suite data contains information from the balance sheets and income statements of all FDIC-insured institutions along with other financial, demographic and historical information. From the financial information in particular, it becomes very straightforward to identify a given bank's loan profiles, deposits, and other relevant financial dimensions commonly used in banking studies.

All variables other than the number of full-time equivalent (FTE) employees are stock variables denoted in thousands of US dollars (USD). Following Henderson et al. (2015), we convert our quarterly data to annual averages over years where four data from all fiscal quarters are available.[66] After converting our data to annual averages, we have a complete panel that extends from 1993 through 2021 containing 211916 observations.

For the purpose of our study, we use a mix of inputs and outputs from Wheelock & Wilson (2018), Restrepo-Tobón & Kumbhakar (2015) and Henderson et al. (2015) described below in Table 40. The outputs, $\mathcal{Y}$, contain three major loan categories used in nearly all banking performance studies: consumer loans (y_1), commercial and industrial loans (y_2), and real estate loans (y_3). Typically, these three loan categories account for the simple majority of the total loans issued by banking institutions. Beyond these loan categories, we have

[66]This is one of many ways to treat the data. While we could use quarterly data with no further temporal aggregation, taking the annual average over four quarters smooths out inexplicable variation quarter-to-quarter that can exist in the data, especially in the case of interest expenses. Papers like Restrepo-Tobón & Kumbhakar (2015) and Wheelock & Wilson (2018) use quarter four data only, as quarter four data can be thought of as the end-of-year annual equivalent. To reiterate parts of Berger & Mester (1997), irrespective of idiosyncracies study-to-study, the handling of commercial banking data falls into a "black box" wherein there are many potential strategies one may take for converting raw data into something suitable for empirical analysis.

two additional outputs: investment securities (y_4), and Fed Funds Sold (y_5). Our inputs, $\mathcal{X}$, are also fairly straightforward and consistent with the literature and include: purchased funds (x_1), total deposits (x_2), nontransaction accounts (x_3), full-time equivalent employees (x_4), and bank premises (x_5). Correspondingly, any given input price, $w_j \in \mathcal{W}$, is derived by dividing its total expense by its input quantity.

Table 40: Full Sample Descriptive Statistics

Variable	Description	10^{th} Pct.	25^{th} Pct.	Median	75^{th} Pct.	90^{th} Pct.	Mean	Pct. of Mean
C	Total Costs	781	1484	3109	7201	18650	29594	0.935
y_1	Consumer Loans	665	1757	4414	11104	31756	144509	0.964
y_2	C&I Loans	983	3168	9142	26078	77486	185823	0.952
y_3	Real Estate Loans	7167	19295	53745	147477	403329	462572	0.913
y_4	Securities	4088	10743	26391	65915	173753	297213	0.939
y_5	Fed Funds Sold	0	130	1826	5532	13843	56096	0.974
x_1	Purchased Funds	0	0	0	1740	15708	70162	0.957
x_2	Deposits	27227	51969	109975	257635	662931	1120808	0.938
x_3	Nontransaction Accounts	18289	35593	75652	187605	536658	792719	0.930
x_4	FTE Employees	10	17.8	36	80.5	196	257	0.924
x_5	Premises	243	714	2096	5488	13915	15277	0.909
w_1	Price of Purchased Funds	0	0	0	0.00959	0.0271	0.0135	0.788
w_2	Price of Deposits	0.00232	0.00509	0.013	0.0207	0.0249	0.0135	0.513
w_3	Price of Nontransaction Accounts	0	0	0.000225	0.00167	0.0042	0.0326	0.996
w_4	Price of FTE Employees	24.6	29	36	42.6	52.9	37.4	0.602
w_5	Price of Premises	0.0749	0.104	0.154	0.254	0.479	0.499	0.906
S_1	$(w_1 \times x_1)/C$	0	0	0	0.00401	0.0229	0.00879	0.818
S_2	$(w_2 \times x_2)/C$	0.129	0.244	0.428	0.567	0.649	0.407	0.467
S_3	$(w_3 \times x_3)/C$	0	0	0.00606	0.0437	0.0979	0.0329	0.698
S_4	$(w_4 \times x_4)/C$	0.247	0.307	0.408	0.579	0.698	0.444	0.566
S_5	$(w_5 \times x_5)/C$	0.0514	0.0716	0.0998	0.135	0.172	0.107	0.560
z_1	Nonperforming Assets	96.7	405	1390	4347	13093	25599	0.944

Note: $y_q \in \mathcal{Y}$ capture our vectors of outputs. $x_j \in \mathcal{X}$ encompass our vector of input quantities while $w_j \in \mathcal{W}$ correspond to each respective input price. Finally, $S_j \in \mathcal{S}$ are our cost share equations that are identified functionally as $(w_j x_j) \div C_i$, where C_i are the total costs for bank i. Finally, total costs are defined as the sum of the products of all input quantities times their respective prices such that $C_i = \sum_{j=1}^{J} w_j x_j$. Additionally, $z_{\#} \in \mathcal{Z}$ captures environmental variables utilized in this study. All variables with the exception of full-time employees are denoted in terms of thousands of US dollars in constant 2012 prices.

As echoed in Wheelock & Wilson (2018) and Restrepo-Tobón & Kumbhakar (2015), we observe from Table 40 a typical finding in the literature in that our data is heavily skewed to the right. As a consequence of our data cleaning procedure, in particular our allowance for $\{\mathcal{Y}, \mathcal{X}, \mathcal{W}\} \in \mathbb{R}^J_{\geqslant 0}$, the left tail of our data will predictably contain a fair amount of zeroes. We argue that this is not problematic and, in fact, closer to reality. Furthermore, given that the focus of our study is on banks most heavily involved in critical markets, the slim left tail of our data is of little concern to our empirical analysis and main results.

Given that our descriptive statistics describe *all* banks regardless of size or complexity, it is unsurprising that some banks rely on certain inputs much more than others and experience considerable variation in their individual input prices accordingly. Furthermore, small local banks or community banks necessarily possess portfolios that differ dramatically in scope and scale relative to large, transnational banks.

3.7 Model

For our empirical approach, we leverage a variation of the seemingly unrelated regressions model (SUR) originally developed in Zellner (1962). A SUR model can be generalized as a system of equations with some degree of cross-equation correlation. Via feasible generalized least squares (FGLS), more efficient estimation can be achieved than through the use of an analogous single-equation ordinary least squares (OLS) model. In most banking studies, the estimation of such a SUR system consists of a cost equation stemming from equation (3.1) along with $J - 1$ cost-share equations implied by the derivative conditions described by equation (3.4).

In theory, the cost-share equations capture the conditional factor demand for any one input x_j and, as such, the representative firm's cost minimization problem necessarily includes the formulation of their input factor demands alongside efforts in minimizing total costs. Works like Henderson et al. (2015) show that gains in precision of elasticity-based scale economies are possible when using a SUR system compared to a single-equation model. Furthermore, Kumbhakar (1997) illustrates that the estimation of certain performance indicators like allocative efficiency can only be achieved through the use of a dual-production [or dual-cost] system, which can only be formulated through the use of system-wise estimation techniques such as a SUR system.

Finally, as discussed at length in the seminal piece Christensen et al. (1973), in a production environment, the only true means to capture the necessary conditions for a producer's equilibrium is through a dual-production system wherein assumptions like symmetry, linear homogeneity, and other similar restrictions can be implemented for significant gains in a model's fit compared to a single-equation cost function. The duality nested in the theory of production necessitates a system-wise econometric specification, rather than single equation. It is for these reasons, we adopt a SUR-type estimator using a translog functional form.

3.7.1 A Brief Technical Overview of the SPSC SUR Estimator

We utilize a semiparametric smooth coefficient seeming unrelated regressions (SPSC SUR) in the form of a translog cost function as our primary econometric approach. The SPSC SUR builds upon the classic Zellner (1962) SUR model by allowing coefficients to vary locally across each equation in the SUR system based on a set of environmental factors, $\mathcal{Z}$.[67] It should be stressed that z is most important component of the SPSC SUR that differentiates it from the Zellner (1962) SUR model. The variation and gains and precision of overall fit are directly tied to variation in z, thus an important assumption going forward is that nonperforming assets as an environmental factor contains significant enough variation to produce a stronger

[67]We provide a detailed breakdown of the Henderson et al. (2015) SPSC SUR estimator in our appendix along with a replication exercise to illustrate its finite-sample properties.

model fit than a traditional linear SUR.

While one could specify more than one environmental factor, a limitation of nonparametric methods is that of the "curse of dimensionality" or slower computation times as the number of z vectors in $\mathcal{Z}$ increases. An additional important consideration for our estimation procedure is the nonparametric weighting of z, which is achieved via a kernel weighting scheme. Technically speaking, our kernel weighting function can be expressed as equation (3.16).

$$K_i(z) = K\left(\frac{z_i - z}{h}\right) \tag{3.16}$$

Where $K(\cdot)$ is our kernel density function and h is our bandwidth. Our bandwidth, h, is selected using least-squares cross-validation and our kernel smoothing function, $K(\cdot)$, is that of a Gaussian or normal kernel.[68] With these considerations in mind, our unrestricted OLS-type estimator takes on the following general form described by equation (3.17).

$$\hat{\beta}(z) = \left[\sum_{i=1}^{n} \tilde{x}_i^T \tilde{x}_i K_i(z)\right]^{-1} \left[\sum_{i=1}^{n} \tilde{x}_i^T \tilde{y}_i K_i(z)\right] \tag{3.17}$$

Where $\tilde{x}_i$ contains our independent variables for each equation in the system while $\tilde{y}_i$ is our stacked vector of dependent variables. Fundamentally, $\tilde{x}_i$ is $m \times l$ in dimension, where m is the maximum number of equations in our SUR system while n is our total number of observations. Finally, l captures the total number of variables across all equations. For example, if there are $m = 2$ equations with $n = 100$ observations each, and each equation contains two variables, we would have $l = 4$ total variables across all equations. Equation (3.18) illustrates the specific $\tilde{x}_i$ matrix associated with this example.

$$\tilde{x}_i = \begin{bmatrix} x_{11,n} & x_{12,n} & 0_n & 0_n \\ 0_n & 0_n & x_{21,n} & x_{22,n} \end{bmatrix} \tag{3.18}$$

Note that the dimension of equation (3.18) would be 2×4. The vectors, 0_n, contain zeroes of length n that partition the variables in our first equation uniquely from the variables in our second equation. Turning towards $\tilde{y}_i$, in a SUR system it is necessary to stack our vectors of dependent variables on top of one another. Continuing with our example described above, if there are $m = 2$ equations, then our $\tilde{y}_i$ matrix, which is $m \times 1$, would be 2×1. Equation (3.19) describes this.

[68] While there are alternative kernel functions that offer modest gains in statistical performance, the Gaussian kernel is incredibly robust and provides more appropriate data regularization for our study (Li & Racine, 2007). Furthermore, the Guassian kernel has an analytical solution to the integration of its second moment, making the calculation of observation-specific standard errors straightforward.

$$\tilde{y}_i = \begin{bmatrix} y_{1,n} \\ y_{2,n} \end{bmatrix} \tag{3.19}$$

While our unrestricted estimator described by equation (3.17) is sufficient for estimation sake, gains in estimation efficiency and precision can be gleaned by accounting for covariance between each equation in our SUR system, particularly if there is reason to believe that cross-equation correlation could exist. We can use our residuals from the first stage of our estimator to identify a variance-covariance matrix that can be used to produce an unrestricted generalized least squares (GLS) estimator described by equation (3.20).

$$\tilde{\beta}(z) = \left[X^T \mathbb{K}_z^{0.5} \Gamma^{-1} \mathbb{K}_z^{0.5} X\right]^{-1} \left[X^T \mathbb{K}_z^{0.5} \Gamma^{-1} \mathbb{K}_z^{0.5} Y\right] \tag{3.20}$$

Where $\Gamma^{-1} = \hat{\Sigma}^{-1} \otimes I_n$ and $\hat{\Sigma}^{-1}$ is our the inverse of our variance-covariance matrix. I_n is an $n \times n$ identity matrix, and $\otimes$ denotes the Kronecker product.[69] Formally, following our example once more, if there are $m = 2$ equations in our system, then we will produce a stacked vector of residuals, U_i that is also $m \times 1$ in length. Subsequently, U_i can be partitioned to produce $\hat{u}_1$ and $\hat{u}_2$, which are the residuals unique to each equation, for $m = 2$ equations. $\hat{\Sigma} = cov(\hat{u}_1, \hat{u}_2)$, which is $m \times m$ or 2×2 in dimension. Equation (3.20) is our baseline unrestricted cross-equation correlation consistent estimator that is typical for SUR models while allowing for nonparametric variation in z across our coefficients.

3.7.2 Econometric Specification

We specify a normalized translog cost function described by equations (3.21) and (3.22) wherein we allow the coefficients of our regressors to vary by a common environmental factor, z. Following our motivation and methodology, we specify z to be the nonperforming assets of a given bank i.

$$\begin{aligned} \log(C_i/w_{5i}) &= \alpha_0(z_i) + \sum_{j=1}^{4} \beta_j(z_i) \log(w_{ji}/w_{5i}) + \sum_{q=1}^{5} \gamma_q(z_i) \log(y_{qi}) \\ &+ \frac{1}{2} \sum_{q=1}^{5} \sum_{l=1}^{5} \gamma_{ql}(z_i) \log(y_{qi}) \log(y_{li}) + \frac{1}{2} \sum_{j=1}^{4} \sum_{v=1}^{4} \eta_{jv}(z_i) \log(w_{ji}/w_{5i}) \log(w_{vi}/w_{5i}) \\ &+ \sum_{j=1}^{4} \sum_{q=1}^{5} \delta_{jq}(z_i) \log(w_{ji}/w_{5i}) \log(y_{qi}) + u_{ji} \end{aligned} \tag{3.21}$$

[69] When applying our estimator to the data, it is not uncommon to encounter det $\left[X^T \mathbb{K}_z^{0.5} \Gamma^{-1} \mathbb{K}_z^{0.5} X\right] = 0$ or issues with invertibility. To remedy this, we adopt the least-additive approach to perturbing diag $\left[X^T \mathbb{K}_z^{0.5} \Gamma^{-1} \mathbb{K}_z^{0.5} X\right]$ by an arbitrarily small quantity $\zeta = n^{-1}$ as needed.

$$S_{ji} = \alpha_j(z_i) + \sum_{v=1}^{4} \eta_{jv}(z_i)\log(w_{vi}/w_{5i}) + \sum_{q=1}^{5} \delta_{jq}(z_i)\log(y_{qi}) + u_{ji} \text{ for } j \in \{1,2,3,4\} \tag{3.22}$$

As we can see, equation (3.21) is equivalent to (3.1) from our production theory setup, but given a functional form. Subsequently, the cost share equations described by (3.22) are also analogous to equation (3.4) from our setup. It also should be stressed that the five-equation model described by both equations (3.21) and (3.22) are estimated via equation (3.20). Beyond our principle SUR system described above, we take additional measures to ensure that our econometric specification falls in-line with its structural underpinnings. These additional measures consist of: normalization to ensure linear homogeneity and the imposition of cross-equation restrictions to emulate features of equations (3.1) and (3.4). Normalization is achieved by dividing our input prices and total costs by a common input price. The imposition of cross-equation restrictions is achieved through the use of a design matrix, R, which is $J \times ml$ in dimension, where J is the total amount of restrictions across coefficients, and a vector r, which is $J \times 1$, and contains the restrictions themselves. By this design, restrictions are imposed formally by setting $R\tilde{\beta}(z) = r$.[70]

Normalization

It is common practice to normalize our prices, w_j, and our total cost, C_i, by a common price. The motivation for this normalization procedure is twofold: firstly, as highlighted in pieces like Embrey (2019), normalization of production [cost] functions in a reduced form ensures linear homogeneity, which is important for both empirical and theoretical consistency; secondly, given that the sum over all J cost share equations equals unity, issues of multicollinearity can arise if our SUR system includes our cost function and all five share equations implied by the derivative conditions, thus, via this normalization procedure, it becomes necessary to drop the S_j equation that corresponds to the normalizing w_j price.

For our application, we normalize by w_5, the input price associated with firm premises. Given that premises are a proxy for fixed costs, there is not as much variation year-to-year for any i bank compared to other input prices, thus, w_5 is a natural choice as our normalizing point.[71]

Restrictions

In theory, there are no new coefficients in our cost share equations described by (3.22), thus we impose cross-equation equality restrictions on η_{jv} and δ_{jq} to satisfy the derivative conditions implied by Shephard's Lemma. Furthermore, we impose symmetry restrictions via Young's Theorem such that $\gamma_{ql} = \gamma_{lq}$, $\eta_{jv} = \eta_{vj}$, and $\delta_{jq} = \delta_{qj}$, which are satisfied through the estimation procedure itself.

[70] In the case of equality restrictions, $r = 0$.

[71] In our appendix, we show kernel density plots of our normalized variables of interest.

Our cross-equation equality restrictions are achieved through the utilization of a design matrix, R, and a vector, r, consisting of the constraints themselves. If our constraints are equality constraints ($r = 0$), then we can rewrite our estimator as in equation (3.23).[72]

$$\tilde{\beta}^*(z) = \Bigg(I_{ml} - \bigg[X^T \mathbb{K}_z^{0.5} \Gamma^{-1} \mathbb{K}_z^{0.5} X\bigg]^{-1} R^T \times \bigg[R(X^T \mathbb{K}_z^{0.5} \Gamma^{-1} \mathbb{K}_z^{0.5} X)^{-1} R^T\bigg]^{-1} R\Bigg)\tilde{\beta}(z) \tag{3.23}$$

Where I_{ml} is an identity matrix wherein m is the number of equations in the SUR system and l represents the total amount of coefficients across all equations.

Additional Notes on Translog Functions

We take this space to acknowledge that translog functions are not without criticism. In particular, there are several papers that advocate against the use of linear system-wise translog cost function estimations due to statistical biases and poor global fits (McAllister & McManus, 1993; White, 1980; Mitchell & Onvural, 1996). These findings have led to a shift in model preferences towards more flexible cost function specifications, including entirely nonparametric cost functions. While relaxing the rigidity of OLS translog functions is useful for improving a given model's fit, the interpretability of results become less clear, particularly when using fully nonparametric methods.

We believe that the SPSC SUR for the purpose of deriving returns-to-scale more than overcomes these concerns and strikes the best balance between relaxing the strong assumptions of OLS while remaining model tractability. For starters, by using a semiparametric model, we allow for far more flexibility in our estimation procedure by allowing coefficient estimates to vary smoothly observation-to-observation in accordance to our environmental factor, z. This alone represents a vast improvement over linear SUR systems that do not allow for such flexibility—in fact, under the assumption that z does not vary over each z_i, the SPSC SUR collapses to a traditional linear SUR. Secondly, single-equation cost functions, such as the flexible fourier form, do not allow for cross-equation restrictions, which are implied via Shepherd's Lemma. As shown in Henderson et al. (2015), SPSC SUR models with cross-equation restrictions that satisfy theory can offer considerable improvement in the fit of each individual equation model coefficient that RTS estimates are built from. Given that returns-to-scale estimates are often elasticity-based, cross-equation restrictions offer a strict improvement in returns-to-scale estimates from coefficients facing such restrictions compared to similar unrestricted models.

[72] See our appendix for a simple version of our estimator, including an explicit cross-equation restriction illustration.

3.7.3 Estimating Scale Economies in a Reduced Form

A translog cost specification such as equation (3.21) allows for scale economies to be obtained for any given bank i as the reciprocal sum of its output elasticities, $\gamma_q(z)$, consistent with the formulation of scale economies described by equations (3.5) through (3.13). Denote $\tilde{\varepsilon^y}_i$ as our estimated returns to scale for bank i described by equation (3.24).

$$\tilde{\varepsilon^y}_i = \left(\sum_{q=1}^{5} \left\{ \frac{\partial \log(C_i/w_{5i})}{\partial \log(y_{qi})} \right\} \right)^{-1} \tag{3.24}$$

Equivalently, we can capture scale economies for a given bank i following the expression described by equation (3.14).

$$\tilde{\xi^y}_i = 1 - \tilde{\varepsilon^y}_i \tag{3.25}$$

An advantage that the Henderson et al. (2015) estimator offers when calculating $\tilde{\xi^y}_i$ is that economies of scale can be thought of as a distribution of estimates that vary with z, rather than a global approximation across all banks. Given how heterogeneous banking institutions are, estimates of economies of scale and the degree to which TBTF banks or otherwise capitalize on scale economies should be as heterogeneous.

3.8 Results & Discussion

Despite efforts to separate critical market from non-critical market banks, there is still a possibility of encountering anomalies in any given subset of banks in a given year. These anomalies could arise from data entry errors or extreme values among other reasons. To rectify this, we purge all outlying returns-to-scale estimates post-estimation. Table 41 shows the evolution of nonperformance-adjusted returns-to-scale estimates from 2001 through 2021. Standard errors for returns-to-scale estimates are constructed in accordance with Henderson et al. (2015) and Geng & Sun (2022) through the construnction of an asymptotic variance-covariance matrix.[73]

[73]Note, as is the norm in a lot of banking RTS studies, particularly ones that use nonparametric or partially linear models (which the SPSC SUR is), RTS estimates are reported in terms of quartiles (Henderson et al., 2015; Geng & Sun, 2022). Standard errors for varying coefficients themselves are also seldom reported. These caveats notwithstanding, we do provide in our appendix two approaches one could use to derive standard errors for a given varying coefficient. Additionally, all constant returns-to-scale numbers are derived by counting each bank whose returns-to-scale estimates are not statistically different from unity in a given year.

Table 41: Nonperformance-Adjusted RTS Estimates for Critical Market Banks

Year	Lower Quartile	Median	Upper Quartile	Mean (μ_t)	Std. Dev. (σ_t)	Obs. (N)	DRS	CRS	IRS
2001–2021	0.59	1.26	1.77	1.19	0.95	6108	3384	3715	3486
2001	1.08***	1.58***	2.43***	1.76	0.95	664	109	437	118
2002	0.79***	1.35***	1.86***	1.33	0.71	523	114	123	286
2003	0.94***	1.08***	1.35***	1.19	0.43	514	43	336	135
2004	-1.39***	0.56*	0.89*	-0.25	1.88	722	464	144	114
2005	0.88***	1.06***	1.25***	1.06	0.27	564	43	440	81
2006	0.72***	2.15***	3.10***	2.14	1.90	650	134	69	447
2001–2006	0.68	1.13	1.86	1.18	1.50	3637	907	1549	1181
2007	0.54*	0.76**	1.36***	0.99	0.55	595	247	193	155
2008	0.39	0.73**	1.57***	1.02	1.16	568	269	123	176
2009	-0.59	0.72**	1.51***	0.59	1.27	598	292	138	168
2007–2009	0.47	0.75	1.49	0.86	1.06	1761	808	454	499
2010	0.74**	0.95***	1.23***	1.04	0.43	463	76	286	101
2011	0.41	0.45*	0.54*	0.53	0.25	456	373	83	0
2012	0.54*	0.67**	1.45***	1.01	0.72	442	211	113	105
2013	0.94***	1.73***	2.50***	1.83	1.04	527	29	179	319
2014	-2.67***	0.81**	1.30***	-0.44	2.61	516	250	138	128
2015	0.50*	0.57**	1.34***	1.02	0.93	470	261	88	121
2016	1.11***	1.46***	1.65***	1.40	0.38	408	20	121	267
2017	0.80***	2.27***	3.60***	3.65	5.10	344	75	42	227
2018	0.38*	0.81***	2.94***	0.66	2.70	413	146	109	158
2019	1.17***	1.72***	2.50***	1.77	0.72	421	33	113	275
2010–2019	0.52	1.03	1.79	1.17	2.18	4460	1474	1272	1701
2020	0.40	0.94***	1.49***	0.54	1.68	370	112	153	105
2021	0.69**	0.81**	0.89***	0.79	0.11	340	83	287	0
2020–2021	0.67	0.86	0.97	0.66	1.19	710	195	440	105

Note: ***$p < 0.01$; **$p < 0.05$; *$p < 0.10$. Our results present RTS estimates described by equation (3.24) that are derived from our cost equation, (3.21), for all years from 2001 through 2021 meaning our results consist of estimates from 21 unique SPSC SUR models indexed from years $t = 2001$ through $T = 2021$. Subsample estimates are computed through aggregation of all index years of interest. While it is possible to apply our model over longer samples or as a true panel model, estimate convergence times and processing requirements limit the feasibility of this option. Critical market banks identified fall in the upper decile (90^{th} percentile) for relative critical market involvement of all banks in any given year, thus there is an allowance for the possibility that the total number of banks operating in critical markets most intensely can vary as the competitive landshape shrinks.

Echoing results found in Henderson et al. (2015), Restrepo-Tobón & Kumbhakar (2015) and Wheelock & Wilson (2018), the median returns-to-scale estimates for critical market banks hover, on average, slightly above one, thus, over long periods of time, most critical market banks exhibit slight increasing returns-to-scale in spite of nonperforming assets and their impact on the production environment. However, year-to-year variation can be quite large. During the Global Financial Crisis, most critical market banks expectedly were exhibiting decreasing returns-to-scale and did not fully revert back to an increasing returns-to-scale state until 2013 only to fall again during 2014 and 2015. Another subsequent reversion to an increasing returns-to-scale state occurred from 2016 through Covid-19's onset. For the purpose of evaluating the distribution of all returns-to-scale estimates over all years, consider illustrative results described by Figure 29.

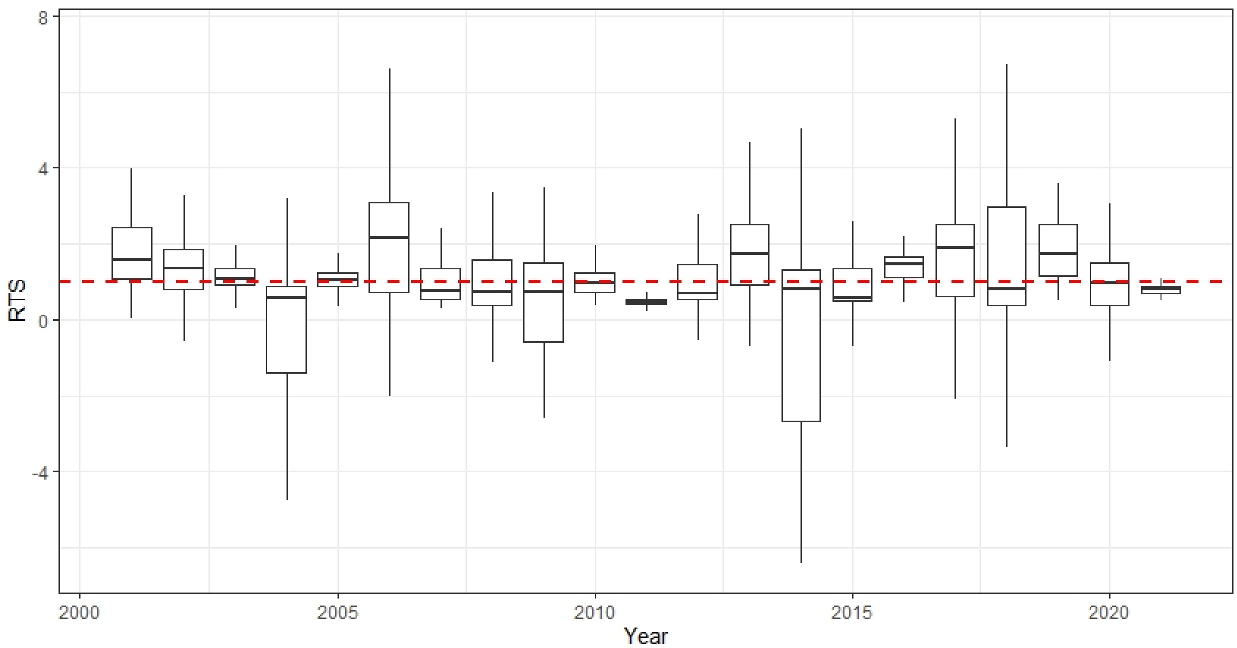


Figure 29: Boxplot of Nonperformance-Adjusted RTS Estimates

We note that, despite consistent median returns-to-scale results, extremities exist at the upper and lower quartiles, respectively. This is not abnormal. In fact, standard error estimates of returns-to-scale estimates tend to be quite large in most banking studies (if reported at all). We note some extreme negative returns-to-scale estimates in the years of 2004 and 2014. These lower quartile estimates in particular should be taken with caution. The implication of negative returns-to-scale estimates is that a unit increase in total costs results in a *reduction* of total output in real terms. As noted in Wheelock & Wilson (2018), these results are rare, but can occur. Interestingly, despite the median critical market bank expressing small increasing returns-to-scale over our sample, most years express median returns-to-scale that are $\leqslant 1$.

The results described by Figure 29 and Table 41 partially address our first motivating question. It would appear that there are strong distributional effects that nonperforming assets can have on the production environment for critical market banks. In most years, there are a minority of banks exhibiting increasing returns-to-scale in spite of nonperforming asset holdings while the majority of critical market banks in any given year are expressing mild evidence of decreasing or constant returns-to-scale. While we look at a different sample of banks and implement a distinct research design, these findings are not far from pieces like Restrepo-Tobón & Kumbhakar (2015) who suggest that mild increasing returns-to-scale are experienced by the median large bank.

Interestingly, however, other works like Sapci & Miles (2019) find that most banks seem to exploit increasing returns-to-scale until they grow "too large." However, Sapci & Miles (2019) points out that banks who have exhausted their returns-to-scale tend to retain their cost efficiencies. Taken with our results and in the context of our study, it is possible that critical market banks who are experiencing decreasing or constant returns-to-scale in the past five years have retained some degree of cost efficiency, although further research is needed to verify this. Figure 30 shows the precise count of increasing, constant, and decreasing returns-to-scale critical market banks in any given year.

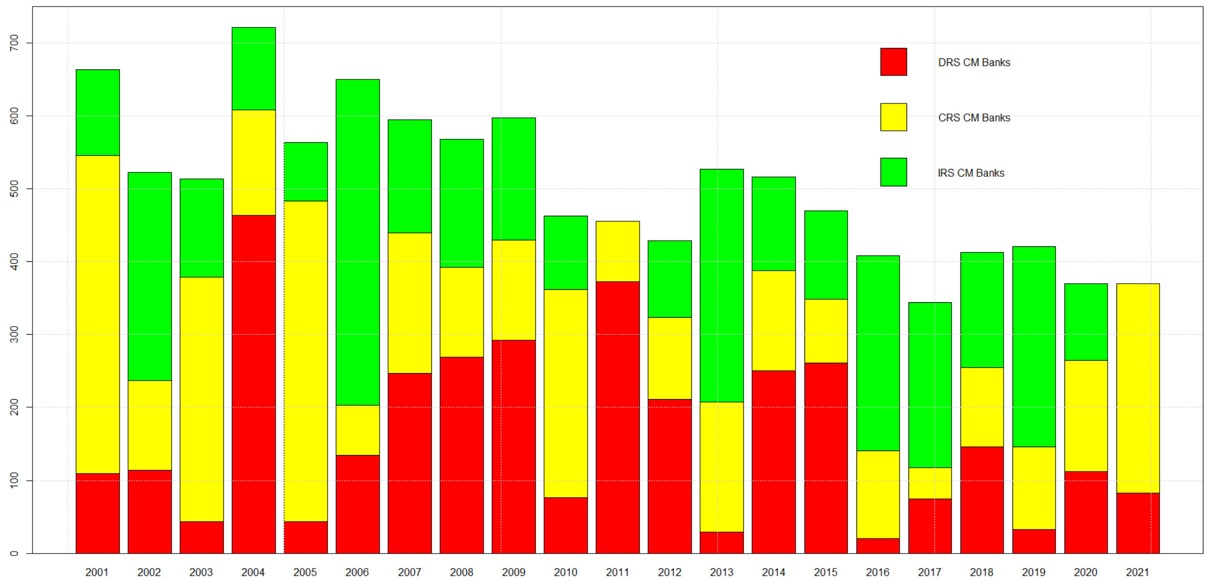


Figure 30: Share of Scale Economies of Critical Market Banks by Year

In terms of addressing whether or not TBTF banks stand to gain from engaging in moral hazard, our nonperformance adjusted returns-to-scale estimates for the upper quartile suggest that *some* critical market banks with large holdings of nonperforming assets are in fact continuing to express increasing returns-to-scale, however, these results should be taken with a grain of salt given the wide-ranging standard deviation of returns-to-scale estimates for any given year—a common finding in the literature as well. It is also possible that there are banks involved in critical markets with very few nonperforming assets, thus critical market banks operating in the upper quartile for nonperformance-adjusted returns-to-scale for any given year may not be engaging in much moral hazard to begin with. Furthermore, Figure 30 further illustrates that the shrinking competitive landscape has diminished opportunities for critical market banks to capitalize on returns-to-scale (in fact, in 2011, and 2021, there are zero critical market banks in our sample that express increasing returns-to-scale).

Finally, it is also worth highlighting that our study is one of the first to examine changes in returns-to-scale since the enactment of Economic Growth, Regulatory Relief, and Consumer Protection Act, which represents the largest overhaul to banking regulation since the Dodd-Frank Act. Very few economic studies have looked specifically at the period of time since its enactment in 2018, thus our results provide some foundational insights that can be built upon by future work. In many ways, the Economic Growth, Regulatory Relief, and Consumer Protection Act functions as a relaxation of the regulatory restrictions imposed by Dodd-Frank.[74] The changes in regulatory burdens and compliance costs may very well effect moral hazard associated with the economies of scale facing critical market banks, thus, our findings for the years of 2018 through 2021 provide some insight into how scale economies have evolved under these relaxed regulatory standards.

While the degree of meaningful inference is confounded by the Covid-19 pandemic, we note that in 2018, the median critical market bank was experiencing decreasing returns to scale at 0.81 and rebounds considerably to 1.72 in 2019. Furthermore, since the introduction of the Dodd-Frank Act in 2011, we observe that most years (with the exception of 2013, 2016, and 2017) show the median critical market bank expressing decreasing returns-to-scale. While 2020 and 2021 show the median critical market bank expressing decreasing returns-to-scale, it is worth noting that the estimates themselves are comparatively larger than the degree of decreasing returns-to-scale faced in the aftermath of the Dodd-Frank Act in the years of 2011 and 2012.

Not incidently, it is possible that the relaxation of regulatory standards in 2018 could have buffered critical market banks engaged in moral hazard from experiencing potentially lower decreasing returns-to-scale levels during the Covid-19 pandemic. Unfortunately, the absence of economic studies examining the impact of Economic Growth, Regulatory Relief, and Consumer Protection Act makes generalizing our results difficult. The need for external validation of our nonperformance-adjusted returns-to-scale estimates from 2018-2021 is a gap best left for future studies along with more focus on the evolution of scale economies in the wake of Covid-19 across banks of all sizes and irrespective of critical market involvement.

3.9 Concluding Remarks

Returning to our motivation, we posed three main questions regarding the performance puzzles pervasive in the TBTF literature. Revisiting these questions, we first asked ourselves *how does moral hazard impact banking performance*? We look at performance through the lens of returns-to-scale and find that over the past two decades, in the aggregate, nonperforming assets do not seem to inhibit the median critical market bank from capitalizing on increasing returns-to-scale. However, year-by-year, the total number of increasing returns-to-scale banks has fallen considerably. In fact, from 2019 through 2021 alone, only the upper quartile of critical market banks seem to exhibit increasing returns-to-scale. One explanation for this is that as the competitive landscape has diminished, fewer opportunities for mergers, acquisitions, and consolidation

[74] Specifically, the Dodd-Frank Act established a threshold for discerning banks that might be TBTF at $50 billion while the Economic Growth, Regulatory Relief, and Consumer Protection Act raised the threshold to $250 billion.

present themselves for otherwise risky banks to capitalize on. Additionally, it is possible that the effects of moral hazard and insolvency do not have a significant contemporaneous impact on returns-to-scale, but could have a considerable dynamic or lagged impact. Finally, there is some evidence from the community banking literature that policies like the Dodd-Frank Act drove strong changes in the cost structure of commercial banks through increased compliance costs, among other burdens (Peirce et al., 2014; Liu, 2019; Minuci, 2021). As a result, it is possible that the low-marginal cost environment the banking industry has previously enjoyed is less favorable since the end of the Global Financial Crisis through the present day. Future research would do well to focus on these shocks to the cost structure of US banks.

Our second question asked *what banks are capable of engaging in moral hazard set by bailout expectations*? This question is more philosophical than anything. Realistically, we cannot observe a bank's leadership and intentions with regards to risky portfolio decisions they may make, however, we postulate that while large banks might be more apt to engage in TBTF risk, they certainly are not the only types of banks that might take excess risks with policy intervention expectations in mind. We believe that critical market banks, which are *mostly* large, are more likely to engage in TBTF moral hazard. We acknowledge that our cutoff (90^{th} percentile) for identifying banks with the highest involvement in critical markets may be somewhat arbitrary, but so are preexisting definitions posed by the FRB that have previously been adopted in other studies to answer this exact same question. If anything, our definition is probably more *inclusive* than *exclusive*. However, a convenient feature of our estimator is that by applying a kernel weight to each observation, we can afford to be more inclusive of banks that might not be TBTF or highly involved in critical markets by assigning low relative weights to those observations. Additionally, if one were interested in the 95^{th}, or 99^{th} percentile as a cutoff for critical market banks, we could simply censor all observations post-estimation that fall below the percentile of interest–this feature is another convenience afforded to us by the SPSC SUR estimator.

Our final question is related to our first and asks *if TBTF banks stand to gain much from engaging in moral hazard*? Looking at our median nonperformance-adjusted returns-to-scale estimates, we see that most TBTF banks exhibit modest increasing returns-to-scale (1.26). This number may suggest that critical market banks actually stand to gain from engaging in moral hazard. However, this number is not too far from results presented in Henderson et al. (2015) who take a random sample of 3,112 banks from 2010 alone and find median returns-to-scale to be 1.049. Wheelock & Wilson (2018) only look at the largest US banks and find returns-to-scale for the most recent quarter in their sample to range from 0.96 to 1.15. Restrepo-Tobón & Kumbhakar (2015) look at large commercial banks from 2000 through 2010 and find that only a minority express increasing returns-to-scale and, even so, such gains are small. To a degree, our findings echo all three of these previous studies, but capitalize on a different definition of TBTF banks and leverage a strictly better and more flexible functional form that incorporates moral hazard directly into the coefficients of our reduced form model from which returns-to-scale are derived.

While some critical market banks have enjoyed increasing returns-to-scale early in our sample, in recent years, only a small minority capitalize on scale economies. In the past decade alone, the vast majority of critical market banks are operating under decreasing or constant returns-to-scale. From a policy standpoint, these results imply that the degree to which regulatory statutes such as the Dodd-Frank and Consumer Protection Acts inhibit or disproportionately suppress banks from achieving economies of scale is minimal, if at all.

3.10 Appendix A: Details on SPSC SUR Estimator

The SPSC SUR estimator described by Henderson et al. (2015) fundamentally marries together features of the seminparametric smooth coefficient model described in Hastie & Tibshirani (1993) with the seemingly unrelated regressions model first penned by Zellner (1962).[75]

3.10.1 General Setup for SPSC SUR

The benchmark linear SUR model was originally penned by Zellner (1962) and is generalized by equation (3.26).

$$y_{si} = x_{si}^T \beta_s + u_{si} \text{ for } s = 1, ..., m \text{ and } i = 1, ..., n \quad (3.26)$$

y_{si} is an $mn \times 1$ vector of stacked dependent variables from $s = 1$ equations to a total of m equations. Each stacked y_{si} variable contains observations from $i = 1$ to a total of n. x_{si} is an $mn \times l$ stacked matrix of independent variables where l is the total amount of coefficients across all equations to be estimated. Finally, β_s is our an $l \times 1$ vector of stacked coefficient estimates for all m equations. Finally, u_{si} is our residual error vector that is $mn \times 1$. Typically, a SUR model is estimated via generalized least squares to account for the cross-equation correlation in the error term. To implement this, the SUR model is estimated via ordinary least squares to obtained the residual vector, u_{si}, to construct the variance-covariance matrix. The variance-covariance matrix is then used in a second-stage SUR model to perform the a similar estimation procedure, but with proper consideration of the covariance between the cross-equation error terms.

With this simple setup in mind, the Henderson et al. (2015) SPSC SUR estimator is described by equation (3.27).

$$y_{si} = x_{si}^T \beta_s(z_{si}) + u_{si} \text{ for } s = 1, ..., m \text{ and } i = 1, ..., n \quad (3.27)$$

The same characteristics of the Zellner (1962) are still in place in equation (3.27), however, our coefficient vector, β_s is a function of z_{si}. Thus, the vector of coefficients produced are $n \times l$ in length, rather than $l \times 1$. Each coefficient from $1, ..., l$ will have n varying coefficient estimates determined via a kernel weighting function applied to z_{si}.[76]

[75] The semiparametric smooth coefficient model is also known as a *varying coefficient model*. See Li & Racine (2007) for explicit details on the varying coefficient model setup.

[76] While one could allow different z terms to exist across equations, a limitation of any nonparametric or semiparametric procedure is the curse of dimensionality or the slow convergence of our estimates with each subsequent addition of other z terms. As a result, for model implementation, it is best to specify one z common across all equations.

3.10.2 Replication Exercise

Given that the SPSC SUR estimator is a relatively new innovation in the nonparametric and semiparametric econometrics literature, there are no explicit built-in packages for traditional econometric software, nor is there publicly available source code for the entirety of the estimator, thus, we build it from scratch carefully following the methodology described by Henderson et al. (2015) and reproduce their simple two-equation model and Monte-Carlo simulation results for the relative performance of the restricted versus unrestricted estimator. This is done to ensure that the construction and performance of our SPSC SUR estimator closely corresponds to Henderson et al. (2015).

To begin this replication, consider a two equation model with two sets of coefficients per equation. Assume that $z_{1i} = z_{2i} = ... = z_{mi} = z$ (the same environmental variable appears in all m equations).

$$y_{1i} = b_{11}(z)x_{1i,1} + b_{12}(z)x_{1i,2} + u_{1i} \tag{3.28}$$

$$y_{2i} = b_{21}(z)x_{2i,1} + b_{22}(z)x_{2i,2} + u_{2i} \tag{3.29}$$

Consider that we know the underlying data generating processes.

$$b_{11}(z) = 3z,\ b_{12}(z) = b_{21}(z) = sin(z), \text{ and } b_{22}(z) = z^3$$

Wherein u_{1i} is IID $\sim N(0,1)$ and $u_{2i} = c \times u_{1i} + v_{2i}$ where v_{2i} is IID $\sim N(0,1)$ and c is the cross-equation correlation coefficient. Furthermore, x_{1i} is IID $\sim U(0,2)$, and x_{2i} is IID $\sim U(0,1)$. Finally, assume our environmental variable, z_i, is common between both equations and is IID $\sim U(-3,2)$.

With this in mind, we can write our unrestricted estimator as $\hat{\beta}(z) = \left[\sum_{i=1}^{n} \tilde{x}_i^T \tilde{x}_i K_i(z)\right]^{-1} \left[\sum_{i=1}^{n} \tilde{x}_i^T \tilde{y}_i K_i(z)\right]$ or $\hat{\beta}(z) = D_z^{-1} \times A_z$. Equations (3.30) through (3.34) describe in explicit detail each component of the unrestricted estimator given our two-equation setup.[77]

$$\tilde{y}_i = \begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix} \tag{3.30}$$

$$\tilde{x}_i = \begin{bmatrix} x_{1i,1} & x_{1i,2} & 0 & 0 \\ 0 & 0 & x_{2i,1} & x_{2i,2} \end{bmatrix} \tag{3.31}$$

[77] Our bandwidth, h, is selected via least-squares cross-validation (LSCV). Furthermore, our kernel density function described by equation (3.32) is specified as a *normal* kernel.

$$K_i(z) = K_{zi,z} = K\left(\frac{z_i - z}{h}\right) \tag{3.32}$$

$$D_z = \begin{bmatrix} \sum_{i=1}^n x_{1i,1}^2 K_i(z) & \sum_{i=1}^n x_{1i,1}x_{1i,2}K_i(z) & 0 & 0 \\ \sum_{i=1}^n x_{1i,2}x_{1i,1}K_i(z) & \sum_{i=1}^n x_{1i,2}^2 K_i(z) & 0 & 0 \\ 0 & 0 & \sum_{i=1}^n x_{2i,1}^2 K_i(z) & \sum_{i=1}^n x_{2i,1}x_{2i,2}K_i(z) \\ 0 & 0 & \sum_{i=1}^n x_{2i,2}x_{2i,1}K_i(z) & \sum_{i=1}^n x_{2i,2}^2 K_i(z) \end{bmatrix} \tag{3.33}$$

$$A_z = \begin{bmatrix} \sum_{i=1}^n x_{1i,1}y_{1i}K_i(z) \\ \sum_{i=1}^n x_{1i,2}y_{1i}K_i(z) \\ \sum_{i=1}^n x_{2i,1}y_{2i}K_i(z) \\ \sum_{i=1}^n x_{2i,2}y_{2i}K_i(z) \end{bmatrix} \tag{3.34}$$

Our fitted values, $\tilde{y}_i^f$, can be obtained by obtained as $\sum_{i=1}^n \tilde{x}_i \odot \hat{\beta}(z)$. Thus, our residual vector is by definition $\hat{u}_i = \tilde{y}_i - \sum_{i=1}^n \tilde{x}_i \odot \hat{\beta}(z)$. Now that we've obtained our residual vector, we can identify our variance-covariance matrix as $\hat{\Sigma} = cov(\hat{u}_1, \hat{u}_2)$. With out fitted values in mind, we can leverage generalized least squares (GLS) to account for the covariance and correlation in the residuals of each regression equation. The second step of producing our GLS-type estimator starts by defining Γ^{-1} as $\hat{\Sigma}^{-1} \otimes I_n$. Furthermore, define $\mathbb{K}_z$ as $I_m \otimes K_i(z)$ and let our stacked y_{si} and x_{si} observations be denoted as Y and X, respectively. With this in mind, we can describe our "true" estimator via equation (3.35).

$$\tilde{\beta}(z) = \left[X^T \mathbb{K}_z^{0.5} \Gamma^{-1} \mathbb{K}_z^{0.5} X\right]^{-1} \left[X^T \mathbb{K}_z^{0.5} \Gamma^{-1} \mathbb{K}_z^{0.5} Y\right] \tag{3.35}$$

A key feature of SUR models is their ability to allow for cross-equation restrictions to better match the underlying economic theory within the system. The SPSC SUR is no different in this regard. Specifically, in a SUR system of any form, we can impose restrictions such that $R\beta(z) = r$ in order to simulate economic theory. In our efforts to replicate Henderson et al. (2015), we impose $b_{12}(z) = b_{21}(z)$. When imposed on our existing estimator, $\tilde{\beta}(z)$, we are arrive equation (3.36).[78]

[78] To be explicit, we have $R = [0, 1, -1, 0]$ and $r = 0$.

$$\begin{aligned}\tilde{\beta}^{*}(z) = \Bigg(I_{ml} - \left[X^T \mathbb{K}_z^{0.5} \Gamma^{-1} \mathbb{K}_z^{0.5} X\right]^{-1} R^T \\ \times \left[R(X^T \mathbb{K}_z^{0.5} \Gamma^{-1} \mathbb{K}_z^{0.5} X)^{-1} R^T\right]^{-1} R\Bigg)\tilde{\beta}(z)\end{aligned} \tag{3.36}$$

Where I_{ml} is an $m \times l$ identity matrix, where l is the maximum number of coefficients in a given equation, thus $I_{ml} = 2 \times 2$. Following Henderson et al. (2015), we present Monte-Carlo simulations results of the relative performance of the unrestricted to restricted estimator. Performance is evaluated via the ratio of the average square error (ASE) across each coefficient in our system as well as the conditional mean of our individual equations. Equation (3.37) describes this evaluation criteria.

$$n^{-1} \sum_{i=1}^{n} \left(\hat{b}_{sl}(z) - b_{sl}(z)\right)^2 \text{ for } s \text{ equations and } l \text{ coefficients} \tag{3.37}$$

Finally, we allow for varying degrees of cross-equation correlation and differing sample sizes. Table 42 reports the results of our Monte-Carlo simulations.

Table 42: ASE Ratio of Unrestricted to Restricted Model

c	n	$b_{11}(z)$ 10 Pct.	$b_{11}(z)$ 50 Pct.	$b_{11}(z)$ 90 Pct.	$b_{12}(z)$ 10 Pct.	$b_{12}(z)$ 50 Pct.	$b_{12}(z)$ 90 Pct.	$b_{21}(z)$ 10 Pct.	$b_{21}(z)$ 50 Pct.	$b_{21}(z)$ 90 Pct.	$b_{22}(z)$ 10 Pct.	$b_{22}(z)$ 50 Pct.	$b_{22}(z)$ 90 Pct.	Function 10 Pct.	Function 50 Pct.	Function 90 Pct.
0	100	0.87	1.06	1.34	0.78	1.18	1.71	1.87	5.54	15.73	0.84	1.21	1.90	0.98	1.00	1.02
	200	0.87	1.07	1.36	0.86	1.21	1.72	2.41	5.44	13.75	0.87	1.20	1.77	0.99	1.00	1.01
	500	0.88	1.08	1.33	0.88	1.21	1.66	2.80	5.37	10.51	0.92	1.25	1.76	0.99	1.00	1.01
0.25	100	0.86	1.06	1.37	0.80	1.14	1.68	1.82	5.67	16.00	0.83	1.20	1.90	0.98	1.00	1.02
	200	0.87	1.08	1.38	0.83	1.18	1.66	2.45	5.43	13.70	0.86	1.20	1.77	0.99	1.00	1.01
	500	0.87	1.09	1.36	0.88	1.18	1.59	2.78	5.45	10.71	0.91	1.25	1.82	0.99	1.00	1.01
0.50	100	0.87	1.06	1.35	0.81	1.12	1.60	1.98	5.87	17.34	0.83	1.20	1.87	0.98	1.00	1.02
	200	0.87	1.07	1.36	0.84	1.14	1.54	2.52	5.93	14.08	0.86	1.20	1.78	0.99	1.00	1.01
	500	0.88	1.09	1.35	0.88	1.14	1.49	2.92	5.89	11.70	0.91	1.24	1.85	0.99	1.00	1.01
1	100	0.88	1.05	1.32	0.85	1.08	1.42	2.50	7.38	21.62	0.85	1.19	1.83	0.98	1.00	1.03
	200	0.88	1.06	1.30	0.87	1.07	1.33	3.18	7.70	18.63	0.87	1.19	1.78	0.98	1.00	1.02
	500	0.89	1.07	1.30	0.90	1.06	1.27	3.95	8.21	15.81	0.92	1.23	1.78	0.99	1.00	1.01

Consistent with Henderson et al. (2015), we find the gains in our restricted estimator are strongest across each individual coefficient, rather than the conditional mean. Our performance results our also on-magnitude similar to Henderson et al. (2015). This replication exercise illustrates the simplest version of the SPSC SUR model that is extended for our main results and subsequent analysis.[79]

3.10.3 Standard Errors in the Restricted SPSC SUR Model

Despite the varying nature of our SPSC SUR estimates, one can also obtain observation-specific standard errors for each varying coefficient. To achieve this, we emulate the methodologies employed in both Henderson

[79]Obviously, there are small details we do not know from the Henderson et al. (2015) Monte-Carlo simulations such as their bandwidth and the type of kernel density specified. Given that we do not know what seed their data is generated from, it is likely our h bandwidth value will be different despite choosing h using the same methodology.

et al. (2015) and Geng & Sun (2022), which entail calculating the asymptotic variance associated with our restricted estimator, $\tilde{\beta}^{*}(z)$. Formally, the asymptotic variance-covariance matrix needed to derive standard errors can be generalized by equation (3.38).

$$\Lambda(z) = v_0 \left[I_{ml} - G_z\right] W_z^{-1} \left[I_{ml} - G_z^T\right] \tag{3.38}$$

Wherein W_z, G_z, and v_0 are described by equations (3.39), (3.40), and (3.41), respectively.[80]

$$W_z = f(z)\mathbb{E}(X^T\Gamma^{-1}X|z_i = z), \text{ where } f(z) = nh^{-1}\sum_{i=1}^{n} K\left(\frac{z_i - z}{h}\right) \tag{3.39}$$

$$G_z = \left\{[X^T\mathbb{K}_z^{0.5}\Gamma^{-1}\mathbb{K}_z^{0.5}X]^{-1}R^T\right\} \times \left\{R[X^T\mathbb{K}_z^{0.5}\Gamma^{-1}\mathbb{K}_z^{0.5}X]^{-1}R^T\right\} R \tag{3.40}$$

$$v_0 = \int_{-\infty}^{\infty} K^2(v)dv, \text{ where } v = (z_i - z)/h \tag{3.41}$$

Geng & Sun (2022) stress that the square root of the diagonal of $\Lambda(z)$ produce the standard errors for observation i associated with each varying coefficient across all m equations. To provide an illustrative example of standard errors derived from this approach for our restricted estimator, consider our Monte-Carlo simulation using the same setup as before for $n = 100$ observations, and a cross-equation correlation coefficient of $c = 0.25$. For $b_{11}(z_i)$, $b_{12}(z_i)$, $b_{21}(z_i)$, and $b_{22}(z_i)$, Figure 31 describes each varying coefficients' observation-specific standard errors.

[80] Note that $\mathbb{E}\hat{f}(z) = \frac{1}{h}\mathbb{E}K\left(\frac{z_i - z}{h}\right) = h^{-1}\int_{-\infty}^{\infty} K\left(\frac{z_i - t}{h}\right) f(t)dt$. Given that $K(\cdot)$ is a Guassian, or normal, kernel, it is symmetric around zero as $h \to \infty$ and integrates to unity. With this in mind, it can be shown that $f(z) + h^2/2f^2(z)\int_{-\infty}^{\infty} K(v)v^2dv$. Given that our kernel is equal to point 0.5 on the $[-1, 1]$ interval, and zero otherwise, we have $\int_{-\infty}^{\infty} K(v)v^2dv = 0.5\int_{-1}^{1} v^2dv = \frac{1}{3}$.

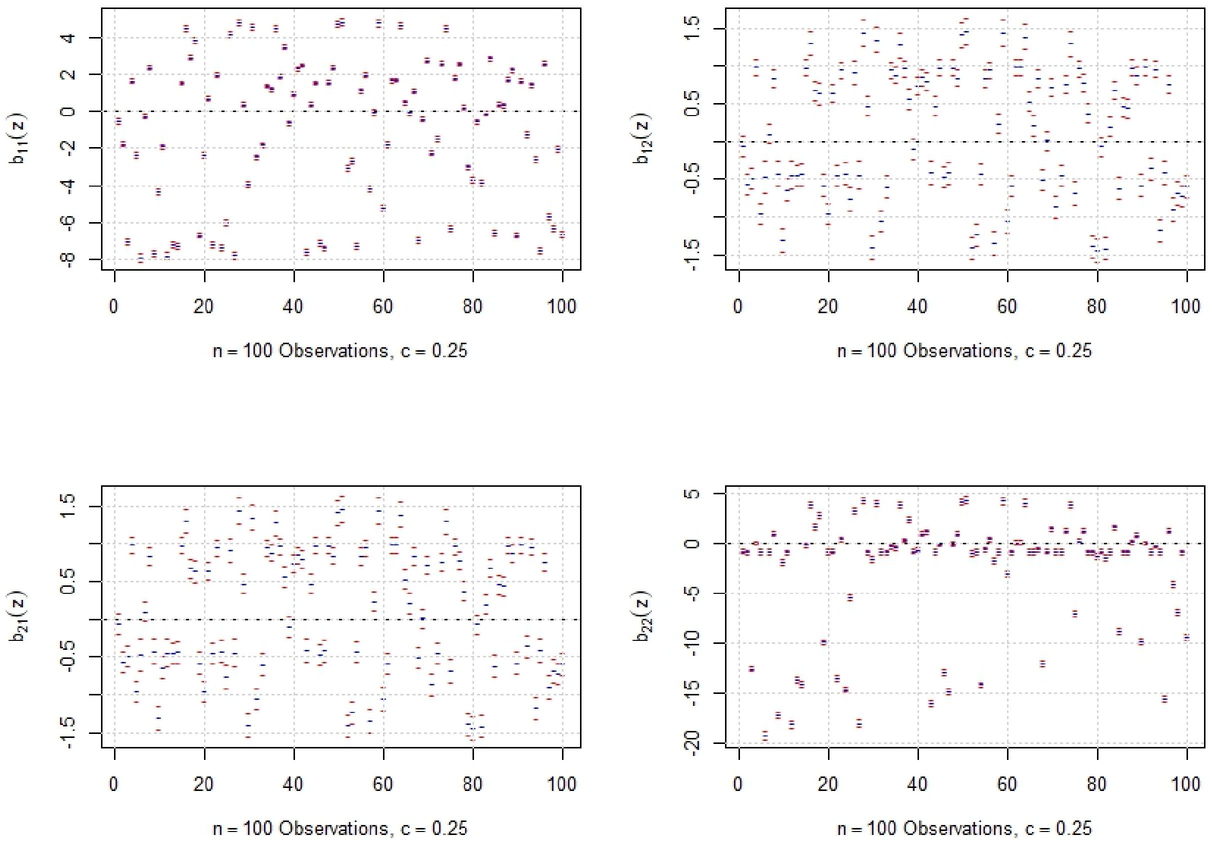


Figure 31: Point Estimates & Standard Errors of $b_{11}(z_i)$, $b_{12}(z_i)$, $b_{21}(z_i)$, and $b_{22}(z_i)$

The dashed blue lines denote the point estimates of observation i, while the dashed red lines illustrate the standard errors for observation i. Note that via our restrictions of $b_{12}(z_i) = b_{21}(z_i)$, both the point estimates standard errors for our restricted coefficients will be identical. Note that we can also add the point estimates and their respective standard errors together (like one would do for calculating elasticity-based returns-to-scale estimates). For adding point estimates, we can simply take their individual sums as described by $\mathbb{E}(\mathbf{X} + \mathbf{Y}) = \mathbb{E}(\mathbf{X}) + \mathbb{E}(\mathbf{X})$. Under the assumption that $\mathbf{X}$, and $\mathbf{Y}$ are independent from one another, we can define their joint standard error as $\sqrt{\text{var}(\mathbf{X} + \mathbf{Y})} = \sqrt{\text{var}(\mathbf{X}) + \text{var}(\mathbf{Y})}$. If we assume equation (3.28) is like a production function, then the sum of b_{11} and b_{12} can be thought of as returns-to-scale estimates. Graphically, $b_{11} + b_{12}$, and its corresponding standard errors are described by Figure 32.

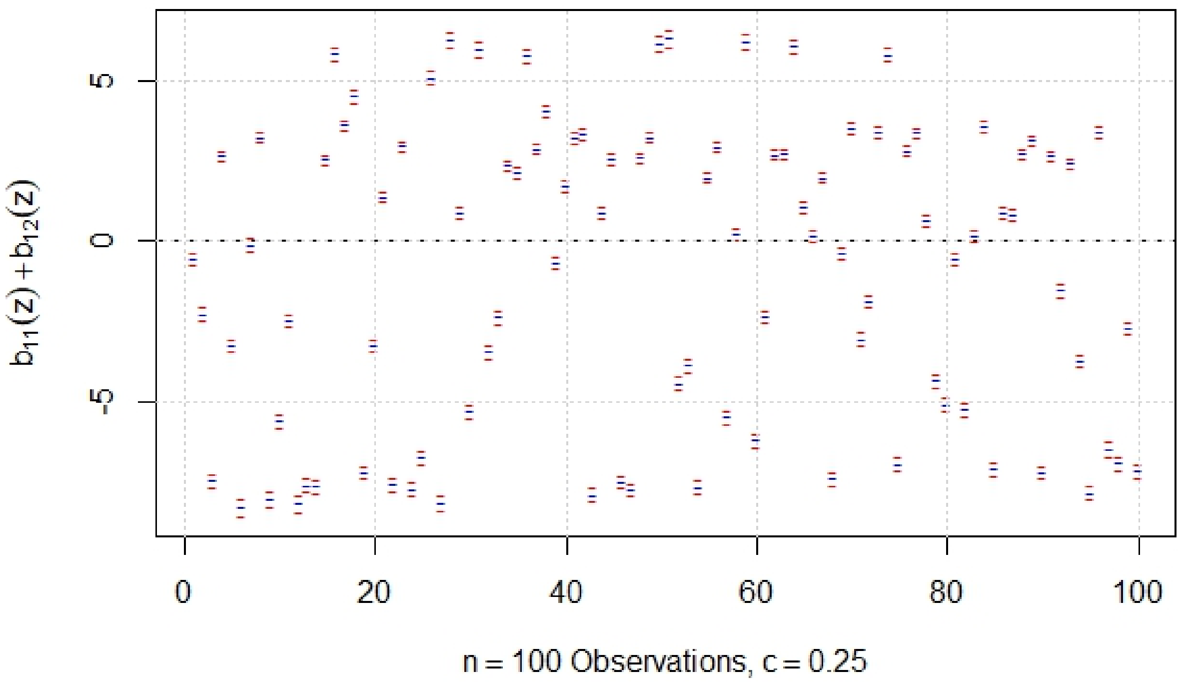


Figure 32: Point Estimates & Standard Errors for $b_{11} + b_{12}$

3.11 Appendix B: Bootstrapped Standard Errors

We take this space to discuss an alternative construction of standard errors from the SPSC SUR estimator. Given that each coefficient from an SPSC SUR or varying coefficient model varies with z_i, it is commonplace to report the distribution of each parameter estimate across each i observation. In a sense, the lower and upper quartiles of varying coefficients themselves are pseudo standard errors. However, we can construct explicit standard errors through the use of a bootstrapping procedure. While we do not formally present standard error estimates from our model coefficients, nor for returns-to-scale, we take this space to illustrate how constructing standard errors are possible in our empirical methodology.

For an illustrative case, we will use our data set of critical market banks from the fiscal year of 2021. Prior to purging outlying RTS estimates post-estimation, we have $n = 465$ observations. Each i observation for any parameter estimate in $\tilde{\beta}^*(z)$ needs a standard error. To achieve this, the following steps are taken in our bootstrapping procedure:

1. For any given indexed year, t, we resample our data with replacements for z, y_1, y_2, y_3, y_4, y_5, w_1, w_2,

w_3, w_4, w_5, x_1, x_2, x_3, x_4, and x_5

2. Using our resampled data, we reconstruct our cost shares, S_1, S_2, S_3, S_4, and S_5, as well as our total costs, C

3. We rerun our SPSC SUR model using our resampled data and store our $\tilde{\beta}^*(z)$ of coefficient estimates[81]

4. We repeat all previous steps $b = 20$ times[82]

5. For a given parameter and observation i, across all b repetitions, we take observation i's point estimates and apply the traditional standard error formula $SE_i = \sigma_i/\sqrt{b}$, where σ_i is the standard deviation of the vary coefficients a given i observation possesses across b simulations

3.11.1 Illustrative Case

Let's say we are interested in the standard errors associated with $\gamma_1(z_i)$ from equation (3.21) for the year $t = 2021$. The first thing we will do is create an object that can store $b = 20$ $\tilde{\beta}^*(z)$ matrices of varying coefficient estimates. Given that $n = 465$ for the year of 2021, and we have $l = 111$ coefficients across all five equations in our SPSC SUR system, $\tilde{\beta}^*(z)$ will be 465×111 in dimension. Denote our storage matrix as $\overline{\gamma}_1(z_i)$, which is $n \times b$ in dimension.

Any column vector b in $\overline{\gamma}_1(z_i)$ is associated with the bootstrapped values of $\gamma_1(z_i) \in \tilde{\beta}^*(z)$ from $b = 1, \ldots, 20$ bootstrapping procedures for $i = 1, \ldots, 465$ observations. For any given row vector i in $\overline{\gamma}_1(z_i)$, we apply our normal standard error formula of $SE_i = \sigma_i/\sqrt{b}$. After applying this formula over all n rows of $\overline{\gamma}_1(z_i)$, we produce an $n \times 1$ vector, $SE(\overline{\gamma}_1(z_i))$, containing bootstrapped standard errors for all varying estimates of $\gamma_1(z_i)$. Figure 33 shows the distribution of bootstrapped standard errors for $\gamma_1(z_i)$ based on $b = 20$ repetitions.

[81] For computational ease, we assume that for any bootstrapping simulation our h bandwidth is fixed at the true bandwidth observed in the actual z data.

[82] We acknowledge that this is a rather small scale of repetition for a bootstrapping procedure, however, computation time takes roughly 36 hours for $b = 20$, thus rendering any b of an arbitrarily large size computationally infeasible.

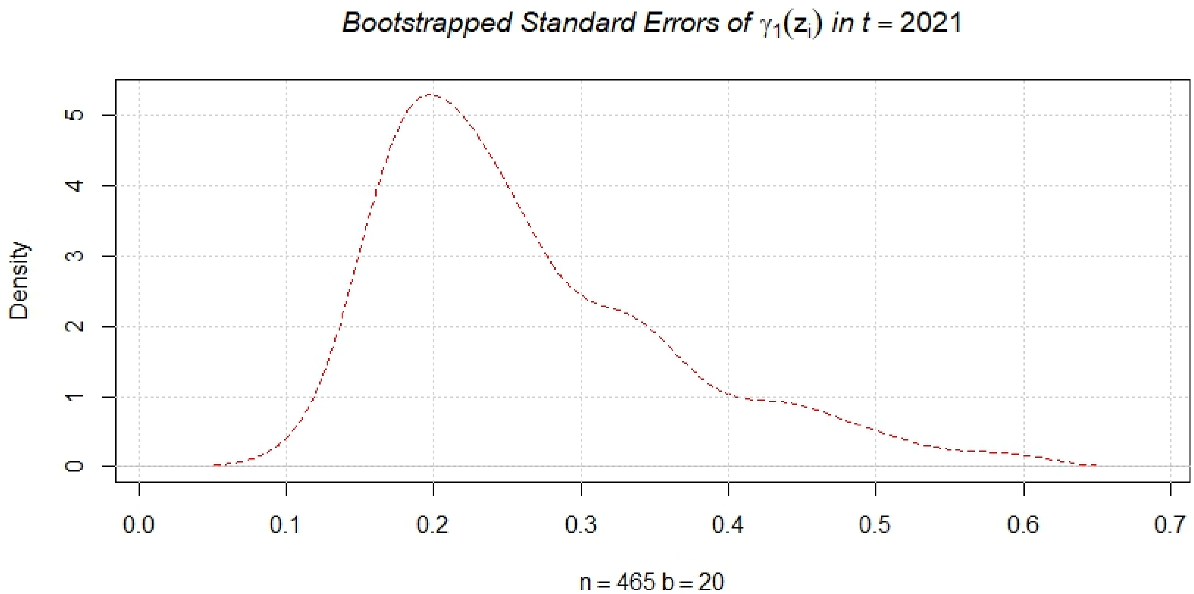


Figure 33: Bootstrapped Standard Errors of $\gamma_1(z_i)$

We can see that the highest density of standard errors associated with $\gamma_1(z_i)$ for any observation i tends to be centered around (0.2) with the median, upper, and lower quartiles carrying estimates of (0.24), (0.32), and (0.19), respectively. Of course, given that we know the estimates of $\gamma_1(z_i)$ from our actual estimation procedure, we can assign all 465 bootstrapped standard errors to their corresponding point estimates. Figure 34 captures this pairing of our bootstrapped standard errors with our point estimates of $\gamma_1(z_i)$ for each observation i.

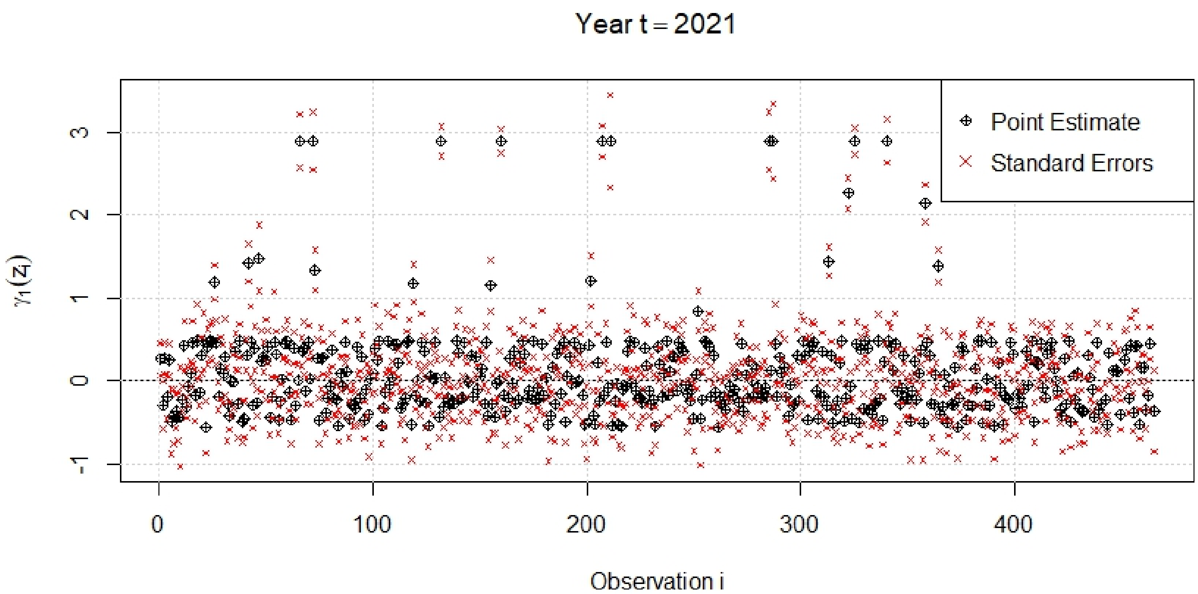


Figure 34: Point Estimates & Standard Errors of $\gamma_1(z_i)$

As we can see from Figure 34, there is considerable variation in the size and significance of $\gamma_1(z_i)$ across our subsample of critical market banks from 2021. This variation lies at the heart of what SPSC and SPSC SUR models seek to exploit. In fact, of the 465 varying coefficients of $\gamma_1(z_i)$, there are 194 whose confidence intervals contain zero while there are 271 that are statistically different from zero. Given that our RTS estimates are constructed partially from $\gamma_1(z_i)$, it will likely be the case that this elasticity has a significant impact for most critical market banks' economies of scale estimates, though this will not always be the case.

3.12 Appendix C: Select Coefficient Estimates

While not relevant to our main results per se, it is typical in most varying coefficient models to report the median, lower, and upper quartiles of key model coefficients. Looking specifically at our total cost function, (3.21), we report varying coefficient results for the Cobb-Douglass elasticities from our SPSC SUR model, β_j and γ_q. Table 43 reports these varying coefficients estimates for our input price elasticities, β_j.

Table 43: Input Prices Varying Coefficients

	β_1			β_2			β_3			β_4		
	25 Pct.	50 Pct.	75 Pct.	25 Pct.	50 Pct.	75 Pct.	25 Pct.	50 Pct.	75 Pct.	25 Pct.	50 Pct.	75 Pct.
2001	-0.18	-0.04	0.14	2.40	4.06	5.21	1.12	1.46	2.18	0.40	0.64	0.80
2002	-0.24	-0.05	0.16	5.40	5.90	7.10	1.88	2.63	3.44	0.35	0.53	0.59
2003	-0.72	-0.20	0.09	6.96	8.68	10.38	2.02	2.97	3.79	0.37	0.49	0.68
2004	-0.39	-0.09	0.27	8.19	9.95	12.30	2.74	3.62	4.01	0.41	0.51	0.56
2005	-0.39	-0.15	0.09	4.90	7.51	9.80	2.94	3.54	4.04	0.44	0.58	0.71
2006	-0.08	0.01	0.15	4.78	5.48	8.33	1.21	1.84	2.72	0.37	0.74	0.88
2007	-0.30	-0.01	0.17	2.76	4.78	6.76	0.90	1.70	3.03	0.55	0.68	0.87
2008	-0.39	0.13	0.41	4.10	4.77	6.15	0.78	1.57	2.26	0.55	0.64	0.81
2009	-0.78	-0.04	0.54	4.99	6.63	8.38	0.01	3.62	4.39	0.58	0.69	0.82
2010	-1.80	-0.80	0.25	11.29	16.19	20.66	6.58	11.96	13.97	0.47	0.56	0.73
2011	-0.12	0.07	0.54	3.36	10.43	12.64	4.33	6.24	10.36	0.46	0.76	0.95
2012	-0.07	0.54	1.43	6.39	9.36	13.54	6.06	7.75	9.22	0.58	0.73	0.87
2013	-0.12	1.07	2.31	0.69	5.36	8.44	8.59	10.78	11.25	0.95	0.99	1.02
2014	-0.26	0.96	3.18	9.30	13.62	14.58	10.99	11.91	12.32	0.71	0.77	0.82
2015	0.63	0.84	1.05	6.84	8.29	13.92	8.85	10.85	14.90	0.67	0.72	0.78
2016	-0.72	-0.34	0.45	11.75	14.22	17.32	7.13	10.25	13.71	0.52	0.54	0.59
2017	0.04	0.05	0.36	11.99	13.66	15.85	5.97	9.02	11.31	0.58	0.65	0.71
2018	-0.21	0.05	0.30	15.02	15.35	15.77	3.57	5.44	10.53	0.61	0.69	0.75
2019	-0.63	-0.11	0.31	8.02	11.01	13.88	1.58	5.22	5.96	0.55	0.73	0.92
2020	-0.87	-0.62	-0.40	12.47	13.15	16.88	1.81	4.15	8.62	0.48	0.56	0.68
2021	-1.80	-0.80	0.25	11.29	16.19	20.66	6.58	11.96	13.97	0.47	0.56	0.73

We can see that the median coefficient estimates are generally positively signed, as theory would imply. However, in the lower quartile in some years for some β_j estimates, we observe negative coefficients. While theory implies $\beta_j > 0$, it is not uncommon to encounter such inconsistencies. For our output elasticities, γ_q, we report a similar distribution of varying coefficient estimates in Table 44.

Table 44: Output Quantities Varying Coefficients

	γ_1			γ_2			γ_3			γ_4			γ_5		
	25 Pct.	50 Pct.	75 Pct.	25 Pct.	50 Pct.	75 Pct.	25 Pct.	50 Pct.	75 Pct.	25 Pct.	50 Pct.	75 Pct.	25 Pct.	50 Pct.	75 Pct.
2001	-0.26	0.06	0.26	0.05	0.17	0.23	0.08	0.26	0.37	-0.11	0.06	0.25	0.03	0.04	0.09
2002	-0.89	-0.49	-0.10	0.08	0.23	0.36	0.00	0.29	0.79	-0.32	0.21	0.62	-0.02	0.03	0.06
2003	-0.36	-0.11	0.20	0.10	0.23	0.64	0.12	0.60	1.05	-0.77	-0.22	0.02	-0.13	0.07	0.12
2004	-0.41	-0.26	0.25	0.15	0.25	0.36	-0.27	0.26	0.78	-1.20	-0.08	0.71	-0.04	0.01	0.12
2005	-0.30	-0.02	0.11	0.13	0.22	0.33	-0.01	0.53	0.82	-0.68	0.06	0.55	-0.10	-0.03	0.01
2006	-0.37	-0.21	0.01	0.07	0.21	0.32	-0.11	0.18	0.49	-0.12	0.24	0.42	-0.17	-0.08	-0.01
2007	-0.28	0.09	0.69	-0.07	0.14	0.38	-0.77	0.24	0.85	-0.41	0.43	1.93	-0.25	-0.07	0.02
2008	-0.43	-0.12	0.05	0.06	0.31	0.48	-0.42	0.27	0.69	-0.57	0.34	1.08	-0.13	-0.01	0.10
2009	-0.33	-0.05	0.53	-0.34	0.10	0.52	-0.49	0.31	0.96	-0.83	-0.30	0.22	-0.08	0.03	0.15
2010	-0.27	-0.04	0.36	-0.12	0.01	0.20	-0.63	-0.42	0.14	0.89	1.52	1.93	-0.06	-0.01	0.11
2011	-0.45	-0.17	0.06	-0.22	0.22	0.43	-0.52	-0.05	0.48	1.06	1.47	1.77	-0.14	-0.02	0.03
2012	-0.33	-0.18	0.23	-0.07	0.09	0.28	-0.58	0.32	1.00	-0.40	0.35	1.01	-0.09	-0.01	0.10
2013	-0.39	-0.27	0.10	-0.06	0.04	0.16	0.42	0.47	0.62	-0.16	0.00	0.18	-0.01	0.02	0.12
2014	-0.45	-0.19	-0.12	0.07	0.21	0.32	-0.44	-0.22	0.04	-0.05	0.42	0.90	-0.07	-0.01	0.06
2015	0.16	0.23	0.37	-0.23	-0.19	-0.12	-0.12	0.39	0.90	0.12	0.49	0.91	-0.03	0.06	0.13
2016	-0.32	-0.04	0.11	-0.06	0.11	0.19	-0.04	0.20	0.50	0.05	0.77	0.99	0.00	0.00	0.03
2017	-0.23	-0.03	0.14	-0.11	-0.07	-0.02	-0.10	0.10	0.58	-0.10	0.17	0.39	-0.05	0.00	0.06
2018	-0.17	0.07	0.17	-0.30	-0.21	-0.04	-0.06	0.12	1.07	-0.18	0.15	0.38	-0.11	0.06	0.14
2019	-0.23	0.03	0.20	-0.22	-0.16	-0.08	0.29	0.53	0.67	-0.07	0.17	0.35	-0.01	0.07	0.11
2020	-0.24	0.02	0.15	-0.22	-0.02	0.07	-0.37	-0.25	0.15	-0.09	0.70	1.02	-0.16	-0.10	0.00
2021	-0.27	-0.04	0.36	-0.12	0.01	0.20	-0.63	-0.42	0.14	0.89	1.52	1.93	-0.06	-0.01	0.11

In a similar vein, all γ_q coefficients should be > 0. As we can see, particularly for γ_1 and γ_5, there are several sign violations across each year. At the moment, there is no way to correct for monotonicity violations in the SPSC SUR estimator. Future work would do well to expand the Henderson et al. (2015)

SPSC SUR estimator to accommodate non-negativity constraints drawing from works like Hall & Huang (2001), Du et al. (2013), and Parmeter et al. (2014).

3.13 Appendix D: Alternative Transformation

As evident by Table 40, banking data is heavily skewed to the right and contains many zeroes. While a log transformation is standard for most banking studies, it is frequently necessary to add one to all data before taking logs to avoid taking the log of zero. An alternative transformation common for data of our variety is that of the inverse hyperbolic sine transformation ($\sinh^{-1}$). Notwithstanding criticisms in from works like Bellemare & Wichman (2020), and Aihounton & Henningsen (2021), the $\sinh^{-1}$ transformation is an appealing alternative to a log transformation that has yet to be used extensively in banking studies.

We take this space to provide some tertiary results using our SPSC SUR estimator wherein our data is transformed using an inverse hyperbolic sine function. Equations (3.42), and (3.43) express our alternative SPSC SUR specification taking the $\sinh^{-1}$ transformation into account.

$$
\begin{aligned}
\sinh^{-1}(C_i/w_{5i}) = \alpha_0(z_i) + \sum_{j=1}^{4} \beta_j(z_i)\sinh^{-1}(w_{ji}/w_{5i}) + \sum_{q=1}^{5} \gamma_q(z_i)\sinh^{-1}(y_{qi}) \\
+\frac{1}{2}\sum_{q=1}^{5}\sum_{l=1}^{5}\gamma_{ql}(z_i)\sinh^{-1}(y_{qi})\sinh^{-1}(y_{li}) + \frac{1}{2}\sum_{j=1}^{4}\sum_{v=1}^{4}\eta_{jv}(z_i)\sinh^{-1}(w_{ji}/w_{5i})\sinh^{-1}(w_{vi}/w_{5i}) \\
+\sum_{j=1}^{4}\sum_{q=1}^{5}\delta_{jq}(z_i)\sinh^{-1}(w_{ji}/w_{5i})\sinh^{-1}(y_{qi}) + u_{ji}
\end{aligned}
\tag{3.42}
$$

$$
S_{ji} = \alpha_j(z_i) + \sum_{v=1}^{4}\eta_{jv}(z_i)\sinh^{-1}(w_{vi}/w_{5i}) + \sum_{q=1}^{5}\delta_{jq}(z_i)\sinh^{-1}(y_{qi}) + u_{ji} \text{ for } j \in \{1,2,3,4\} \tag{3.43}
$$

Table 45 reports new RTS estimates from equations (3.42), and (3.43).

Table 45: Nonperformance-Adjusted RTS Estimates for Critical Market Banks ($\sinh^{-1}$ Transformation)

Year	Lower Quartile	Median	Upper Quartile	Mean (μ_t)	Std. Dev. (σ_t)	Obs. (N)	IRS	DRS
2001–2021	0.51	0.94	1.67	0.85	2.25	10606	4983 (47%)	5623 (53%)
2001	0.97	1.67	2.66	1.90	1.26	613	456 (74%)	157 (26%)
2002	0.65	1.38	1.91	1.28	0.93	544	341 (63%)	203 (37%)
2003	0.90	1.07	1.32	1.16	0.42	508	328 (65%)	180 (35%)
2004	-1.46	0.53	0.86	-0.20	1.59	734	150 (20%)	584 (80%)
2005	0.86	1.08	1.24	1.05	0.28	542	324 (60%)	218 (40%)
2006	0.69	2.18	3.15	2.01	1.71	598	406 (68%)	192 (32%)
2001–2006	0.64	1.10	1.82	1.15	1.44	3539	2005 (57%)	1534 (43%)
2007	0.54	0.74	1.26	0.98	0.57	599	191 (32%)	408 (68%)
2008	0.28	0.65	1.27	0.81	1.12	532	175 (33%)	357 (67%)
2009	-0.57	0.64	1.39	0.52	1.35	609	215 (35%)	394 (65%)
2007–2009	0.45	0.70	1.30	0.77	1.08	1740	581 (33%)	1159 (67%)
2010	0.69	0.91	1.21	0.98	0.54	479	201 (42%)	278 (48%)
2011	0.39	0.43	0.53	0.53	0.28	451	59 (13%)	392 (87%)
2012	0.51	0.65	1.30	0.91	0.59	412	150 (36%)	262 (64%)
2013	0.95	1.71	2.80	2.02	1.39	584	394 (67%)	154 (33%)
2014	-2.60	0.58	1.27	-0.44	2.39	516	190 (37%)	326 (63%)
2015	0.48	0.53	1.16	0.89	0.75	440	129 (29%)	311 (71%)
2016	0.94	1.42	1.68	1.32	0.52	457	319 (70%)	138 (30%)
2017	-7.87	0.83	2.51	-2.31	8.13	421	202 (48%)	219 (52%)
2018	-0.18	0.72	2.85	0.64	2.96	436	168 (39%)	268 (61%)
2019	1.15	1.79	2.90	1.93	0.99	442	382 (86%)	60 (14%)
2010–2019	0.47	0.94	1.71	0.68	3.06	4602	2194 (48%)	2408 (52%)
2020	0.59	1.00	1.57	0.63	1.65	353	177 (50%)	176 (50%)
2021	0.72	0.85	0.93	0.83	0.12	372	26 (7%)	346 (93%)
2020–2021	0.71	0.87	1.02	0.73	1.16	725	203 (28%)	522 (72%)

Note: our results present RTS estimates described by equation (3.42) that are derived from our cost equation, (3.43), for all years from 2001 through 2021 meaning our results consist of estimates from 21 unique SPSC SUR models indexed from years t = 2001 through T = 2021. Subsample estimates are computed through aggregation of all index years of interest. While it is possible to apply our model over longer samples or as a true panel model, estimate convergence times and processing requirements limit the feasibility of this option. Critical market banks identified fall in the upper decile (90^{th} percentile) for relative critical market involvement of all banks in any given year, thus there is an allowance for the possibility that the total number of banks operating in critical markets most intensely can vary as the competitive landshape shrinks.

Consistent with our original results described in Table 41, we observe that over the full sample, most individual critical market banks express decreasing returns-to-scale. Unlike our original results that suggest that the median critical market bank enjoys modest increasing returns-to-scale, Table 45 suggests that the median critical market bank expresses slight decreasing returns-to-scale. Furthermore, banks expressing increasing returns-to-scale in spite of nonperforming assets are most frequently observed over the subsample preempting the Financial Crisis. Again, echoing our original results, the total amount of banks expressing increasing returns-to-scale seems to decrease concurrent with the shrinking competitive landscape. These results suggest that the use of the inverse hyperbolic sine transformation can serve as an appealing alternative to log transformations while maintaining result interpretability.

3.14 Appendix E: Additional Descriptive Statistics

Full Sample Density Plots of Variables of Interest

Logs of Output Quantities

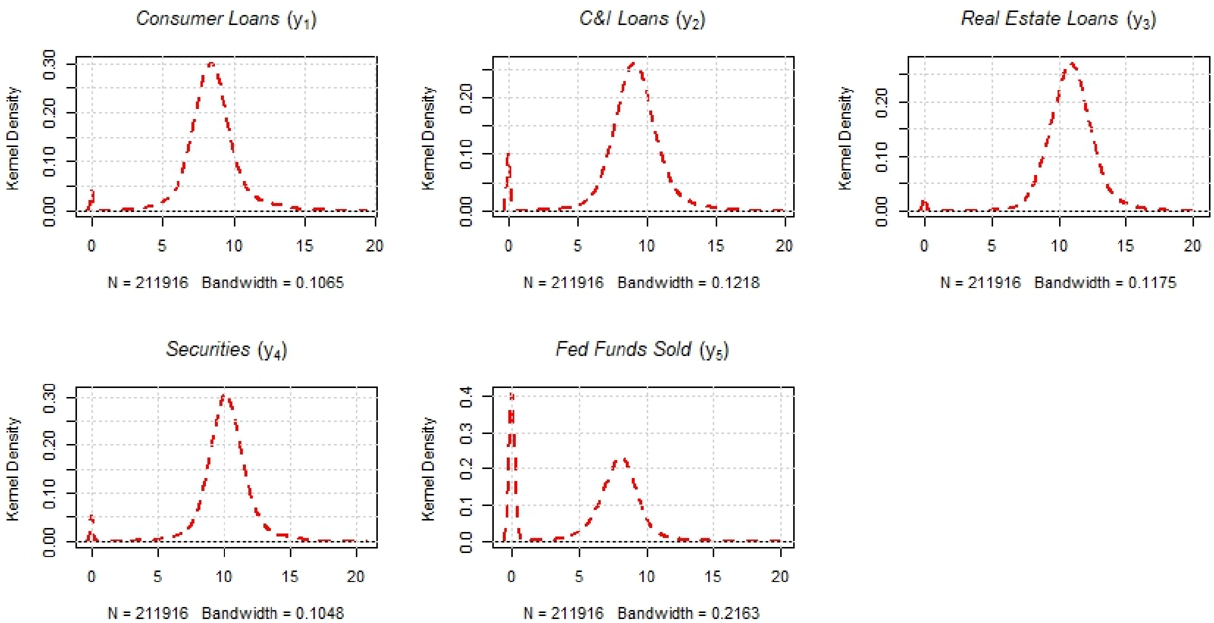


Figure 35: Kernel Density Estimates of Banking Outputs

Logs of Input Quantities

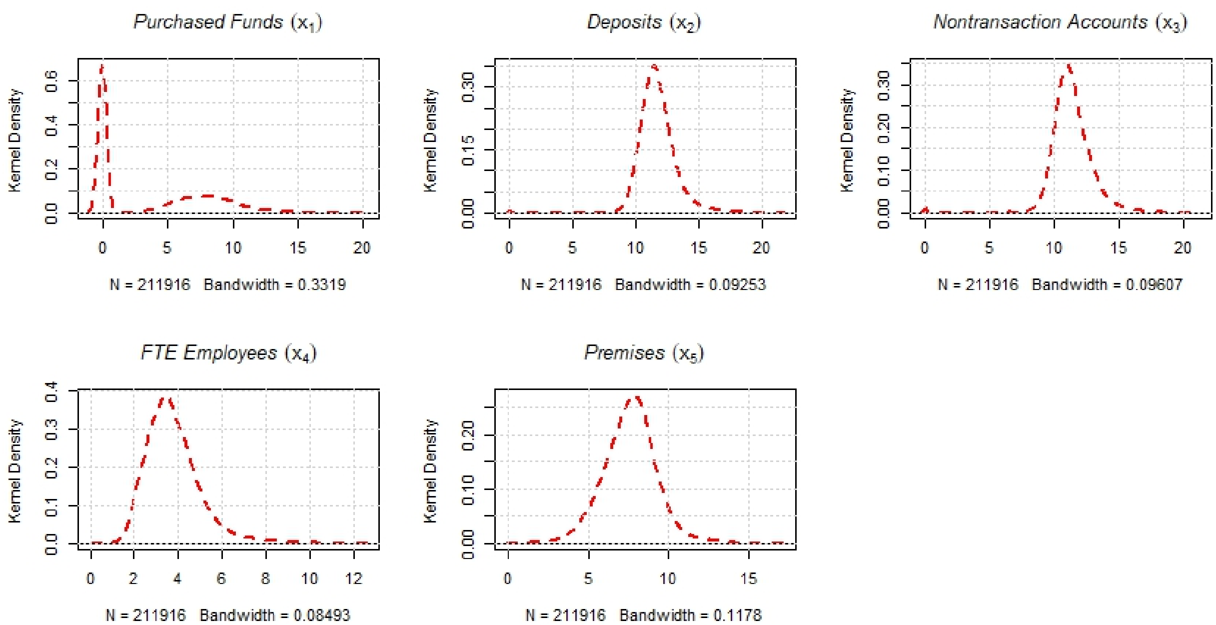


Figure 36: Kernel Density Estimates of Input Quantities

Logs of Normalized Costs & Input Prices

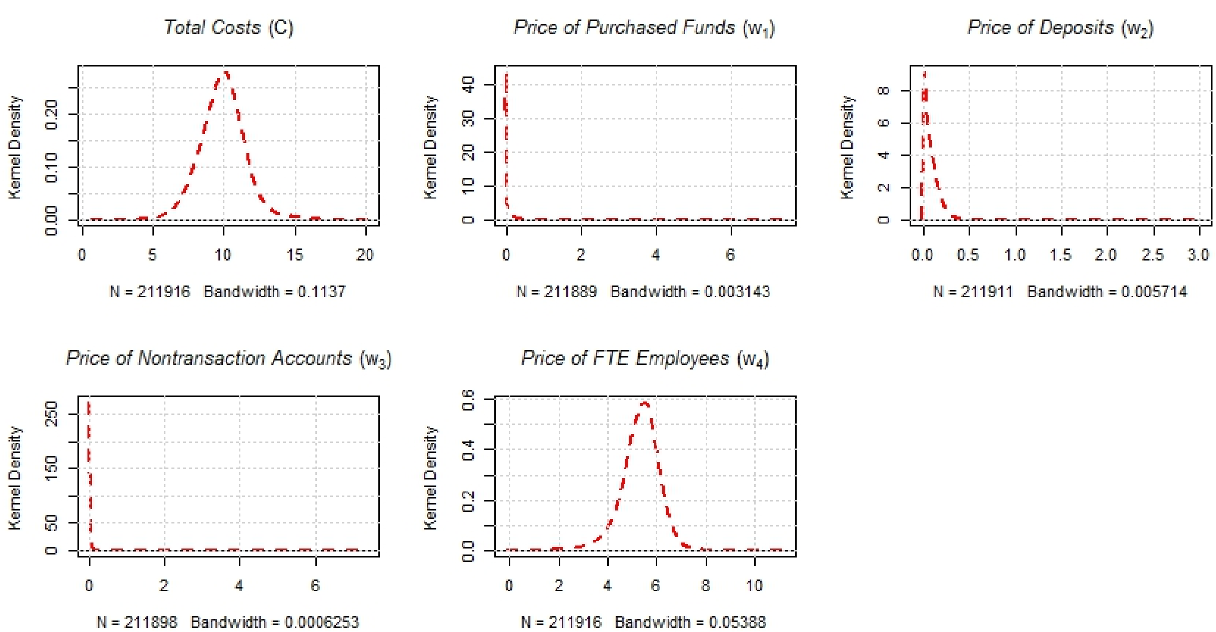


Figure 37: Kernel Density Estimates of Normalized Banking Costs & Unit Prices

Cost Shares

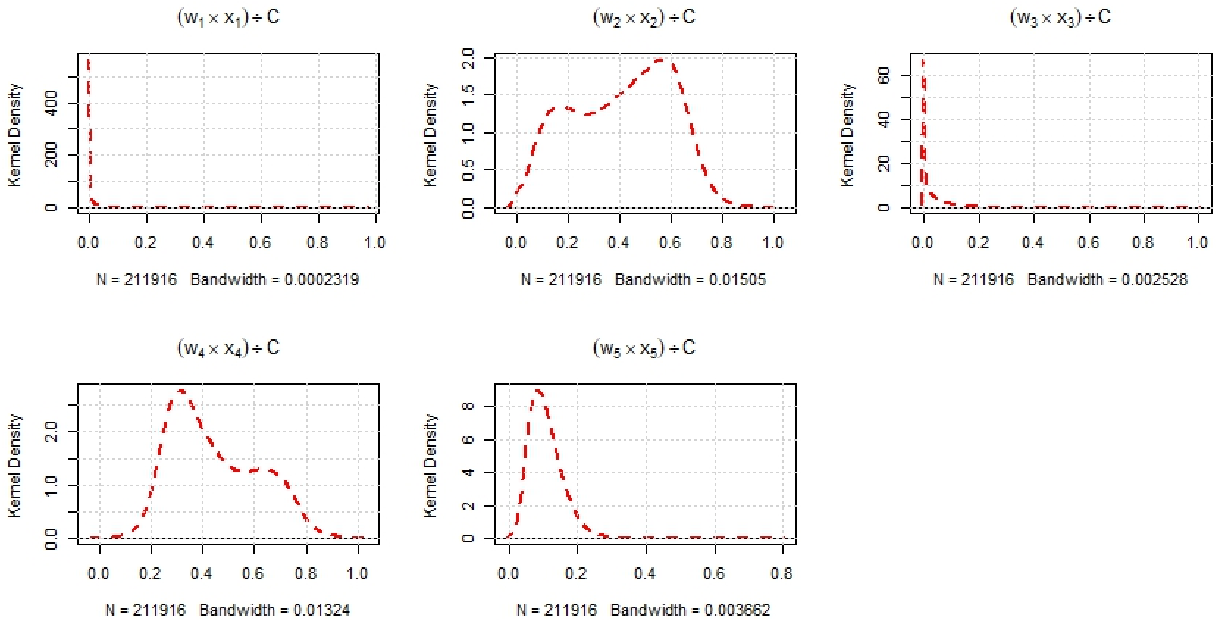


Figure 38: Kernel Density Estimates of Cost Shares

Logs of Environmental Variables

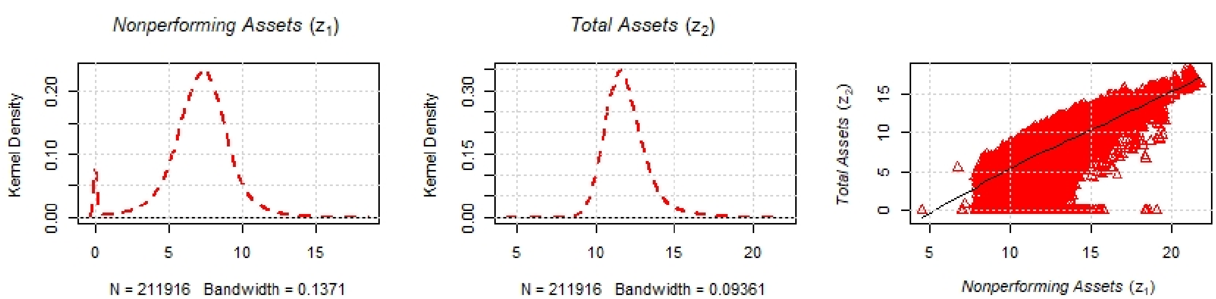


Figure 39: Environmental Factors

Correlation Matrix

Table 46: Correlation Matrix of Model Variables

Variable	C/w_5	w_1/w_5	w_2/w_5	w_3/w_5	w_4/w_5	y_1	y_2	y_3	y_4	y_5
C/w_5	1.0000	0.1914	0.3070	0.1013	0.5837	0.5743	0.5597	0.72130	0.5842	0.09339
w_1/w_5	0.1914	1.0000	0.2921	0.0341	0.1375	0.1289	0.1168	0.07838	0.0904	0.06497
w_2/w_5	0.3070	0.2921	1.0000	0.0563	0.4271	0.1193	-0.0309	-0.01199	-0.0189	0.18214
w_3/w_5	0.1013	0.0341	0.0563	1.0000	0.0280	0.0523	-0.0413	-0.05709	-0.0171	-0.02082
w_4/w_5	0.5837	0.1375	0.4271	0.0280	1.0000	0.0890	0.1835	0.31127	0.1108	-0.08217
y_1	0.5743	0.1289	0.1193	0.0523	0.0890	1.0000	0.5262	0.43962	0.4882	0.18915
y_2	0.5597	0.1168	-0.0309	-0.0413	0.1835	0.5262	1.0000	0.58337	0.4455	0.14474
y_3	0.7213	0.0784	-0.0120	-0.0571	0.3113	0.4396	0.5834	1.00000	0.5101	-0.00982
y_4	0.5842	0.0904	-0.0189	-0.0171	0.1108	0.4882	0.4455	0.51005	1.0000	0.09184
y_5	0.0934	0.0650	0.1821	-0.0208	-0.0822	0.1892	0.1447	-0.00982	0.0918	1.00000

3.15 Appendix F: Additional Tabulated Statistics

3.15.1 Banks With Highest Levels of Involvement in Critical Markets

Table 47: Full Sample Descriptive Statistics: Highest Critical Market Involvement

Variable	Description	10^{th} Pct.	25^{th} Pct.	Median	75^{th} Pct.	90^{th} Pct.	Mean
C	Total Costs	680	1245	2646	6516	20454	22604
y_1	Consumer Loans	523	1483	3677	9213	25463	93937
y_2	C&I Loans	333	1632	5032	15197	54932	161673
y_3	Real Estate Loans	3649	10122	30831	96848	346436	361608
y_4	Securities	15155	27468	58445	142748	412174	516087
y_5	Fed Funds Sold	0	0	1509	4658	12623	28128
x_1	Purchased Funds	0	0	0	2079	26466	62292
x_2	Deposits	25969	48657	102830	260523	763300	1150311
x_3	Nontransaction Accounts	16556	31408	67820	189463	639939	876130
x_4	FTE Employees	9	15.5	31	74.2	206	201
x_5	Premises	156	473	1458	4505	13461	13564
w_1	Price of Purchased Funds	0	0	0	0.00843	0.0262	0.00903
w_2	Price of Deposits	0.00178	0.00434	0.013	0.0198	0.0241	0.0125
w_3	Price of Nontransaction Accounts	0	0	0.0000475	0.00133	0.00502	0.00486
w_4	Price of FTE Employees	24	28.3	31	41.2	50.6	36.3
w_5	Price of Premises	0.0836	0.119	0.18	0.3	0.538	0.686
S_1	$(w_1 \times x_1)/C$	0	0	0	0.00473	0.0329	0.0129
S_2	$(w_2 \times x_2)/C$	0.116	0.232	0.425	0.577	0.666	0.407
S_3	$(w_3 \times x_3)/C$	0	0	0.00139	0.0374	0.117	0.0351
S_4	$(w_4 \times x_4)/C$	0.229	0.296	0.402	0.586	0.711	0.44
S_5	$(w_5 \times x_5)/C$	0.0451	0.0667	0.0973	0.136	0.174	0.105

Note: $y_q \in \mathcal{Y}$ capture our vectors of outputs. $x_j \in \mathcal{X}$ encompass our vector of input quantities while $w_j \in \mathcal{W}$ correspond to each respective input price. Finally, $S_j \in \mathcal{S}$ are our cost share equations that are identified functionally as $(w_j x_j) \div C_i$, where C_i are the total costs for bank i. Finally, total costs are defined as the sum of the products of all input quantities times their respective prices such that $C_i = \sum_{j=1}^{J} w_j x_j$. All variables with the exception of full-time employees are denoted in terms of thousands of US dollars in constant 2012 prices.

3.15.2 Banks With Lowest Levels of Involvement in Critical Markets

Table 48: Full Sample Descriptive Statistics: Lowest Critical Market Involvement

Variable	Description	10^{th} Pct.	25^{th} Pct.	Median	75^{th} Pct.	90^{th} Pct.	Mean
C	Total Costs	592	1194	2615	6274	17219	33295
y_1	Consumer Loans	124	752	2664	8098	31442	319918
y_2	C&I Loans	11.5	1878	7222	21342	60352	166068
y_3	Real Estate Loans	4150	16497	47541	120427	302409	361608
y_4	Securities	0	344	2779	10756	33499	59496
y_5	Fed Funds Sold	0	0	1781	6648	18282	64804
x_1	Purchased Funds	0	0	0	65.9	3035	75989
x_2	Deposits	16426	36809	80875	187418	463628	723945
x_3	Nontransaction Accounts	10924	25772	58072	135757	358771	432603
x_4	FTE Employees	8	14	27.8	64.2	164	226
x_5	Premises	158	480	1482	3853	9496	11472
w_1	Price of Purchased Funds	0	0	0	0	0.02	0.00905
w_2	Price of Deposits	0.0024	0.00527	0.0126	0.021	0.0262	0.0148
w_3	Price of Nontransaction Accounts	0	0	0.000221	0.00202	0.00588	0.292
w_4	Price of FTE Employees	24.8	27.8	30.2	48.1	62.1	42
w_5	Price of Premises	0.0689	0.104	0.173	0.391	0.886	1.07
S_1	$(w_1 \times x_1)/C$	0	0	0	0	0.0087	0.00855
S_2	$(w_2 \times x_2)/C$	0.095	0.213	0.391	0.54	0.642	0.379
S_3	$(w_3 \times x_3)/C$	0	0	0.00544	0.0491	0.118	0.0433
S_4	$(w_4 \times x_4)/C$	0.242	0.314	0.433	0.601	0.725	0.46
S_5	$(w_5 \times x_5)/C$	0.0476	0.0703	0.101	0.138	0.179	0.109

Note: $y_q \in \mathcal{Y}$ capture our vectors of outputs. $x_j \in \mathcal{X}$ encompass our vector of input quantities while $w_j \in \mathcal{W}$ correspond to each respective input price. Finally, $S_j \in \mathcal{S}$ are our cost share equations that are identified functionally as $(w_j x_j) \div C_i$, where C_i are the total costs for bank i. Finally, total costs are defined as the sum of the products of all input quantities times their respective prices such that $C_i = \sum_{j=1}^{J} w_j x_j$. All variables with the exception of full-time employees are denoted in terms of thousands of US dollars in constant 2012 prices.

3.16 Appendix G: Additional Contour Plots

Total Costs & Nonperforming Assets

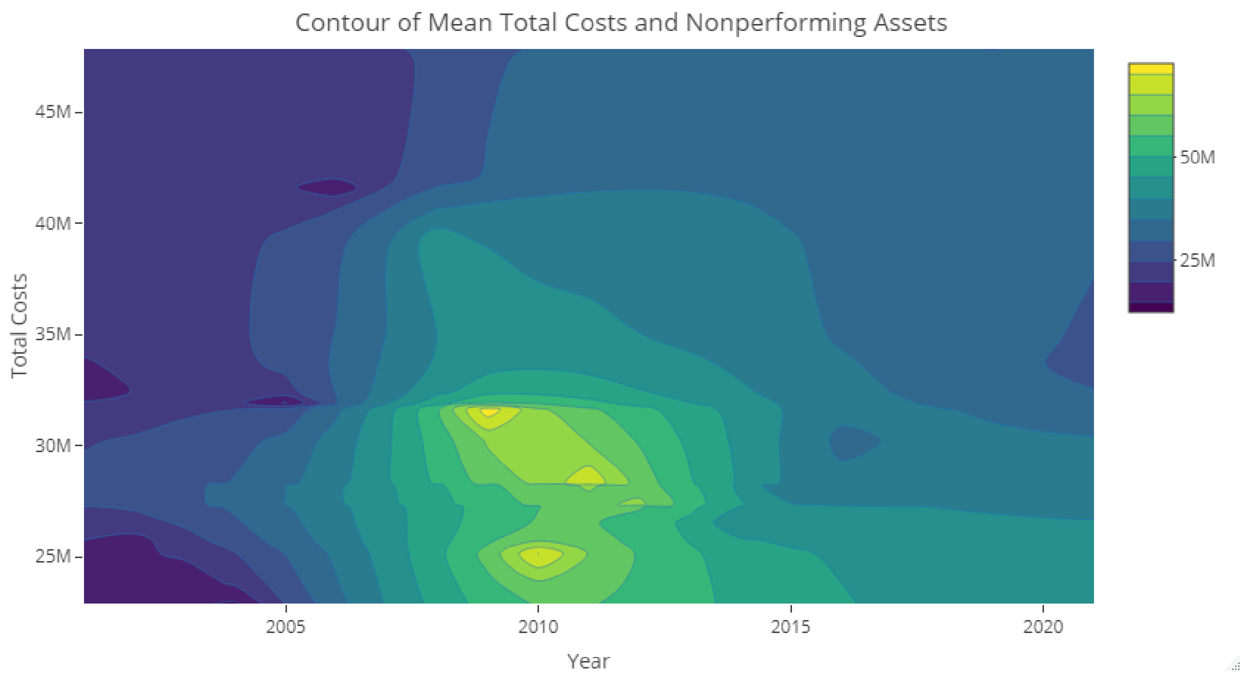


Figure 40: Contour of Total Costs & Nonperforming Assets Over Time

Total Costs & Critical Market Involvement

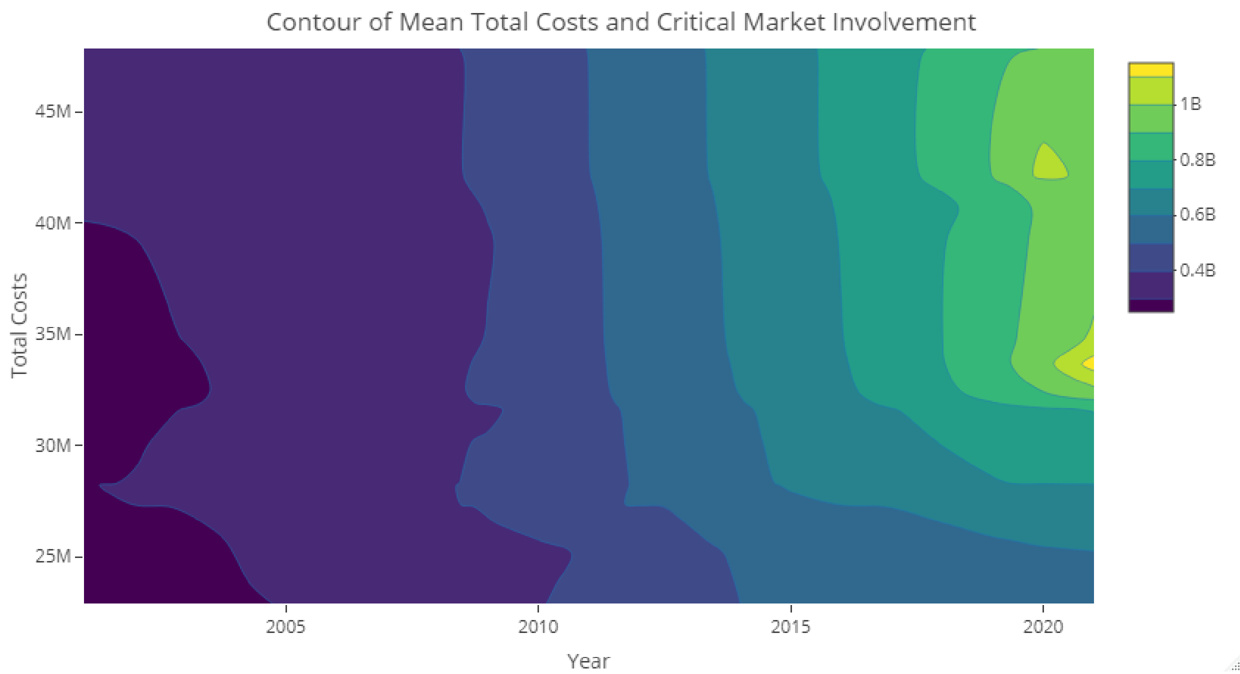


Figure 41: Contour of Total Costs & Critical Market Involvement Over Time

3.17 Appendix H: Critical Market Versus Non-Critical Market Banks

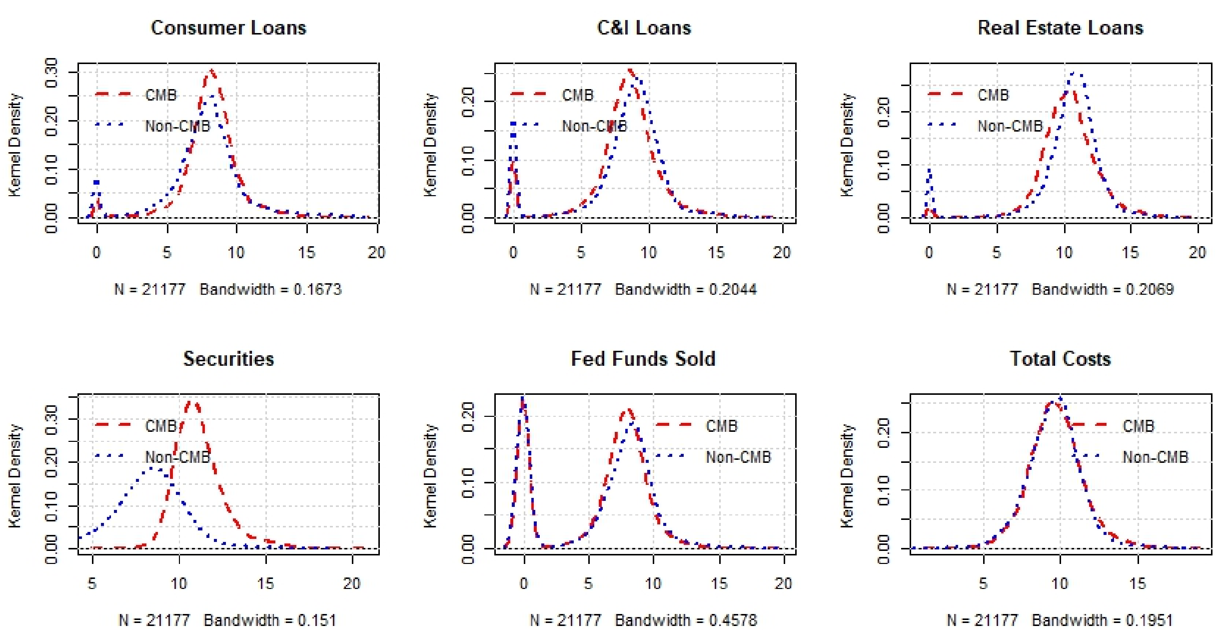


Figure 42: Outputs & Total Cost Distributions Between CMB and Non-CMB Banks

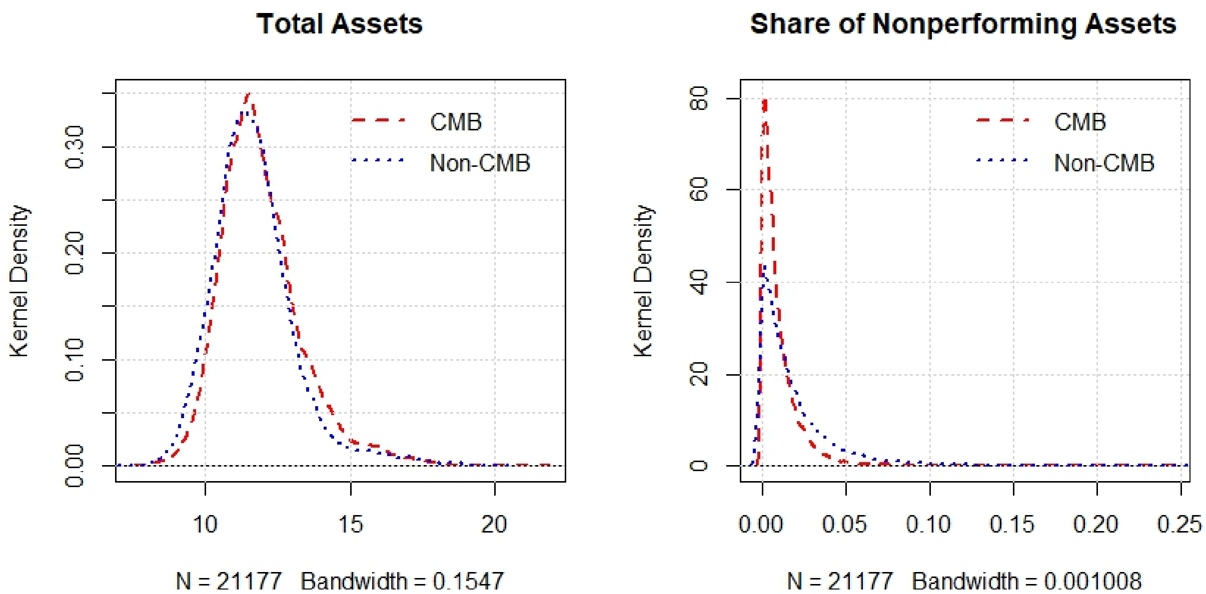


Figure 43: Assets & Nonperforming Assets Between CMB and Non-CMB Banks

3.17.1 Ridgeline Graph

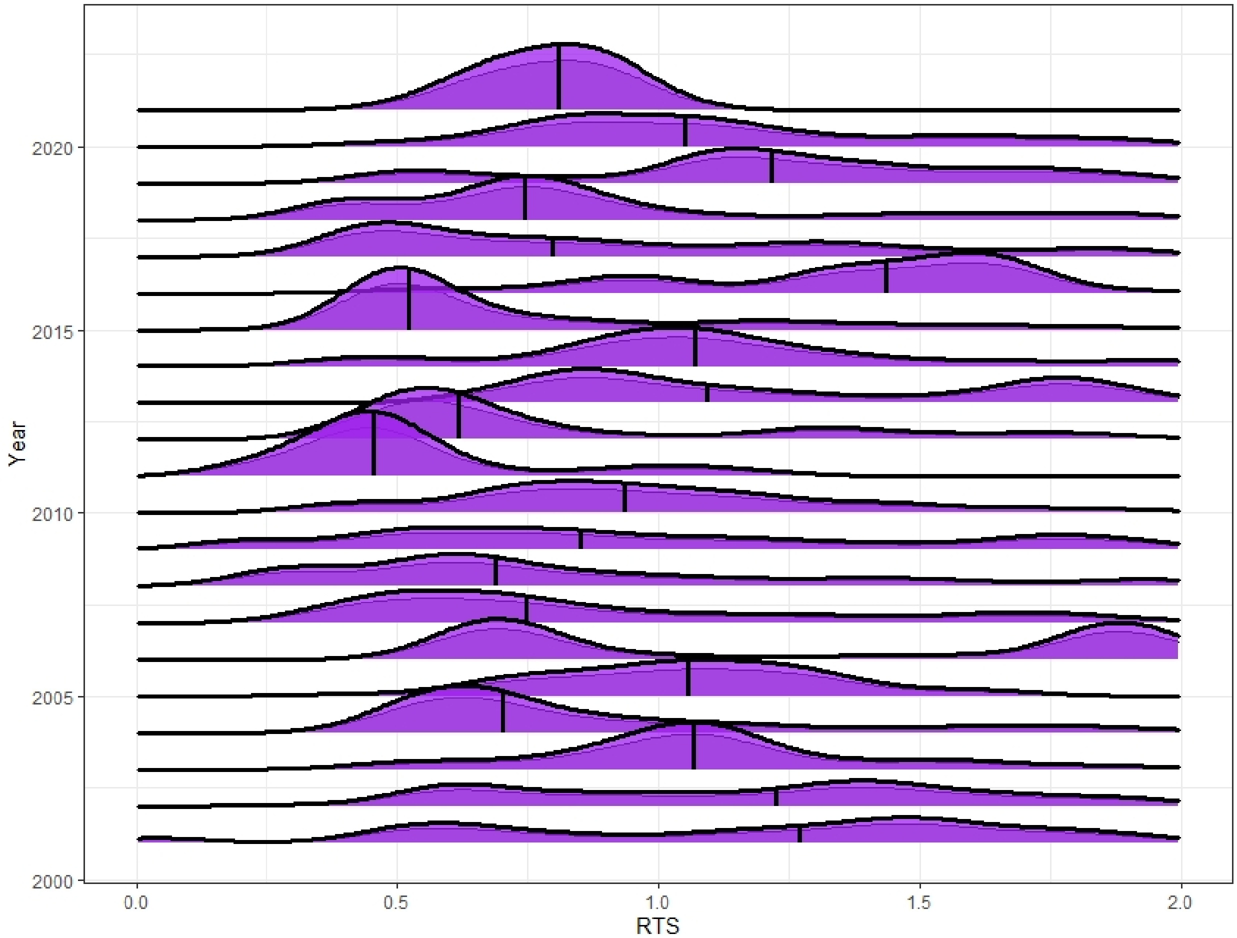


Figure 44: Densities of Nonperformance-Adjusted RTS Estimates for Critical Market Banks

3.18 Appendix I: Data Cleaning Procedure

For our study, we use the FDIC BankFind Suite: API for Data Miners & Developers. The data is publicly available in quarterly buckets and contains balance sheet and income statement data for all FDIC-insured depository institutions. This data is close in nature to the FDIC Call Report and, in fact, contains many of the same core income statement and balance sheet variables. The dimension of the BankFind Suite data is smaller when compared to the Call Report data, however. That being said, the common banking inputs and outputs utilized in the banking performance literature can be found in both data sets. Below we detail our explicit data cleaning procedure:

1. Before making variable-specific alterations, we convert quarterly stock variables in nominal terms to annual averages using the averages over years where data on all four quarters are available

2. For any output, y_q, we drop observations that have missing data
3. For the remaining outputs, we convert them to constant national prices using the GDP deflator with 2012 as our base year
4. As a precaution, we drop outputs that may take on negative values
5. For any given input quantity, x_j, we also deflate their nominal values by the GDP deflator with 2012 as our base year
6. We drop any observations that have neither employees or bank premises, which is to say we drop observations that correspondingly have $x_4 = 0$ and $x_5 = 0$
7. For any given expense, we deflate their nominal values by the GDP deflator with a base year of 2012
8. We derive unit input prices, w_j, by dividing each expense by their corresponding input quantities
9. Any input price that comes up as missing is recoded as zero with the exception of w_4 and w_5, which are dropped if missing
10. All $w_j < 0$ are dropped
11. We generate our cost-shares for any given input j as $S_j = (w_j \times x_j)/C$
12. We drop any observations where the sum of the individual cost shares are greater than one, equal to zero or less zero—this is an incredibly small quantity, but is done as a precaution

3.19 Appendix J: FDIC Data Glossary

Definitions for the specific variables used in this study were obtained from the FDIC BankFind Suite: API for Data Miners & Developers. This API has replaced the older portal for statistics on depository institutions (SDI), but fundamentally contains precisely the same data.

A **ASSET**: total assets on a consolidated basis

B **BKPREM**: premises and fixed assets

D **DEP**: total deposits

E **EDEP**: interest expense on total deposits

EFREPP: interest expense on Federal Funds purchased and securities sold under agreements to repurchase on a consolidated basis

EINTEXP: total interest expense on a consolidated basis

EPREMAGG: expenses of premises and fixed assets

ESAL: salaries and employee benefits on a consolidated basis

F **FREPP**: Federal Funds purchased and securities sold under agreement to repurchase on a consolidated basis

FREPO: Federal Funds sold and securities purchased under agreements to resell on a consolidated basis

L **LNCI**: commercial and industrial loans on a consolidated basis

LNCON: loans to individuals for household, family or other personal expenditures (consumer loans) on a consolidated basis

LNRE: loans secured by real estate on a consolidated basis

N **NALNLS**: total nonaccrual loans and leases financing receivables on a consolidated basis

NONII: total non-interest income on a consolidated basis

NONIX: total non-interest expense on a consolidated basis

NTR: represents deposits that are not included in the definition of transaction accounts above or that do not satisfy the criteria necessary to be defined as a transaction account

NUMEMP: number of full-time equivalent (FTE) employees on payroll at the end of the current period

O **ORE**: other real estate owned investments on a consolidated basis

P **P3LNLS**: total loans and leases financing receivables past due 30 through 89 days and still accruing interest on a consolidated basis

P9LNLS: total loans and leases financing receivables past due 90 days or more and still accruing interest on a consolidated basis

P9CI: dollar value of all commercial and industrial loans $>$ 90 past due

P9CON: dollar value of all consumer loans $>$ 90 past due

P9RE: dollar value of all real estate loans $>$ 90 past due

S **SC**: total investment securities (book value)

SCAGE: total US government agency and corporation obligations on a consolidated basis

SCMTGBK: mortgage-backed securities on a consolidated basis

SCUST: holdings of U.S. Treasury securities

T **TRN**: domestic deposits transactions; represents all demand deposits, NOW accounts, ATS accounts, accounts from which payments may be made to third parties by means of an automated teller machine, a remote service unit, or another electric device, and accounts that permit third party payments through the use of checks, drafts, negotiable instruments or other similar instrument

All variables, with the exception of **NUMEMP**, are denoted in thousands of current US dollars.

References

Afonso, G., Santos, J. A., & Traina, J. (2014). Do'too-big-to-fail'banks take on more risk? *Economic Policy Review*, *20*(2), 41–58.

Aihounton, G. B., & Henningsen, A. (2021). Units of measurement and the inverse hyperbolic sine transformation. *The Econometrics Journal*, *24*(2), 334–351.

Akcay, S. (2011). The causal relationship between producer price index and consumer price index: Empirical evidence from selected european countries. *International Journal of Economics and Finance*, *3*(6), 227–232.

Alchian, A. (1963). Reliability of progress curves in airframe production. *Econometrica: Journal of the Econometric Society*, 679–693.

Andrews, D., Criscuolo, C., & Gal, P. (2016). The global productivity slowdown, technology divergence and public policy: a firm level perspective. *Brookings Institution Hutchins Center Working Paper*, *24*.

Arrow, K. J. (1962). The economic implications of learning by doing. *The Review of Economic Studies*, *29*(3), 155–173.

Assaf, A. G., Berger, A. N., Roman, R. A., & Tsionas, M. G. (2019). Does efficiency help banks survive and thrive during financial crises? *Journal of Banking & Finance*, *106*, 445–470.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47–78.

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, *18*(1), 1–22.

Baily, M. N., Gordon, R. J., Nordhaus, W. D., & Romer, D. (1988). The productivity slowdown, measurement issues, and the explosion of computer power. *Brookings papers on economic activity*, *1988*(2), 347–431.

Barr, R. S., Seiford, L. M., & Siems, T. F. (1994). Forecasting bank failure: A non-parametric frontier estimation approach. *Recherches Économiques de Louvain/Louvain Economic Review*, *60*(4), 417–429.

Barth, J. R., & Wihlborg, C. (2016). Too big to fail and too big to save: dilemmas for banking reform. *National Institute Economic Review*, *235*, R27–R39.

Barzel, Y. (1972). The rate of technical progress: The "Indianapolis 500". *Journal of Economic Theory*, *4*(1), 72–81.

Basu, S., Fernald, J. G., & Kimball, M. S. (2006). Are technology improvements contractionary? *American Economic Review*, *96*(5), 1418–1448.

Bellemare, M. F., & Wichman, C. J. (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, *82*(1), 50–61.

Berger, A. N. (1993). "distribution-free" estimates of efficiency in the us banking industry and tests of the standard distributional assumptions. *Journal of productivity Analysis*, *4*(3), 261–292.

Berger, A. N., & Bouwman, C. H. (2013). How does capital affect bank performance during financial crises? *Journal of financial economics*, *109*(1), 146–176.

Berger, A. N., & DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks. *Journal of Banking & Finance*, *21*(6), 849–870.

Berger, A. N., Hancock, D., & Humphrey, D. B. (1993). Bank efficiency derived from the profit function. *Journal of Banking & Finance*, *17*(2-3), 317–347.

Berger, A. N., & Humphrey, D. B. (1997). Efficiency of financial institutions: International survey and directions for future research. *European journal of operational research*, *98*(2), 175–212.

Berger, A. N., & Mester, L. J. (1997). Inside the black box: What explains differences in the efficiencies of financial institutions? *Journal of banking & finance*, *21*(7), 895–947.

Berndt, E. R., & Wood, D. O. (1975). Technology, prices, and the derived demand for energy. *The review of Economics and Statistics*, 259–268.

Boppart, T., & Li, H. (2021). Productivity slowdown: reducing the measure of our ignorance.

Brynjolfsson, E., & Hitt, L. M. (2003). Computing productivity: Firm-level evidence. *Review of economics and statistics*, *85*(4), 793–808.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Byrne, D. M., Fernald, J. G., & Reinsdorf, M. B. (2016). Does the united states have a productivity slowdown or a measurement problem? *Brookings Papers on Economic Activity*, *2016*(1), 109–182.

Byrne, J. P., Perman, R., et al. (2006). *Unit roots and structural breaks: a survey of the literature.* University of Glasgow, Department of Economics.

Campa, J. M., & Goldberg, L. S. (2005). Exchange rate pass-through into import prices. *Review of Economics and Statistics*, *87*(4), 679–690.

Campbell, J. Y., & Mankiw, N. G. (1987). Are output fluctuations transitory? *The Quarterly Journal of Economics*, *102*(4), 857–880.

Caporale, G. M., Katsimi, M., & Pittis, N. (2002). Causality links between consumer and producer prices: some empirical evidence. *Southern Economic Journal*, *68*(3), 703–711.

Carrion-i Silvestre, J. L., Kim, D., & Perron, P. (2009). Gls-based unit root tests with multiple structural breaks under both the null and the alternative hypotheses. *Econometric theory*, *25*(6), 1754–1792.

Chaffai, M. E., Dietsch, M., & Lozano-Vivas, A. (2001). Technological and environmental differences in the european banking industries. *Journal of Financial Services Research*, *19*(2), 147–162.

Chambers, R. G., et al. (1988). *Applied production analysis: a dual approach.* Cambridge University Press.

Chandler, K. (1952). The distribution and frequency of record values. *Journal of the Royal Statistical Society: Series B (Methodological)*, *14*(2), 220–228.

Chen, S.-S. (2009). Oil price pass-through into inflation. *Energy economics*, *31*(1), 126–133.

Christensen, L. R., Jorgenson, D. W., & Lau, L. J. (1973). Transcendental logarithmic production frontiers. *The review of economics and statistics*, 28–45.

Christiano, L. J. (1992). Searching for a break in gnp. *Journal of Business & Economic Statistics*, *10*(3), 237–250.

Clark, J. A. (1996). Economic cost, scale efficiency, and competitive viability in banking. *Journal of Money, Credit and Banking*, *28*(3), 342–364.

Clark, T. E., et al. (1995). Do producer prices lead consumer prices? *Economic Review-Federal Reserve Bank of Kansas City*, *80*, 25–25.

Colwell, R. J., & Davis, E. P. (1992). Output and productivity in banking. *The Scandinavian Journal of Economics*, S111–S129.

De Gregorio, J., Landerretche, O., Neilson, C., Broda, C., & Rigobon, R. (2007). Another pass-through bites the dust? oil prices and inflation [with comments]. *Economia*, *7*(2), 155–208.

DeVaughn, M. L., & Leary, M. M. (2010). Antecedents of failure for newly chartered banks in the us banking industry. *Group & Organization Management*, *35*(5), 666–695.

Dietsch, M., & Lozano-Vivas, A. (2000). How the environment determines banking efficiency: A comparison between french and spanish industries. *Journal of Banking & Finance*, *24*(6), 985–1004.

Diewert, W. E. (1980). Capital and the theory of productivity measurement. *The American Economic Review*, *70*(2), 260–267.

Dosi, G., & Nelson, R. R. (2010). Technical change and industrial dynamics as evolutionary processes. *Handbook of the Economics of Innovation*, *1*, 51–127.

Drake, L., Hall, M. J., & Simper, R. (2006). The impact of macroeconomic and regulatory factors on bank efficiency: A non-parametric analysis of hong kong's banking system. *Journal of Banking & Finance*, *30*(5), 1443–1466.

Du, P., Parmeter, C. F., & Racine, J. S. (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statistica Sinica*, 1347–1371.

Durlauf, S. N., & Phillips, P. C. (1988). Trends versus random walks in time series analysis. *Econometrica: Journal of the Econometric Society*, 1333–1354.

Embrey, I. (2019). On the benefits of normalization in production functions.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251–276.

Fagerberg, J. (1994). Technology and international differences in growth rates. *Journal of economic Literature*, *32*(3), 1147–1175.

Fellner, W. (1969). Specific interpretations of learning by doing. *Journal of Economic Theory*, *1*(2), 119–140.

Feng, G., & Serletis, A. (2010). Efficiency, technical change, and returns to scale in large us banks: Panel data evidence from an output distance function satisfying theoretical regularity. *Journal of Banking & Finance*, *34*(1), 127–138.

Feng, G., & Zhang, X. (2012). Productivity and efficiency at large and community banks in the us: A bayesian true random effects stochastic distance frontier analysis. *Journal of Banking & Finance*, *36*(7), 1883–1895.

Feng, G., & Zhang, X. (2014). Returns to scale at large banks in the us: A random coefficient stochastic frontier approach. *Journal of Banking & Finance*, *39*, 135–145.

Fernald, J. G. (2015). Productivity and potential output before, during, and after the great recession. *NBER macroeconomics annual*, *29*(1), 1–51.

Fernald, J. G., Hall, R. E., Stock, J. H., & Watson, M. W. (2017). The disappointing recovery of output after 2009. *Brookings Papers on Economic Activity*, *2017*(1), 1–81.

Foerster, A., Hornstein, A., Sarte, P.-D., & Watson, M. W. (2019). *Aggregate implications of changing sectoral trends* (Working Paper No. 25867). National Bureau of Economic Research.

Foster, L., Haltiwanger, J. C., & Krizan, C. J. (2001). Aggregate productivity growth: Lessons from microeconomic evidence. In *New developments in productivity analysis* (pp. 303–372). University of Chicago Press.

Fuhrer, J. C. (1995). The phillips curve is alive and well. *New England Economic Review*, 41–57.

Gagnon, J. E., & Ihrig, J. (2004). Monetary policy and exchange rate pass-through. *International Journal of Finance & Economics*, *9*(4), 315–338.

García, J., & Rodríguez, P. (2002). The determinants of football match attendance revisited: Empirical evidence from the spanish football league. *Journal of Sports Economics*, *3*(1), 18–38.

Geng, X., & Sun, K. (2022). Estimation of a partially linear seemingly unrelated regressions model: application to a translog cost system. *Econometric Reviews*, 1–39.

Ghazali, M. F., Yee, O. A., & Muhammad, M. Z. (2008). Do producer prices cause consumer prices? some empirical evidence. *International Journal of Business and Management*, *3*(11), 78–82.

Gopinath, G., Itskhoki, O., & Rigobon, R. (2010). Currency choice and exchange rate pass-through. *American Economic Review*, *100*(1), 304–36.

Gordon, R. J. (1975). Alternative responses of policy to external supply shocks. *Brookings Papers on Economic Activity*, *1975*(1), 183–206.

Gordon, R. J. (2013). *The phillips curve is alive and well: Inflation and the nairu during the slow recovery* (Tech. Rep.). National Bureau of Economic Research.

Grammatikos, T., & Papanikolaou, N. I. (2018). "too-small-to-survive" versus "too-big-to-fail" banks: The two sides of the same coin. *Financial Markets, Institutions & Instruments*, *27*(3), 89–121.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438.

Griliches, Z. (1980). R & d and the productivity slowdown. *The American Economic Review*, *70*(2), 343–348.

Hahn, F. H., & Matthews, R. C. (1964). The theory of economic growth: a survey. *The Economic Journal*, *74*(296), 779–902.

Hall, P., & Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics*, *29*(3), 624–647.

Hamilton, J. D. (2018). Why you should never use the hodrick-prescott filter. *Review of Economics and Statistics*, *100*(5), 831–843.

Hansen, B. E. (2001). The new econometrics of structural change: dating breaks in us labour productivity. *Journal of Economic perspectives*, *15*(4), 117–128.

Hansen, H., & Gauthier, R. (1989). Factors affecting attendance at professional sport events. *Journal of sport management*, *3*(1), 15–32.

Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *55*(4), 757–779.

Henderson, D. J., Kumbhakar, S. C., Li, Q., & Parmeter, C. F. (2015). Smooth coefficient estimation of a seemingly unrelated regression. *Journal of Econometrics*, *189*(1), 148–162.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, 1–16.

Huang, K. X., & Liu, Z. (2005). Inflation targeting: What inflation rate to target? *Journal of Monetary Economics*, *52*(8), 1435–1462.

Hughes, J. P., & Mester, L. J. (1993). A quality and risk-adjusted cost function for banks: Evidence on the "too-big-to-fail" doctrine. *Journal of productivity analysis*, *4*(3), 293–315.

Hughes, J. P., Mester, L. J., & Moon, C.-G. (2001). Are scale economies in banking elusive or illusive?: Evidence obtained by incorporating capital structure and risk-taking into models of bank production. *Journal of Banking & Finance*, *25*(12), 2169–2208.

Hulten, C. R. (2001). Total factor productivity: a short biography. In *New developments in productivity analysis* (pp. 1–54). University of Chicago Press.

Ireland, P. (2015). Comment on: on the stability of money demand, by robert e. lucas jr. and juan pablo nicolini. *Journal of Monetary Economics*, *73*(7), 66–9.

Ireland, P. N. (1999). Does the time-consistency problem explain the behavior of inflation in the united states? *Journal of Monetary Economics*, *44*(2), 279–291.

Ireland, P. N., & Schuh, S. (2008). Productivity and us macroeconomic performance: Interpreting the past and predicting the future with a two-sector real business cycle model. *Review of Economic Dynamics*, *11*(3), 473–492.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, *12*(2-3), 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica: journal of the Econometric Society*, 1551–1580.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models.* OUP Oxford.

Jorgenson, D. W. (1991). Productivity and economic growth. In *Fifty years of economic measurement: The jubilee of the conference on research in income and wealth* (pp. 19–118).

Jorgenson, D. W., Ho, M. S., Samuels, J. D., & Stiroh, K. J. (2007). Industry origins of the american productivity resurgence. *Economic systems research*, *19*(3), 229–252.

Jorgenson, D. W., Ho, M. S., & Stiroh, K. J. (2008). A retrospective look at the us productivity growth resurgence. *Journal of Economic perspectives*, *22*(1), 3–24.

Jovanovic, B. (1996). Learning by doing and the choice of technology. *Econometrica*, *64*(6), 1299–1310.

Kaufman, G. G. (2014). Too big to fail in banking: What does it mean? *Journal of Financial Stability*, *13*, 214–223.

Kennedy, C., & Thirlwall, A. P. (1972). Surveys in applied economics: technical progress. *The Economic Journal*, *82*(325), 11–72.

King, R. G., Plosser, C. I., Stock, J. H., & Watson, M. W. (1987). *Stochastic trends and economic fluctuations.* National Bureau of Economic Research Cambridge, Mass., USA.

Kumbhakar, S. C. (1997). Modeling allocative inefficiency in a translog cost function and cost share equations: an exact relationship. *Journal of Econometrics*, *76*(1-2), 351–356.

Kydland, F. E., & Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, 1345–1370.

Levitt, S. D., List, J. A., & Syverson, C. (2013). Toward an understanding of learning by doing: Evidence from an automobile assembly plant. *Journal of political Economy*, *121*(4), 643–681.

Li, Q., & Racine, J. S. (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press.

Liu, K. (2019). The impact of the dodd-frank act on small us banks. *Available at SSRN 3419586*.

Long Jr, J. B., & Plosser, C. I. (1983). Real business cycles. *Journal of political Economy*, *91*(1), 39–69.

Lozano-Vivas, A., Pastor, J. T., & Pastor, J. M. (2002). An efficiency comparison of european banking systems operating under different environmental conditions. *journal of Productivity Analysis*, *18*(1), 59–77.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, *1*, 19-46. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167223176800036 doi: https://doi.org/10.1016/S0167-2231(76)80003-6

Lucas Jr, R. E., & Nicolini, J. P. (2015). On the stability of money demand. *Journal of Monetary Economics*, *73*, 48–65.

Makri, V., Tsagkanos, A., & Bellas, A. (2014). Determinants of non-performing loans: The case of eurozone. *Panoeconomicus*, *61*(2), 193–206.

Mantel Jr, S. J., Rosegger, G., & Mantel, S. P. (1995). Managing technology at the Indianapolis 500. *Technological Forecasting and Social Change*, *48*(1), 59–76.

McAllister, P. H., & McManus, D. (1993). Resolving the scale efficiency puzzle in banking. *Journal of Banking & Finance*, *17*(2-3), 389–405.

Menon, J. (1995). Exchange rate pass-through. *Journal of Economic Surveys*, *9*(2), 197–231.

Minuci, E. G. (2021). *Innovations in money and banking markets*. West Virginia University.

Mishkin, F. S. (2006). How big a problem is too big to fail? a review of gary stern and ron feldman's too big to fail: the hazards of bank bailouts. *Journal of economic literature*, *44*(4), 988–1004.

Mitchell, K., & Onvural, N. M. (1996). Economies of scale and scope at large commercial banks: Evidence from the fourier flexible functional form. *Journal of Money, Credit and Banking*, *28*(2), 178–199.

Morrison, C. J. (2012). *A microeconomic approach to the measurement of economic performance: Productivity growth, capacity utilization, and related performance indicators*. Springer Science & Business Media.

Munasinghe, L., O'Flaherty, B., & Danninger, S. (2001). Globalization and the rate of technological progress: What track and field records show. *Journal of Political Economy*, *109*(5), 1132–1149.

Nelson, C. R., & Plosser, C. R. (1982). Trends and random walks in macroeconmic time series: some evidence and implications. *Journal of monetary economics*, *10*(2), 139–162.

Newey, W. K., & West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.

Nkusu, M. M. (2011). *Nonperforming loans and macrofinancial vulnerabilities in advanced economies*. International Monetary Fund.

Nordhaus, W. D. (2021). Are we approaching an economic singularity? information technology and the future of economic growth. *American Economic Journal: Macroeconomics*, *13*(1), 299–332.

Parmeter, C. F., Sun, K., Henderson, D. J., & Kumbhakar, S. C. (2014). Estimation and inference under economic restrictions. *Journal of productivity analysis*, *41*, 111–129.

Peirce, H., Robinson, I., & Stratmann, T. (2014). How are small banks faring under dodd-frank?

Prescott, E. C. (1998). Lawrence r. klein lecture 1997: Needed: A theory of total factor productivity. *International economic review*, 525–551.

Preston, E. C., & Johnson, D. K. (2015). Fastest in the pool: The role of technological innovation on swimming record breaks. *International Journal of Innovation, Management and Technology*, *6*(5).

Restrepo-Tobón, D., & Kumbhakar, S. C. (2015). Nonparametric estimation of returns to scale using input distance functions: an application to large us banks. *Empirical Economics*, *48*(1), 143–168.

Romer, P. M. (1987). Crazy explanations for the productivity slowdown. *NBER macroeconomics annual*, *2*, 163–202.

Sandmo, A. (1970). Returns to scale and the average cost curve. *The Swedish Journal of Economics*, *72*(2), 149–152.

Sapci, A., & Miles, B. (2019). Bank size, returns to scale, and cost efficiency. *Journal of Economics and Business*, *105*, 105842.

Shapiro, A., & Wilson, D. J. (2019). The evolution of the fomc's explicit inflation target. *Evolution*, *2019*, 12.

Sharma, D., Sharma, A. K., & Barua, M. K. (2013). Efficiency and productivity of banking sector: A critical analysis of literature and design of conceptual model. *Qualitative Research in Financial Markets*.

Sichel, D. E. (2019). Productivity measurement: Racing to keep up. *Annual Review of Economics*, *11*(1), 591-614.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics*, *70*(1), 65–94.

Solow, R. M. (1997). *Learning from 'learning by doing': Lessons for economic growth.* Stanford University Press.

Stern, G. H. (2012). Too-big-to-fail and the dodd-frank legislation. *Hamline L. Rev.*, *35*, 339.

Stern, G. H., & Feldman, R. J. (2004). *Too big to fail: The hazards of bank bailouts.* Brookings Institution Press.

Stock, J. H. (1994). Unit roots, structural breaks and trends. *Handbook of econometrics*, *4*, 2739–2841.

Stock, J. H., & Watson, M. W. (1988). Testing for common trends. *Journal of the American statistical Association*, *83*(404), 1097–1107.

Stock, J. H., & Watson, M. W. (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and banking*, *39*, 3–33.

Swan, T. W. (1956). Economic growth and capital accumulation. *Economic record*, *32*(2), 334–361.

Syverson, C. (2017). Challenges to mismeasurement explanations for the us productivity slowdown. *Journal of Economic Perspectives*, *31*(2), 165–86.

Toda, H. Y., & Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of econometrics*, *66*(1-2), 225–250.

Topuz, Y. V., Yazdifar, H., & Sahadev, S. (2018). The relation between the producer and consumer price indices: a two-country study. *Journal of Revenue and Pricing Management*, *17*(3), 122–130.

Van Beveren, I. (2012). Total factor productivity estimation: A practical review. *Journal of economic surveys*, *26*(1), 98–128.

Vu, K., Hanafizadeh, P., & Bohlin, E. (2020). Ict as a driver of economic growth: A survey of the literature and directions for future research. *Telecommunications Policy*, *44*(2), 101922.

Wang, J. C. (2003). Productivity and economies of scale in the production of bank service value added. *Available at SSRN 648142*.

Wheelock, D. C., & Wilson, P. W. (1999). Technical progress, inefficiency, and productivity change in us banking, 1984-1993. *Journal of Money, Credit, and Banking*, 212–234.

Wheelock, D. C., & Wilson, P. W. (2009). *Are U.S. banks too large?* (Tech. Rep.).

Wheelock, D. C., & Wilson, P. W. (2012). Do large banks have lower costs? new estimates of returns to scale for us banks. *Journal of Money, Credit and Banking*, *44*(1), 171–199.

Wheelock, D. C., & Wilson, P. W. (2018). The evolution of scale economies in us banking. *Journal of Applied Econometrics*, *33*(1), 16–28.

White, H. (1980). Using least squares to approximate unknown regression functions. *International economic review*, 149–170.

Wickens, M. R. (1996). Interpreting cointegrating vectors and common stochastic trends. *Journal of econometrics*, *74*(2), 255–271.

Wolff, E. N. (1985). Industrial composition, interindustry effects, and the us productivity slowdown. *The Review of Economics and Statistics*, 268–277.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association*, *57*(298), 348–368.

Zhou, C. (2009). Are banks too big to fail? measuring systemic importance of financial institutions. *Measuring Systemic Importance of Financial Institutions (December 1, 2009)*.