

FEDERAL RESERVE BANK OF ST. LOUIS

# REVIEW

SECOND QUARTER 2022  
VOLUME 104 | NUMBER 2

**Global Supply Chain Disruptions and Inflation  
During the COVID-19 Pandemic**

Ana Maria Santacreu and Jesse LaBelle

**Who Should Work from Home During a Pandemic?  
The Wage-Infection Trade-Off**

Sangmin Aum, Sang Yoon (Tim) Lee, and Yongseok Shin

**Failing to Provide Public Goods:  
Why the Afghan Army Did Not Fight**

Rohan Dutta, David K. Levine, and Salvatore Modica

**Venture Capital: A Catalyst for Innovation and Growth**

Jeremy Greenwood, Pengfei Han, and Juan M. Sánchez

**On the Relative Performance of Inflation Forecasts**

Julie K. Bennett and Michael T. Owyang

# REVIEW

Volume 104 • Number 2



## President and CEO

James Bullard

## Director of Research

Carlos Garriga

## Deputy Director of Research

B. Ravikumar

## Review Editors-in-Chief

Michael T. Owyang

Juan M. Sánchez

## Special Policy Advisor

David C. Wheelock

## Economists

David Andolfatto

Subhayu Bandyopadhyay

Serdar Birinci

Yu-Ting Chiang

YiLi Chien

Riccardo DiCecio

William Dupor

Maximiliano Dvorkin

Miguel Faria-e-Castro

Charles S. Gascon

Victoria Gregory

Nathan Jefferson

Kevin L. Kliesen

Julian Kozlowski

Fernando Leibovici

Oksana Leukhina

Fernando M. Martin

Michael W. McCracken

Amanda M. Michaud

Alexander Monge-Naranjo

Christopher J. Neely

Serdar Ozkan

Paulina Restrepo-Echavarría

Hannah Rubinton

Ana Maria Santacreu

Guillaume Vandenbroucke

Christian M. Zimmermann

## Managing Editor

Lydia H. Johnson

## Contributing Editors

George E. Fortier

Jennifer M. Ives

## Designer | Production Coordinator

Donna M. Stiller

78

### Global Supply Chain Disruptions and Inflation During the COVID-19 Pandemic

*Ana Maria Santacreu and Jesse LaBelle*

92

### Who Should Work from Home During a Pandemic? The Wage-Infection Trade-Off

*Sangmin Aum, Sang Yoon (Tim) Lee, and Yongseok Shin*

110

### Failing to Provide Public Goods: Why the Afghan Army Did Not Fight

*Rohan Dutta, David K. Levine, and Salvatore Modica*

120

### Venture Capital: A Catalyst for Innovation and Growth

*Jeremy Greenwood, Pengfei Han, and Juan M. Sánchez*

131

### On the Relative Performance of Inflation Forecasts

*Julie K. Bennett and Michael T. Owyang*

### **Review**

*Review* is published four times per year by the Research Division of the Federal Reserve Bank of St. Louis. Full online access is available to all, free of charge.

### **Online Access to Current and Past Issues**

The current issue and past issues dating back to 1967 may be accessed through our Research Division website: <http://research.stlouisfed.org/publications/review>. All nonproprietary and nonconfidential data and programs for the articles written by Federal Reserve Bank of St. Louis staff and published in *Review* also are available to our readers on this website.

*Review* articles published before 1967 may be accessed through our digital archive, FRASER: <http://fraser.stlouisfed.org/publication/?pid=820>.

*Review* is indexed in Fed in Print, the catalog of Federal Reserve publications (<http://www.fedinprint.org/>), and in IDEAS/RePEc, the free online bibliography hosted by the Research Division (<http://ideas.repec.org/>).

### **Subscriptions and Alerts**

The *Review* is no longer printed or mailed to subscribers. Our last printed issue was the first quarter of 2020.

Our monthly email newsletter keeps you informed when new issues of *Review*, *Economic Synopses*, *Regional Economist*, and other publications become available; it also alerts you to new or enhanced data and information services provided by the St. Louis Fed. Subscribe to the newsletter here: <http://research.stlouisfed.org/newsletter-subscribe.html>.

### **Authorship and Disclaimer**

The majority of research published in *Review* is authored by economists on staff at the Federal Reserve Bank of St. Louis. Visiting scholars and others affiliated with the St. Louis Fed or the Federal Reserve System occasionally provide content as well. *Review* does not accept unsolicited manuscripts for publication.

The views expressed in *Review* are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

### **Copyright and Permissions**

Articles may be reprinted, reproduced, republished, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. In these cases, there is no need to request written permission or approval. Please send a copy of any reprinted or republished materials to *Review*, Research Division of the Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166-0442; [STLS.Research.Publications@stls.frb.org](mailto:STLS.Research.Publications@stls.frb.org).

Please note that any abstracts, synopses, translations, or other derivative work based on content published in *Review* may be made only with prior written permission of the Federal Reserve Bank of St. Louis. Please contact the *Review* editor at the above address to request this permission.

### **Economic Data**

General economic data can be obtained through FRED®, our free database with over 800,000 national, international, and regional data series, including data for our own Eighth Federal Reserve District. You may access FRED through our website: <https://fred.stlouisfed.org>.

# Global Supply Chain Disruptions and Inflation During the COVID-19 Pandemic

*Ana Maria Santacreu and Jesse LaBelle*

We investigate the role supply chain disruptions during the COVID-19 pandemic played in U.S. producer price index (PPI) inflation. We exploit pre-pandemic cross-industry variation in sourcing patterns across countries and interact it with measures of international supply chain bottlenecks during the pandemic. We show that exposure to global supply chain disruptions played a significant role in U.S. cross-industry PPI inflation between January and November 2021. If bottlenecks had followed the same path as in 2019, PPI inflation in the manufacturing sector would have been 2 percentage points lower in January 2021 and 20 percentage points lower in November 2021. (JEL F13, F14, F44)

Federal Reserve Bank of St. Louis *Review*, Second Quarter 2022, 104(2), pp. 78-91.  
<https://doi.org/10.20955/r.104.78-91>

---

## 1 INTRODUCTION

The COVID-19-pandemic recession and recovery have been unique compared with previous recessions, largely due to policies that led to behavioral changes. Lockdowns meant people were traveling less both for work and for leisure, eating out less, and going to fewer entertainment venues, among other things. At the same time, work from home and fiscal stimulus packages increased the demand for certain goods such as technological goods, cars, and furniture. These changes resulted in an overall shift away from consumption of services and toward consumption of durable goods.

The rapid increase in the demand for durable goods, together with the global nature of the pandemic, has exposed vulnerabilities in the current production structure of these goods. Over the past several decades, production of durable goods has become more fragmented, relying heavily on global value chain (GVCs).<sup>1</sup> Instead of doing everything in-house, firms can outsource parts of their production processes to other countries. Figure 1 shows that GVC participation has been rising steadily over time, though it has plateaued in recent years (see Antrás, 2020).

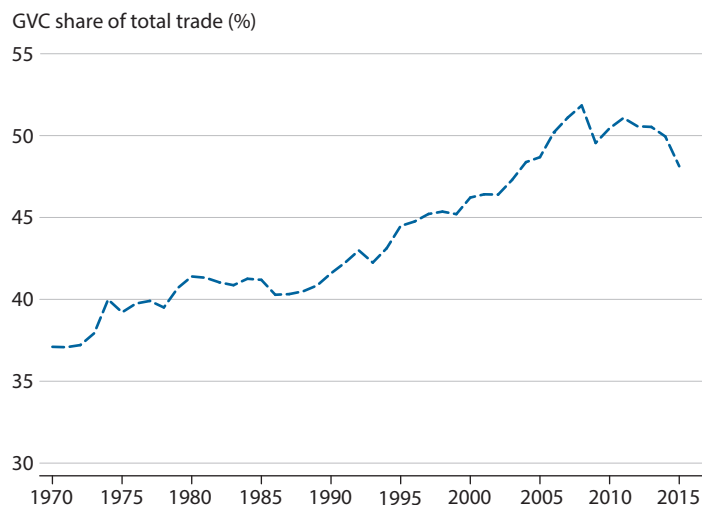
While GVC participation has advantages, as firms can benefit from outsourcing production to regions with a comparative advantage, it comes with risks (Santacreu and LaBelle, 2021a,b). Shocks

Ana Maria Santacreu is a senior economist and Jesse LaBelle is a research associate at the Federal Reserve Bank of St. Louis. We thank Julian Kozlowski for very insightful comments and suggestions.

© 2022, Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

## Figure 1

### GVC Participation Over Time, 1970-2015



NOTE: The figure represents the evolution in the share of total trade that requires inputs from at least two countries.

SOURCE: World Bank *World Development Report 2020*.

that hit a particular stage of the production process can propagate along the chain and expose firms dependent on inputs from these regions. Some of these risks did materialize during the current pandemic through global lockdowns (Leibovici, and Santacreu, and LaBelle, 2021), low vaccination rates in emerging countries (Çakmaklı et al., 2021), and large shipping costs and disruptions in some key ports, putting additional pressure on supply chains.<sup>2</sup>

These risks can be exacerbated when supply chains rely heavily on critical inputs from one or a few regions. Take the example of semiconductors. The advancement of technology in nearly every product has made semiconductors a vastly important input for the entire economy; however, their production largely relies on a few countries, such as Taiwan and China. A sharp increase in the demand for products that use this input may create large bottlenecks in semiconductor-dependent industries. Therefore, due to the global nature of supply chains, even a relatively small demand shock to a critical sector can propagate into a larger supply/demand disruption. This mismatch between supply and demand puts upward pressure on prices. In this article, we address the following question: To what extent has the global nature of supply chain disruptions contributed to producer price index (PPI) inflation across U.S. sectors?<sup>3</sup>

The main challenge to answer this question is the limited access to real-time data on supply chain disruptions. We rely on the Purchasing Managers' Index data from S&P Global. These data, which are available with a subscription, comprise monthly surveys sent to senior executives at private firms in 44 countries. We focus on two measures from this survey that capture supply chain disruptions: backlogs and supplier delivery times. Backlogs measure how much the number of unfulfilled new orders has changed from the previous month; delivery times measure how much the average time it takes for suppliers to deliver inputs has changed from the previous month. Each variable represents a rate of change over the previous month, and both capture demand and supply effects.

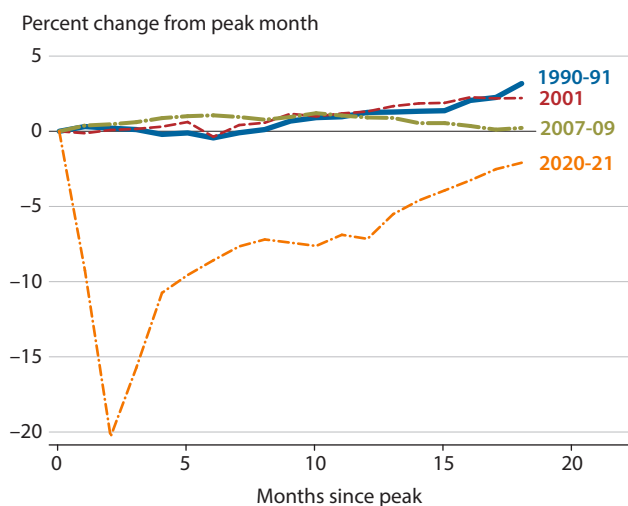
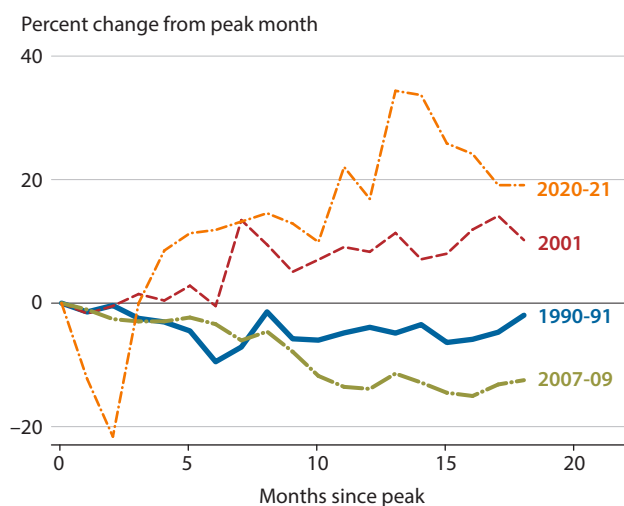
Higher backlogs typically indicate that demand is increasing at a rate producers cannot meet, while the opposite indicates unused production capacity resulting from a lack of demand. Hence, backlogs measure how quickly suppliers can keep up with demand. The same logic applies to delivery times. As such, these measures can be used to infer demand and supply mismatches that contribute to price increases and inflation.

We begin by documenting three salient features of the data on supply chain disruptions. First, bottlenecks have become worse since January 2021, as implied by an increasing number of unfulfilled orders and longer delivery times. Second, backlogs and delivery times track PPI inflation quite well, with each having a correlation of about 90 percent for the period January 2020 to November 2021. Third, supply chain disruptions and their contribution to PPI inflation have been heterogeneous across industries. Backlogs increased sharply in the automobile and technology equipment industries. These increases were followed by large increases in PPI inflation. In the pharmaceutical industry, however, bottlenecks remained relatively steady, which were reflected in a steady increase in PPI inflation over the same period. These results suggest that the supply and demand mismatch was worse in the technology equipment sector and the automobile and auto parts sector than in the pharmaceutical sector.

We then ask the following question: Did U.S. industries more exposed to global supply chain disruptions experience higher PPI inflation over this time period? To answer this question formally, we construct measures of industry exposure to supply chain disruptions, both domestic and foreign.<sup>4</sup> In particular, we exploit heterogeneous variation in an industry's sourcing patterns across countries and interact it with our measures of supply chain disruption: changes in backlogs and in delivery times, respectively. If an industry in the United States relies heavily on intermediate inputs from a country where supply chain disruptions are severe, this industry will be more exposed. We consider both the manufacturing and non-manufacturing sectors. To the extent that for each industry we keep the value-added shares fixed at the levels in 2018, the interaction with the bottleneck variables in our exposure measure captures the role of the supply shock in that particular industry.

Our empirical strategy consists of regressing industry PPI inflation on our measures of domestic and foreign exposure, including industry fixed effects. We focus on the period January 2021 to November 2021. We find that both domestic and foreign exposure have a positive effect on industry PPI inflation. However, only foreign exposure is statistically significant. These results hold when using either backlogs or supplier delivery times as the measure of disruption. Moreover, the effects of global supply chain disruptions on PPI inflation are larger if the exposure variables are lagged by one month, suggesting that supply chain disruptions have a delayed impact on inflation. We then conduct the same regression analysis but separate the industries into manufacturing and non-manufacturing sectors. In the non-manufacturing sector, both domestic and foreign exposure have a positive and statistically significant effect on PPI inflation. In the manufacturing sector, however, only foreign exposure is statistically significant.

Finally, we ask the following question: What would PPI inflation have been during 2021 if backlogs in each country had followed their 2019 path? To answer this question, we do a back-of-the-envelope calculation in which we take the results from our regression analysis and compute a counterfactual PPI inflation rate using the data on bottlenecks from 2019. We find that PPI inflation in the manufacturing sector during 2021 would have been 2 percentage points lower in January and 20 percentage points lower in November 2021.

**Figure 2****Consumption Patterns During Recessions****A. Months following business cycle peaks: Services****B. Months following business cycle peaks: Durables**

NOTE: The figure shows the evolution of real consumption in services (Panel A) and nondurable goods (Panel B) 18 months after the business cycle peaks of the 1990-91, 2001, 2007-09, and 2020-21 recessions.

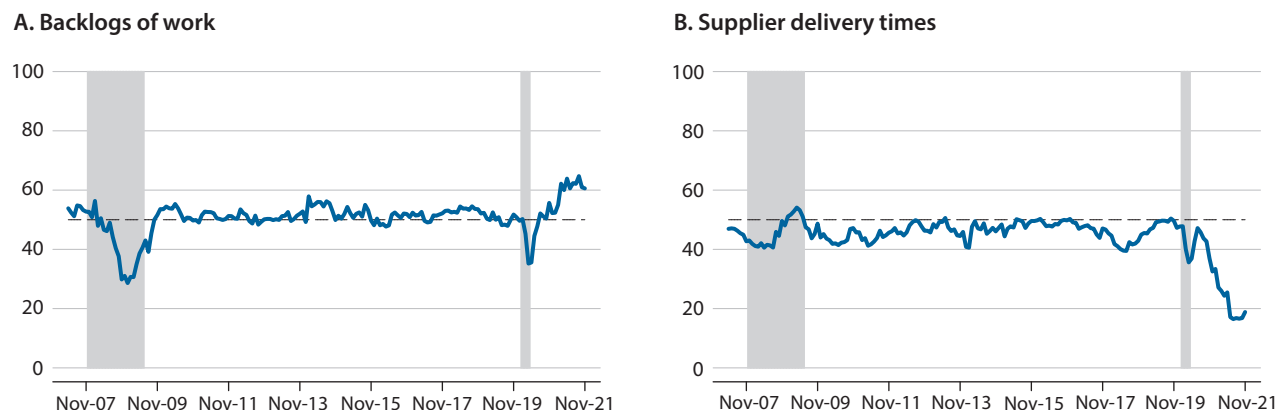
SOURCE: FRED®, Federal Reserve Bank of St. Louis; National Bureau of Economic Research; and authors' calculations.

Our results show that supply chain disruptions during the COVID-19 pandemic recession and recovery have been unprecedented. The shift in demand toward durable goods consumption and the heavy reliance on foreign suppliers to produce these goods has created a mismatch between supply and demand resulting in price increases. Sectors that rely more heavily on foreign inputs from countries that faced stronger disruptions experienced larger increases in PPI inflation.

This article complements a short but growing literature on inflation and supply chain disruptions. Ha, Kose, and Ohnsorge (2021) analyze the driving forces of global inflation, focusing on the 2020 global recession. Comin and Johnson (2020) study the role of trade integration and offshoring on inflation. Leibovici and Dunn (2021) discuss the extent to which supply chain disruptions account for the recent rise in inflation, focusing on the case of semiconductors. Finally, Amiti, Heise, and Wang (2021) study the effects of rising import prices on U.S. producer prices.

## 2 SUPPLY CHAIN DISRUPTIONS DURING COVID-19

The COVID-19 recession and recovery have been different from previous recessions and recoveries along several dimensions. One is related to consumption: There has been a shift in consumption away from services and toward durable goods. Figure 2 plots for the COVID-19 recession and recovery and three earlier ones, the evolution of real consumption of services (Panel A) and durable goods (Panel B) for 18 months after the business cycle peak (with consumption normalized to 1 in each business cycle peak).<sup>5</sup>

**Figure 3****Delivery Times and Backlogs Over Time, May 2007–November 2021**

NOTE: The figure shows the monthly evolution of backlogs (Panel A) and supplier delivery times (Panel B) in the United States. Gray bars indicate recessions as determined by the NBER.

SOURCE: S&P Global.

During a typical recession, services consumption tends to remain stable. In contrast, the COVID-19 recession was characterized by a sharp decline in consumption of services during the first months (over 20 percent in April 2020 from the peak in February 2020), which started recovering steadily after the initial shock. This recovery was helped when lockdowns were lifted and vaccines became widely available. Durable goods consumption, on the other hand, tends to drop and stay low for the duration of a recession and into the recovery, as consumers typically postpone consumption of these types of goods. During the COVID-19 recession, however, there was an initial sharp drop in durable goods consumption as expected; however, durables consumption recovered quickly and remained 19 percent higher than the peak even 18 months later.

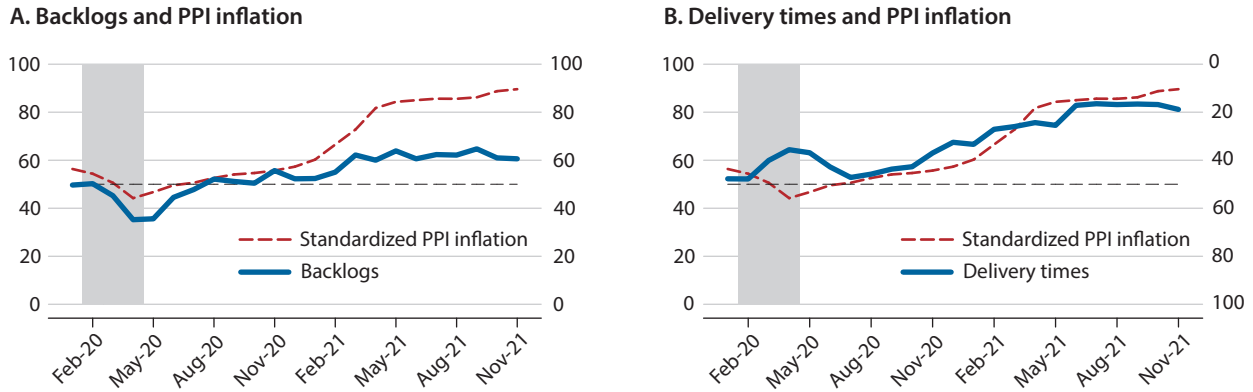
The shift of demand toward durables consumption, together with the fact that production of these goods takes place along complex supply chains, has translated into large bottlenecks. We show evidence of supply chain disruptions by plotting the evolution of backlogs (i.e., new orders that have not been completed or started yet) and supplier delivery times (i.e., the time it takes for a manufacturer to receive inputs from suppliers) from May 2007 to November 2021 in Figure 3. Index values greater than 50 represent increased backlogs (or lower delivery times) relative to those in the previous month, with the reverse being true for index values less than 50.<sup>6</sup>

Backlogs of work (Panel A) and supplier delivery times (Panel B) during the COVID-19 recession and recovery have behaved differently than during the Great Recession. The previous recession represents, in many ways, a typical demand shock: The rate of change of delivery times slightly increased before quickly recovering (see Figure 3). By the end of the recession in June 2009, delivery times had been actually getting shorter on a monthly basis for the previous six months. During this time, backlogs were gradually disappearing as the recession deepened. In fact, there were no month-over-month backlog increases, denoted by index values greater than 50 from May 2008 to October 2009, four months after the recession officially ended. For the COVID-19 recession, delivery times consistently worsened on a monthly basis but this rate flattened starting in June 2021. At the same



**Figure 4**

**Backlogs, Delivery Times, and PPI Inflation, January 2020–November 2021**

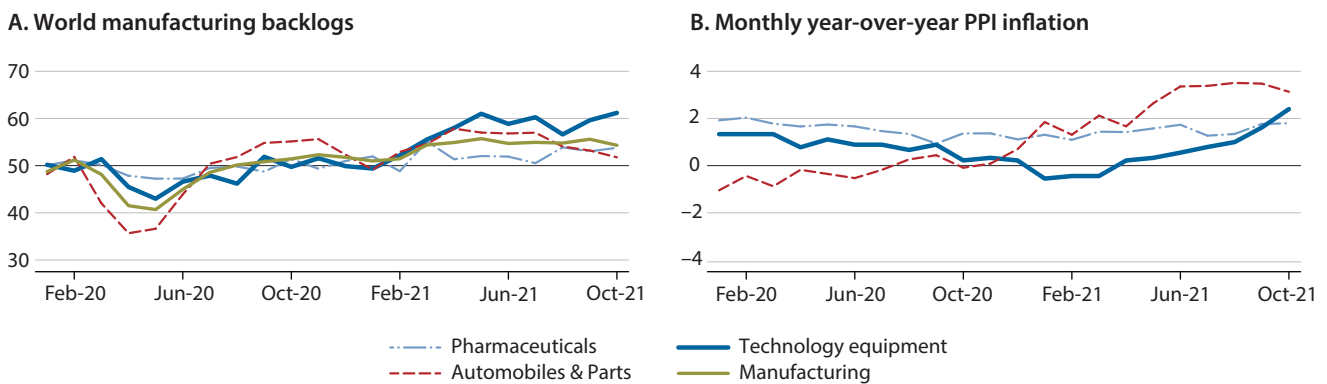


NOTE: The figure shows the monthly evolution of backlogs and PPI inflation (Panel A) and suppliers’ delivery times and PPI inflation (Panel B) in the United States. Delivery times are plotted on the right y-axis with an inverted scale so that higher values represent longer delivery times. Gray bars indicate the COVID-19 recession as determined by the NBER.

SOURCE: S&P Global, BLS, and authors’ calculations.

**Figure 5**

**Backlogs and PPI Inflation by Sector, January 2020–October 2021**



NOTE: The figure shows the monthly evolution of world backlogs (Panel A) and US PPI inflation (Panel B) for three sectors: automobiles and parts, technology equipment, and pharmaceuticals.

SOURCE: S&P Global, BLS, and authors’ calculations.

time, backlogs initially experienced the typical loosening associated with a drop in demand before the supply chain shocks caused an even larger distortion and forced higher levels of backlogs. Backlogs have consistently worsened on a monthly basis since August 2020.

These findings illustrate an unprecedented supply and demand mismatch, contributing to price increases and, hence, inflation. Focusing on the period from January 2020 to November 2021, Figure 4 shows that bottlenecks, measured either with backlogs (Panel A) or delivery times (Panel B; the y-axis is inverted so that higher values represent longer delivery times) track current PPI inflation closely.<sup>7</sup> Indeed, the correlation of PPI inflation with backlogs and with delivery times between January 2020 and November 2021 are each about 90 percent.

The evidence reported in Figure 4 masks large cross-sector heterogeneity. Figure 5 plots backlogs (Panel A) for the world in three manufacturing sectors: automobiles and auto parts; technology equipment; and pharmaceuticals.<sup>8</sup> The world automobile and auto parts sector started experiencing tightening of supply chains by July 2020, which manifested as consistent increases in the monthly rate of change in backlogs. Strong demand for cars in the months following the start of the pandemic, paired with disruptions in the supply of certain key inputs such as semiconductors, led to large supply chain disruptions in this sector. In the case of the technology equipment sector, unused capacity remained relatively stable for months after bottoming out before beginning to tighten at a steep slope after the turn of the new year. The COVID-19 pandemic substantially increased demand for computers and electronics, as people started working from home and fiscal stimulus increased consumption of these goods. As a result, PPI inflation increased substantially in these sectors (Panel B). The pharmaceutical sector behaved differently from the automobile and auto parts sector and the technology equipment sector: Bottlenecks and PPI inflation remained relatively steady in comparison. Hence, sectors that faced worse supply chain disruptions (i.e., automobile and auto parts, and technology equipment) also experienced steeper price increases.<sup>9</sup>

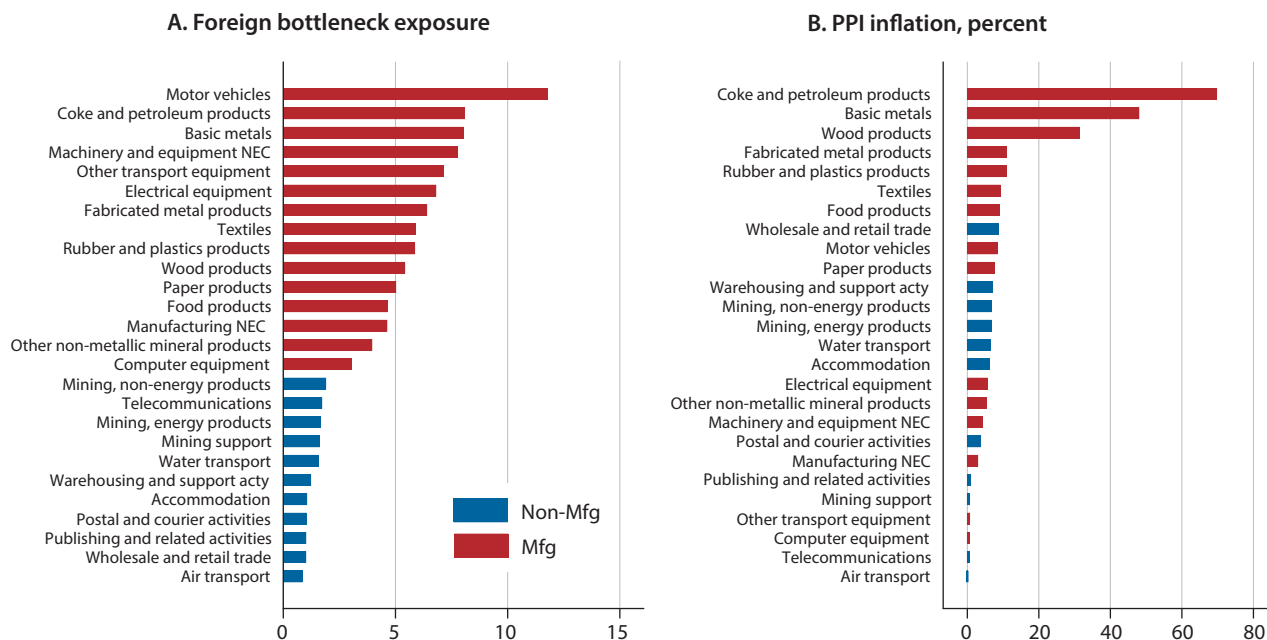
### 3 GLOBAL SUPPLY CHAIN DISRUPTIONS AND PPI INFLATION

In this section, we investigate the channels through which an exposure to global supply chain disruptions may have contributed to inflation during the COVID-19 pandemic. Bottlenecks in an industry can be driven by domestic or foreign factors or both. For instance, if an industry relies heavily on intermediate inputs from countries that experience more bottlenecks, that industry will be more exposed to foreign supply chain disruptions. If demand for that industry's products increases quickly, then foreign exposure may lead to price increases. In this section, we ask the following question: To what extent did exposure to domestic and foreign supply chain disruptions contribute to U.S. PPI inflation between January and November 2021? To answer this question we construct, for each industry of the United States, a domestic measure and a foreign measure of exposure to supply chain disruptions. Our empirical strategy consists of regressing PPI inflation on the exposure measures.

We follow the same methodology employed in Leibovici, Santacreu, and LaBelle (2021) and compute for each industry in the United States a measure of GVC participation—the share of gross exports (GE) produced with foreign value added (FVA) in 2018—for 32 countries and 26 industries, 15 of which correspond to the manufacturing sector.<sup>10</sup> This measure captures how much of the U.S. GE in a particular industry rely on intermediate imports from other countries. We then interact

**Figure 6**

**Average Foreign Bottleneck Exposure and PPI Inflation, January-November 2021**



NOTE: The figure shows for 26 U.S. industry averages for exposure to foreign backlogs (Panel A) and year-over-year PPI inflation (Panel B). Red bars represent manufacturing industries; blue bars represent non-manufacturing services industries. NEC, not elsewhere classified; acty, activity.

SOURCE: TIVA, S&P Global, BLS, and authors' calculations.

this variable with a measure of supply chain disruptions for each foreign supplier.<sup>11</sup> Our conjecture is that industries that are more exposed to global bottlenecks through GVCs experienced larger increases in PPI inflation.<sup>12</sup>

Industry  $i$ 's exposure to foreign ( $f$ ) bottlenecks at time  $t$ ,  $E_{it}^f$ , is computed as

$$(1) \quad E_{it}^f = \sum_{j=1}^N \frac{FVA_i^j}{GE_i} B_t^j,$$

where  $\frac{FVA_i^j}{GE_i}$  is the share of GE from industry  $i$  that are composed of value added from country  $j$  in that industry;  $B_t^j$  represents bottlenecks, either backlogs or delivery times, in country  $j$  at time  $t$ ; and  $N$  is the number of foreign suppliers. A period is a month. We restrict the analysis to the period January 2021 to November 2021.

Similarly, we compute a measure of industry's exposure to domestic bottlenecks defined as

$$(2) \quad E_{it}^d = \frac{DVA_i^{US}}{GE_i} B_t,$$

where  $\frac{DVA_i^{US}}{GE_i}$  is the share of value added embedded in U.S. GE supplied by the United States itself.  $B_t$  is the U.S. bottleneck variable, either backlogs or delivery times.

**Table 1****Exposure to Supply Chain Disruptions and PPI Inflation: Backlogs vs. Delivery Times, January–November 2021**

	Backlogs	Delivery times (inverse)
Domestic exposure	0.00569 (0.004)	0.00314 (0.003)
Foreign exposure	0.239*** (0.071)	0.255*** (0.070)
Constant	−0.598*** (0.115)	2.272*** (0.565)
Industry FE	YES	YES
<i>N</i>	286	165
<i>R</i> <sup>2</sup>	0.752	0.756

NOTE: FE, fixed effects. Standard errors are in parentheses. \*\*\*  $p < 0.001$ .

SOURCE: Authors' calculations.

Figure 6 plots our measure of foreign exposure computed in equation (1) (Panel A) and PPI inflation (Panel B) for the 26 industries in the United States, averaged for January to November 2021. Manufacturing industries are, on average, more exposed to foreign bottlenecks than services industries. In the manufacturing sector, motor vehicles, coke and petroleum products, and basic metals are the most exposed industries. The reason is twofold. On the one hand, these industries rely heavily on foreign intermediate inputs. On the other hand, the main suppliers in these industries have faced strong supply chain disruptions during the pandemic. Consistent with the measure of foreign exposure, manufacturing industries experienced higher PPI inflation than services industries. In the manufacturing sector, the coke and petroleum products industry and the basic metals industry are among the industries with the highest increases in prices. Therefore, there appears to be a positive correlation between exposure to foreign supply chain disruptions and PPI inflation.

### 3.1 Empirical Strategy

Next, we study more formally the extent to which domestic and foreign exposure to supply chain disruptions may have contributed to U.S. PPI inflation. In particular, we conduct the following linear regression:

$$(3) \quad \pi_{it}^{\text{PPI}} = \alpha_0 + \alpha_1 E_{it}^f + \alpha_2 E_{it}^d + I_i + u_{it},$$

where  $\pi_{it}^{\text{PPI}}$  represents the year-over-year PPI increase in industry  $i$  at time  $t$ ;  $E_{it}^f$  is exposure to foreign bottlenecks at time  $t$ ;  $E_{it}^d$  is the exposure to domestic bottlenecks in industry  $i$ ;  $I_i$  captures industry fixed effects; and  $u_{it}$  is the error term.

Table 1 reports the results. Foreign exposure, both in terms of supplier delivery times and backlogs, has a statistically significant effect on PPI inflation. For backlogs, increasing the month-over-month backlogs by 1 percent increases the industry inflation rate by 0.24 percentage points, while the same increase for delivery times causes an increase of about 0.26 percentage points. The

**Table 2****Exposure to Supply Chain Disruptions and PPI Inflation: Lags, January–November 2021**

	Backlogs	Delivery times
Domestic exposure	0.00755* (0.004)	0.00029 (0.004)
Foreign exposure	-0.154 (0.094)	0.0630 (0.087)
Domestic exposure ( $t - 1$ )	0.00340 (0.003)	0.00069 (0.004)
Foreign exposure ( $t - 1$ )	0.260*** (0.078)	0.205* (0.118)
Constant	-0.628*** (0.145)	2.356*** (0.664)
$N$	260	150
$R^2$	0.824	0.819

NOTE: Standard errors are in parentheses. \*  $p < 0.05$ , \*\*\*  $p < 0.001$ .

SOURCE: Authors' calculations.

**Table 3****Exposure to Supply Chain Disruptions and PPI Inflation: Manufacturing vs. Non-manufacturing, January–November 2021**

	Manufacturing	Non-manufacturing
Domestic exposure	-0.00015 (0.005)	0.0161*** (0.003)
Foreign exposure	0.309** (0.099)	0.397** (0.122)
Constant	-3.549*** (0.978)	-1.355*** (0.172)
Industry FE	YES	YES
$N$	165	121
$R^2$	0.747	0.574

NOTE: FE, fixed effects. Standard errors are in parentheses. \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

$R$ -squared is about 75 percent. Exposure to domestic bottlenecks, either measured as backlogs or delivery times, has no statistically significant effect on an industry's PPI inflation. This result may be capturing the high costs of restructuring a global supply chain that relies heavily on foreign suppliers. High fixed costs of setting up global supply chains could be resulting in stronger downstream production effects. For example, an industry heavily dependent on imported intermediates may not be able to efficiently identify new sources for those inputs.

Since supply chain disruptions may have a delayed effect on PPI inflation, we conduct the same regression in equation (3) but lag the domestic and foreign exposure measures. The results

are reported in Table 2. Foreign supply chain disruptions get propagated to domestic PPI inflation with a one-month lag. This result is robust to the use of both backlogs and delivery times as the measure of supply chain disruptions.<sup>13</sup> Hence, supply chain disruptions tend to have a delayed impact on PPI inflation.

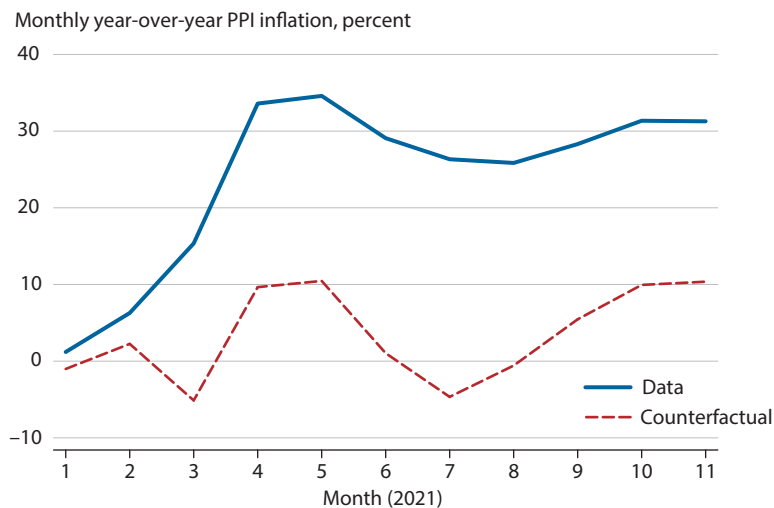
Table 3 reports the results with the industries separated into manufacturing and non-manufacturing sectors. Data on delivery times are only available for manufacturing sectors. Hence, we focus on backlogs as a measure of supply chain disruptions in Table 3. Notably, domestic exposure is only statistically significant when restricting the sample to the non-manufacturing sector. This result may reflect the fact that non-manufacturing sectors typically rely far less on FVA than manufacturing sectors do. Therefore, they are more susceptible to domestic fluctuations overall. In the manufacturing sector, both domestic and foreign bottlenecks have a positive effect on PPI inflation, but the effect is only significant for foreign bottlenecks. The *R*-squared is about 75 percent in the manufacturing sector and 58 percent in the non-manufacturing sector.

Our results suggest that global supply chain disruptions, which reflect a mismatch between demand and supply shocks, can propagate to domestic PPI inflation. The propagation is larger in those sectors where GVCs are more important.

**3.1.1 Back-of-the-Envelope Calculation.** We now ask the following question: What would PPI inflation have been during 2021 if bottlenecks in each country had followed their 2019 path? To answer this question, we compute for each U.S. industry in the manufacturing sector a counterfactual PPI inflation rate that uses country-level data on supply chain disruptions from January to November 2019. The focus is on the manufacturing sector, which has experienced worse supply chain disruptions due to its higher dependence on GVCs.

We proceed in several steps. First, we recalculate our measures of domestic and foreign exposure in equations (2) and (1), using country-level backlogs for each month of 2019. These counterfactual measures capture the exposure of each U.S. manufacturing industry through GVC participation if backlogs had followed the same paths of monthly changes as in 2019. Second, we substitute these measures into equation (3), using the estimated coefficients and industry fixed effects from the first column of Table 3. The result is a counterfactual measure of PPI inflation for each manufacturing industry from January to November 2021. Third, we compute aggregate manufacturing PPI inflation, both in the data and in the counterfactual, as a weighted average across industries' PPI inflation in each month of 2021. The weights are provided by the BLS for December 2020.<sup>14</sup>

Figure 7 plots the evolution of year-over-year monthly manufacturing PPI inflation from January to November 2021, both in the data (solid line) and in the counterfactual (dashed line). PPI inflation is always lower in the counterfactual than in the data, suggesting that bottlenecks during 2021 significantly contributed to inflation. Differences between the data and the counterfactual were larger between June and September 2021 and then started narrowing in October 2021. In particular, we find that manufacturing PPI inflation would have been 2 percentage points lower in January 2021 and 20 percentage points lower in November 2021 if the monthly change in bottlenecks had followed the 2019 path.

**Figure 7****Manufacturing Year-Over-Year PPI Inflation in the Data and in the Counterfactual, January–November 2021**

SOURCE: BLS and authors' calculations.

**4 CONCLUSIONS**

In this article, we investigated the role of global supply chain disruptions in PPI inflation across U.S. industries during the COVID-19 pandemic. We find that exposure to foreign bottlenecks through GVCs played a significant role in transmitting the effects of supply chain disruptions to U.S. prices. Our findings are driven by a combination of demand and supply shocks and the heterogeneous exposure to these shocks across industries. Industries that rely on inputs from countries whose production has been most affected by disruptions also experienced large price increases due to the inability to keep up with demand. Whether the inflation caused by supply chain disruptions is temporary (i.e., a rise in the cost of living) or a more permanent phenomenon will depend—absent any policy intervention—on the ability of supply chain disruptions to ease in order to meet the higher demand. The unequal distribution of vaccines in emerging countries, the rise of new variants, and disruptions in shipping could add some additional pressure on supply chains, creating pessimism about inflation disappearing in the near future. ■

## NOTES

- <sup>1</sup> See Santacreu and LaBelle (2021a,b) for a discussion on the rise of GVCs.
- <sup>2</sup> See Leibovici and Dunn (2021).
- <sup>3</sup> We focus on PPI inflation rather than consumer price index inflation, as the channels explored in the article (bottlenecks and delivery times) are likely to have a more direct effect on producer prices. Increases in producer prices may then be passed onto consumers with a lag. As such, the PPI serves as a leading indicator for the consumer price index.
- <sup>4</sup> We follow the methodology developed in LaBelle, Leibovici, and Santacreu (2021).
- <sup>5</sup> The plot shows monthly real consumption expenditures by major product type as reported on BEA release Table 2.8.3 of durable goods, seasonally adjusted; the date of the business cycle peak for each of the four recessions is from the National Bureau of Economic Research (NBER) Business Cycle Dating Committee (<https://www.nber.org/research/business-cycle-dating>). Plotting total consumption (i.e., including both durables and non-durables consumption) shows a similar evolution as that of durables consumption, but the changes are less striking (see Barnes, Bauer, and Edelberg, 2021).
- <sup>6</sup> Data are from S&P Global, which surveys upper-level executives in different industries across the world. The questions asked focus on the level of different aspects of production compared with one-month ago. The results are summarized into a diffusion index: 50 indicates no change; values above (below) 50 signal an expansion (contraction).
- <sup>7</sup> We use the BLS PPI year-over-year change rate as our measure of PPI inflation. We then standardize it to bring the scale in line with the S&P Global data. In particular, we first de-mean and divide by its sample standard deviation; second, we multiply the resulting series by the standard deviation of the S&P Global backlogs measure and add the mean.
- <sup>8</sup> We do not have data on sectoral backlogs for the United States; instead we use S&P Global data that only report disaggregated data for a few countries. To the extent that these countries have a similar production structure as the United States, we can assume that cross-sector bottlenecks in the United States follow the same pattern.
- <sup>9</sup> A value that goes from 42 to 47, for example, does not mean that bottlenecks are increasing but rather that the rate of loosening of the supply chain is slowing down.
- <sup>10</sup> Data are from the OECD Trade in Value Added (TIVA) dataset, which reports the value-added content from each origin country in the production of U.S. goods and services that are consumed worldwide.
- <sup>11</sup> Note that the backlog and delivery time measures are at the country-period level, whereas the FVA measure is at the country-sector level.
- <sup>12</sup> The list of countries is Australia, Austria, Brazil, Canada, China, Colombia, the Czech Republic, France, Germany, Greece, Indonesia, India, Ireland, Italy, Japan, Kazakhstan, Korea, Malaysia, Mexico, Myanmar, the Netherlands, the Philippines, Poland, the Russian Federation, Spain, Switzerland, Taiwan, Thailand, Turkey, the United Kingdom, the United States, and Vietnam.
- <sup>13</sup> The results (not included here) are robust to including two-month lags.
- <sup>14</sup> See <https://www.bls.gov/ppi/#tables>.

## REFERENCES

- Amiti, Mary; Heise, Sebastian and Wang, Aidan. "High Import Prices along the Global Supply Chain Feed Through to U.S. Domestic Prices." Federal Reserve Bank of New York *Liberty Street Economics*, November 8, 2021; <https://libertystreeteconomics.newyorkfed.org/2021/11/high-import-prices-along-the-global-supply-chain-feed-through-to-u-s-domestic-prices/>.
- Antrás, Pol. "Conceptual Aspects of Global Value Chains." World Bank *Economic Review*, 2020, 34(3), pp. 551-74; <https://doi.org/10.1093/wber/lhaa006>.
- Barnes, Mitchell; Bauer, Lauren, and Edelberg, Wendy. "11 Facts on the Economic Recovery from the COVID-19 Pandemic." Hamilton Project *Economic Facts*, September 2021; [https://www.hamiltonproject.org/assets/files/COVID\\_Facts.pdf](https://www.hamiltonproject.org/assets/files/COVID_Facts.pdf).
- Çakmaklı, Cem; Demiralp, Selva; Kalemli-Özcan, Şebnem; Yeşiltaş, Sevcan and Yıldırım, Muhammed A. "The Economic Case for Global Vaccinations: An Epidemiological Model with International Production Networks." NBER Working Paper 28395, National Bureau of Economic Research, 2021; <https://doi.org/10.3386/w28395>.



- Comin, Diego A. and Johnson, Robert C. "Offshoring and Inflation." NBER Working Paper 27957, National Bureau of Economic Research, October 2020; <https://doi.org/10.3386/w27957>.
- Ha, Jongrim; Kose, M. Ayhan and Ohnsorge, Franziska. "Inflation During the Pandemic: What Happened? What Is Next?" 2021; <https://doi.org/10.2139/ssrn.3881502>.
- Leibovici, Fernando and Dunn, Jason. "Supply Chain Bottlenecks and Inflation: The Role of Semiconductors." Federal Reserve Bank of St. Louis *Economic Synopses*, 2021, No. 28; <https://doi.org/10.20955/es.2021.28>.
- Santacreu, Ana Maria; Leibovici, Fernando and Jesse, LaBelle. "Global Value Chains and U.S. Economic Activity During COVID-19." Federal Reserve Bank of St. Louis *Review*, Third Quarter 2021, 103(3), pp. 271-88; <https://doi.org/10.20955/r.103.271-88>.
- Santacreu, Ana Maria and LaBelle, Jesse. "Rethinking Global Value Chains During COVID-19: Part 1." Federal Reserve Bank of St. Louis *Economic Synopses*, 2021a, No. 16; <https://research.stlouisfed.org/publications/economic-synopses/2021/07/01/rethinking-global-value-chains-during-covid-19-part-1>.
- Santacreu, Ana Maria and Jesse LaBelle. "Rethinking Global Value Chains During COVID-19: Part 2." Federal Reserve Bank of St. Louis *Economic Synopses*, 2021b, No. 17; <https://doi.org/10.20955/es.2021.17>.

# Who Should Work from Home During a Pandemic? The Wage-Infection Trade-off

Sangmin Aum, Sang Yoon (Tim) Lee, and [Yongseok Shin](#)

Shutting down the workplace is an effective means of reducing contagion but can induce large economic losses. We harmonize the American Time Use Survey and O\*NET data to construct a measure of infection risk (exposure index) and a measure of the ease with which a job can be performed remotely (work-from-home index) across both industries and occupations. The two indexes are negatively correlated but distinct, so the economic costs of containing a pandemic can be minimized by sending home only those workers that are highly exposed to infection risk but that can perform their jobs easily from home. Compared with a lockdown of all non-essential jobs, which includes many jobs not easily performed from home, a more selective policy can attain the same reduction in aggregate infection risk (32 percent) with one-third fewer workers sent home to work (24 percent vs. 36 percent) and only half the aggregate wage loss (15 percent vs. 30 percent). In addition, moving to such a policy reduces the infection risk of low-wage workers the most and the wage losses of high-wage workers the most. Our crosswalk between the American Time Use Survey and O\*NET data can be applied to a broader set of topics. (JEL E24, I14, J21 )

Federal Reserve Bank of St. Louis *Review*, Second Quarter 2022, 104(2), pp. 92-109.

<https://doi.org/10.20955/r.104.92-109>

## 1 INTRODUCTION

Most countries have implemented lockdowns and social distancing to varying degrees to contain the COVID-19 pandemic. The obvious downside is the economic costs, since most economic activities depend on in-person interaction. Thus, at least in the short run, policymakers face an inherent trade-off between the risk of contagion and economic losses.

To analyze this trade-off, it is important to know, first, the exposure-to-infection risk from performing a given job and, second, the ease with which the job can be performed remotely. The actual trade-off will depend on how jobs are distributed along these two dimensions, which is the focus of this article.

Sangmin Aum is an assistant professor of economics at Kyung Hee University. Sang Yoon (Tim) Lee is a reader at Queen Mary University of London and an affiliate at the Center for Economic and Policy Research. Yongseok Shin is a professor of economics at Washington University in St. Louis, a research fellow at the Federal Reserve Bank of St. Louis, and a faculty research fellow at the National Bureau of Economic Research. We are grateful for many helpful suggestions from an anonymous referee. Sang Yoon (Tim) Lee gratefully acknowledges the financial support from the British Academy (grant number COV19\201483).

© 2022, Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

We begin by constructing an index of exposure-to-infection risk across occupations by using O\*NET data and an index of how easily a job can be performed remotely (from home) across industries and occupations by using the “time worked from home” variable in the American Time Use Survey (ATUS). Such indexes are not new, but our approach is novel in that we harmonize the two datasets through our own crosswalk to quantify the trade-off between infection risk and the economic losses present in the economy, using the distribution of workers across jobs in the American Community Survey (ACS).

We find that, although jobs not easily performed from home (low work-from-home [WFH] ability) tend to be more exposed to infection risk on average, the negative correlation between the two indexes across occupations and industries is far from tight. First, there are a number of jobs with high exposure that can be easily performed from home (high WFH ability), such as IT sales agent. Second, infection risk varies widely even among jobs with the same WFH ability: For example, neither medical therapists nor experimental physicists can work from home, but the latter’s workplaces pose almost no risk of contagion. In addition, even the same occupation can have a very different WFH ability depending on the industry: For example, a registered nurse employed by a hospital has low WFH ability, but one in consulting services has high WFH ability.

In light of these findings, we consider an optimal policy that selectively sends home specific occupations in specific industries to minimize the aggregate wage loss subject to a given reduction in the aggregate exposure-to-infection risk. Intuitively, it is optimal to first send home workers with jobs with high exposure at work and small productivity and wage losses when working from home, the latter of which can be computed from a job’s wage and WFH ability. Mathematically, this translates into a linear threshold in the two-dimensional plane of exposure and wage loss.

The aggregate wage loss under the optimal policy is much smaller than under a lockdown of all non-essential jobs as implemented in many U.S. states and European economies. Our version of the real-world lockdown reduces aggregate exposure by 32 percent by sending home 36 percent of all workers, costing 30 percent of aggregate wages. Our optimal policy attains the same reduction in aggregate exposure by sending home only 24 percent of all workers, costing only 15 percent of aggregate wages. That is, the optimal policy achieves the same reduction in aggregate infection risk for half the economic cost. Under a constrained optimal policy in which healthcare-related workers must continue to work normally, the aggregate wage loss is 20 percent, still a third smaller than under a real-world lockdown. These gains are possible because the optimal policy exploits the large variation in WFH ability across occupations and industries for any given level of exposure—the novel fact we establish in this article.

It has become clear that low-wage workers are not only bearing the brunt of the pandemic economically but are also bearing the brunt of the infection risk. Compared with a lockdown of all non-essential jobs, the optimal policy reduces low-wage workers’ infection risk the most. On the other hand, a move from such a lockdown to the optimal policy generates the largest wage gains for high-wage workers, although more workers across the entire wage distribution are allowed to work normally and thus earn more than when working from home. In fact, under all policy scenarios we consider, high-wage workers are the least exposed to infection risk and also lose the least economically, pointing to the importance of redistributive policies during a pandemic.

A word of caution: By design, our optimal policy is simple and static, abstracting from the essentiality of certain jobs (other than healthcare-related workers) that need to be performed even in the midst of a pandemic; the complementarity among jobs that need to be performed in person;

the economic propagation across jobs and sectors; and the possibility that people switch to jobs with lower exposure or higher WFH ability (to help prevent wage loss). It also assumes that the indexes we construct are constant, ignoring the potential change in exposure or WFH ability of specific jobs due to more subtle non-pharmaceutical interventions such as wearing masks, reorganizing the workplace, and changing individual behaviors as the population adjusts to the pandemic. However, it does capture the direct trade-offs given the work patterns at the onset of the pandemic. Furthermore, our analysis presents simple guidance for policymakers that is easy to implement in practice, while also providing a benchmark for structural economic models that consider some of the dimensions that we abstract from.<sup>1</sup>

It should also be noted that our industry-occupation crosswalk between ATUS and ACS/O\*NET is not specific to exposure and WFH ability and can be applied to a broader set of topics.

## 2 DATA

### 2.1 *Employment and Wages by Occupation and Industry*

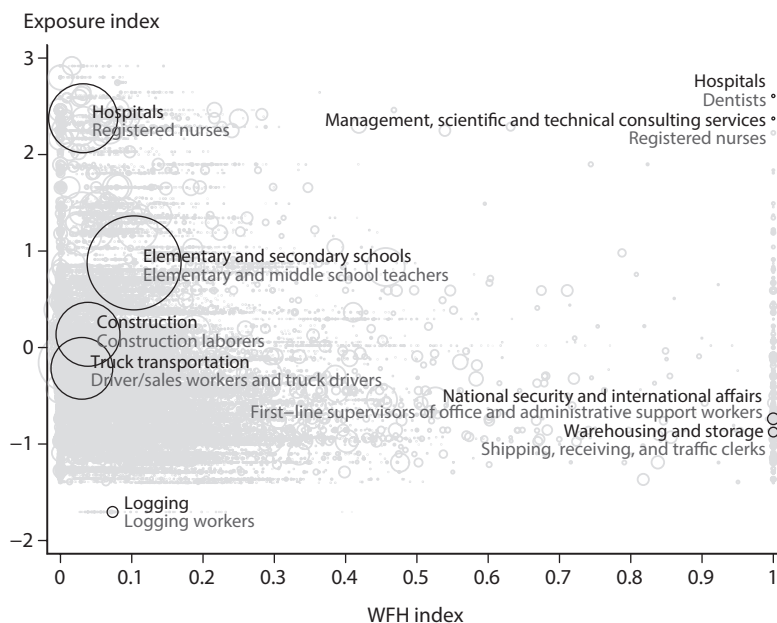
We compute employment weights and mean hourly wages by occupation and industry from the ACS. We include only civilian prime-age workers (between 16 and 65 years old). An individual's employment weight is their sampling weight multiplied by their usual annual hours worked (usual hours worked in a week times usual weeks worked in a year). For each year, we multiply top-coded wages by 1.5 and bottom-code the lowest hourly wage percentiles. Then, the hours-adjusted employment weight and mean hourly wage of each occupation-industry combination are computed using consistent industry and occupation codes following Autor and Dorn (2013), modified to incorporate changes in the Census industry and occupation codes from 2014 to 2018 as in Lee and Shin (2017). Finally, we take a simple average of the employment weights and hourly wages over the five years.

### 2.2 *Constructing the Exposure and WFH Indexes*

**Exposure Index.** The O\*NET asks experts and workers to give numerical answers to questions that capture detailed characteristics of an occupation, where an occupation is defined by its Standard Occupation Classification (SOC) code. To construct our exposure index, we take the weighted average of the answers to two questions: one about physical proximity (PP) to other people in the workplace and the other about potential exposure to disease or infection (EDI) in the workplace.<sup>2</sup>

We first convert O\*NET titles to SOC codes using the accompanying crosswalk.<sup>3</sup> The SOC codes are then mapped to ACS OCC (occupation) codes using a crosswalk from IPUMS USA, which is heavily modified so that each ACS OCC code has a unique value for both descriptors.<sup>4</sup> Finally, we normalize PP and EDI to have a mean of zero and standard deviation of 1, and take the average of the two as our exposure index value.<sup>5</sup>

**WFH Index.** Some earlier studies on COVID-19 used O\*NET descriptors to construct a WFH index (e.g., Dingel and Neiman, 2020, and Mongey, Pilossoph, and Weinberg, 2020). However, a more accurate measure of the ease with which a job can be performed from home may be whether people actually do the job from home, such as reported in the ATUS. While Mongey, Pilossoph, and Weinberg (2020) do show that the former is positively correlated with the latter, there are both qualitative and quantitative reasons to favor the latter. Qualitatively, some descriptors included in O\*NET-based WFH indexes are misleading: For example, O\*NET's "outdoor" categories (implying

**Figure 1****Relationship Between Exposure and WFH Ability**

NOTE: Each circle represents a specific occupation within a specific industry (458 occupations  $\times$  254 industries). The size of the circle denotes the hours-adjusted employment share, averaged from 2014 to 2018. Black circles indicate examples of occupations either with very large employment shares or with extreme values of exposure and/or WFH ability. The first description for these examples is the industry, and the second is the occupation.

low WFH ability) include farmers who in fact have a high incidence of actually working from home, since they work in self-owned plots and land.<sup>6</sup> Quantitatively, the ATUS allows WFH ability to vary across industries, as well as across occupations, and indeed even the same occupation has very different WFH ability across industries in the data.<sup>7</sup>

For each industry-occupation combination in the ATUS, we compute the total time spent working and total time spent working from home across all individuals in the corresponding cell for each year from 2014 to 2018, using each year's sampling weight. We then take a simple average over all five years, and the ratio of the average time worked from home to the average time spent working is our WFH index value. Each cell is then matched to the ACS; in the process, we merge and impute missing cells using ACS employment weights and also ensure that each OCC code has at least one O\*NET descriptor.

All in all, we end up with 458 occupations across 254 industries (458  $\times$  254), each with an exposure index value, a WFH index value, hours worked, and hourly wages.

### 2.3 The Relationship Between Exposure and WFH Ability

Are jobs with higher infection risk harder to perform from home? In Figure 1, each circle represents a job, defined as a specific occupation in a specific industry, and its location on the WFH index (horizontal axis) and exposure index (vertical axis). The size of the circle denotes the (hours-weighted) employment share of that job, averaged from 2014 to 2018.

**Table 1****WFH Index**

	Industry-occupation pairs	By occupation	By industry
Exposure	−0.033 (0.004)	−0.033 (0.005)	−0.045 (0.010)
Observations	54,108	458	254
$R^2$	0.034	0.109	0.126

NOTE: The table shows the result of regressing WFH ability on exposure, with observations weighted by their hours-adjusted employment shares. Robust standard errors are in parentheses. The units of observation are an industry-occupation pair (first column), occupation (second column), and industry (last column).

Three patterns emerge. First, consistent with conventional wisdom, jobs with higher infection risk tend not to be performed from home. For example, hospital nurses have high infection risk and rarely work from home. Second, the negative correlation is not very tight. In Figure 1, there are many industry-occupation pairs that do not work from home, regardless of their exposure-to-infection risk. Most notably, loggers have one of the least exposed occupations, but for obvious reasons they do not work from home. Third, even the same occupation shows substantial variations in WFH ability across industries. For example, while hospital nurses do not work from home, a small number of registered nurses in the consulting services industry exclusively work from home.

Table 1 shows the coefficients from regressing the WFH index on the exposure index for our whole industry-occupation sample (first column), only by occupation (second column), and only by industry (third column). For the latter cases, the exposure and WFH indexes are computed by averaging across each occupation's exposure and WFH ability within an industry, respectively, weighted by each occupation's within-industry employment share. In all three cases, the negative correlation between the exposure index and the WFH index is statistically significant. However, a large fraction of the dispersion in one index still remains unexplained by the other, as represented by the low  $R^2$ 's in all regressions. In particular, the  $R^2$  is as low as 0.034 for the full sample of industry-occupation pairs. Since the exposure index varies only across occupations and not industries, by construction, the large increase in  $R^2$  from the first to the second column mirrors the wide variation in WFH ability across industries—even for the same occupation.<sup>8</sup>

### 3 OPTIMAL POLICY

We interpret the fact that a job is more likely to be worked from home (high WFH ability) to mean that it will have little productivity loss when worked from home. Thus, the large dispersion in WFH ability among jobs with similar levels of exposure implies that the economic cost of sending home workers to reduce infection risk depends on whether their jobs have high or low WFH ability.

Let  $(s_i, w_i, e_i, h_i)$  denote the employment share, average wage, exposure index, and WFH index for each industry-occupation pair  $i$ . The optimal policy minimizes the economic cost from sending home a fraction  $0 \leq x_i \leq 1$  of each industry-occupation pair, subject to reducing aggregate exposure by at least a given fraction  $0 \leq y \leq \psi$ :

$$(1) \quad \min_{\{x_i \in [0,1]\}_{i=1}^I} \psi \sum_{i=1}^I s_i (1-h_i) w_i x_i \quad \text{s.t.} \quad \psi \sum_{i=1}^I s_i e_i x_i \geq y \sum_{i=1}^I s_i e_i,$$

where  $\psi$  is a “compliance” constant that measures how many workers in the jobs sent home comply with the policy—that is, they work from home—and  $I$  is the total number of occupation-industry pairs.<sup>9</sup>

Designing the optimal policy is merely a high-dimensional linear programming problem.<sup>10</sup> The  $h_i$  in the objective function is our WFH index value that lies between 0 and 1. For the problem to be well defined, all  $e_i$ 's need to be nonnegative, so we simply shift our exposure index by subtracting off its minimum value. Since the problem is linear, the solution  $x_i^*$  will be either 0 or 1 for all  $i$ , unless the job happens to fall exactly on the threshold.

The objective function is meant to capture the *aggregate* economic loss from sending home workers. As a result, the economic loss is larger not only when workers low on the WFH index are sent home, but also when industry-occupation pairs with a high wage and a large number of workers are sent home.<sup>11</sup> This finding does not necessarily mean that we are not interested in the distributive consequences of such policies. In fact, we discuss the unequal impact across workers in Section 4. We are merely separating the question of efficiency (i.e., minimizing the aggregate economic loss) from the question of redistribution, although we do not consider any redistributive policies in this article.

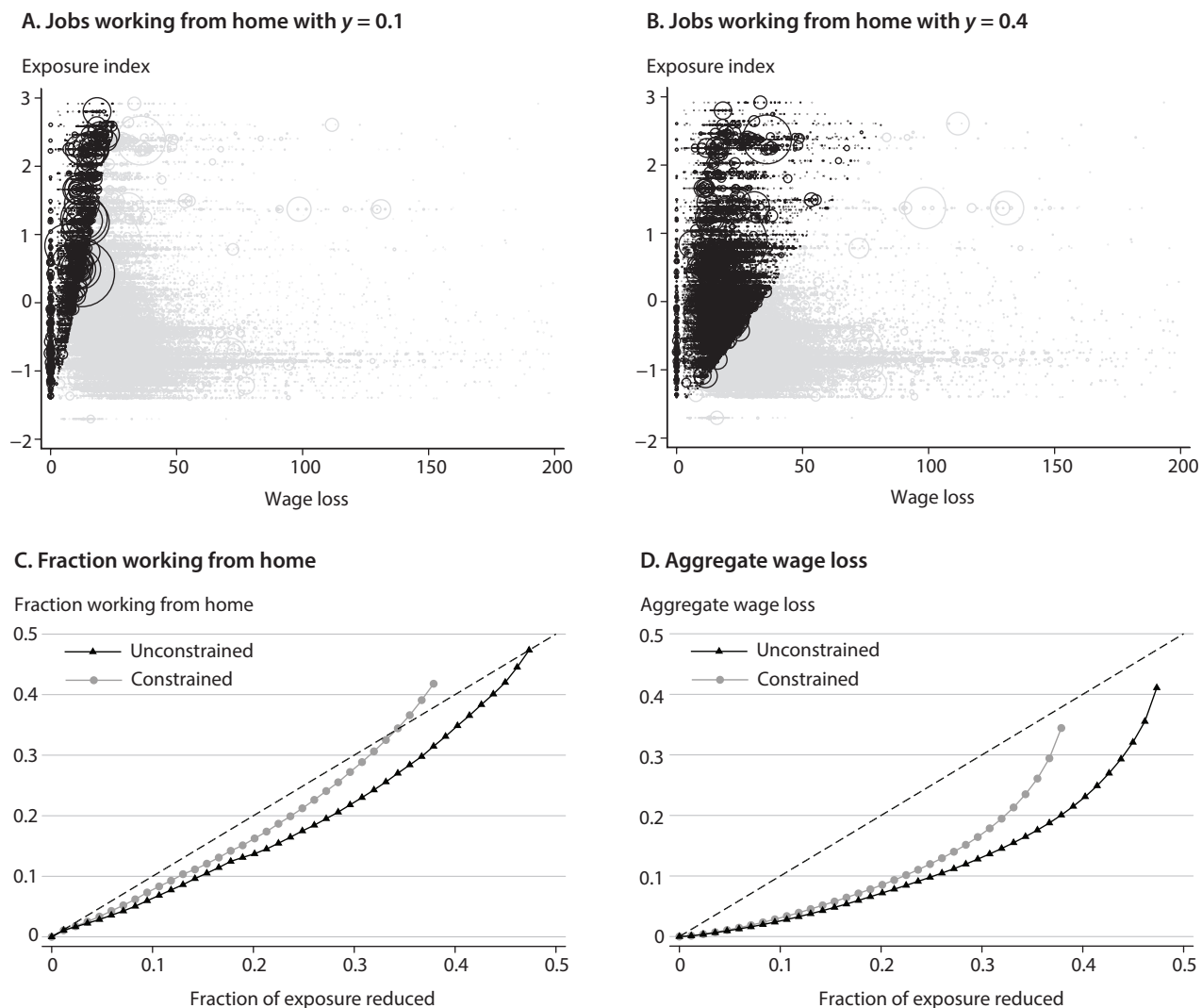
Figure 2 shows the optimal policy when  $\psi = 0.473$ .<sup>12</sup> In Panels A and B, each industry-occupation pair is plotted as a circle along the exposure ( $e_i$ , vertical axis) and the wage losses from sending home workers ( $w_i(1-h_i)$ , horizontal axis) dimensions. The size of a circle represents job  $i$ 's employment share. The optimal policy is a linear threshold in this two-dimensional space, where only workers in jobs that are above the threshold are sent home (black circles). The slope of the threshold is positive since the optimal policy takes into account both exposure and wage losses from sending home workers. That is, even if a job has high infection risk, it is not sent home if it cannot be done remotely (low WFH ability) and has a high average wage, implying large wage losses from working remotely. Panel A is the solution when aggregate exposure is reduced by 10 percent and Panel B by 40 percent.

Panel C plots the fraction of workers sent home under the optimal policy (black lines with triangles) for each level of reduction in aggregate exposure ( $y$ ) on the horizontal axis. Panel D shows the aggregate wage loss from the optimal policy (black lines with triangles against the same horizontal axis). Both black lines with triangles are upward sloping, since reducing exposure requires sending more workers home, which also leads to larger wage losses. More important, the lines are convex, because the optimal policy sends home the jobs with highest exposure and lowest wage losses first.

One unpleasant feature of the optimal policy we solve above is that many healthcare-related workers are sent home because they tend to have very high exposure, even though most of them cannot work from home and thus incur large wage losses. For a more realistic problem during a pandemic, we also solve a constrained optimal policy in which healthcare-related workers must work normally. These are workers in the healthcare and social assistance sectors (industry codes 7970-8390) and healthcare practitioners, technical, and support occupations (OCC codes 3000-3655), who together comprise about 11 percent of (hours-adjusted) aggregate employment. Combined, these jobs account for about 20 percent of aggregate exposure.

**Figure 2**

**Optimal Policy**



NOTE: In Panels A and B, each circle represents a specific occupation within a specific industry (458 occupations  $\times$  254 industries). The size of the circle denotes the hours-adjusted employment share, averaged from 2014 to 2018. The black circles represent jobs sent home by the optimal policy (group  $i$  such that  $x_i^* = 1$ ), and the gray circles indicate jobs performed normally. Panel A is when the aggregate exposure is reduced by 10 percent and Panel B by 40 percent. In Panels C and D, for a fraction  $y$  of aggregate exposure reduction on the horizontal axis, the fraction of workers sent home and the aggregate wage loss (with the scaling constant  $\psi = 0.473$  under the optimal policy) are shown as black lines with triangles in Panels C and D, respectively. The constrained optimal policy solves the same problem (1), but with the added constraint that all healthcare-related workers must work normally, as shown by the gray lines with circles.

In Panels C and D of Figure 2, the gray lines with circles plot the fractions of workers sent home under the constrained optimal policy in Panel C and the aggregate wage loss in Panel D, both against the reduction in aggregate exposure  $y$  on the horizontal axis. Compared with the unconstrained optimal policy, the constrained policy sends more workers home and causes higher wage losses for any given reduction in aggregate exposure, but it is still visibly convex.



The analysis is based on the pre-pandemic employment structure and wages of the United States as well as our exposure and WFH indexes for the United States. To the extent that the employment and wage distributions across occupations and industries differ across countries, the trade-off between infection and economic losses will differ across countries, even if countries share the same exposure and WFH indexes. In fact, trade-off patterns will also vary across regions or states within the United States.

## 4 OPTIMAL POLICY VS. REAL-WORLD LOCKDOWNS

We now compare our optimal policy with a lockdown that mimics those implemented in many U.S. states and most European countries. Although lockdowns were implemented with varying degrees of severity across countries and U.S. states, they did share common features. Most often, governments classified industries as essential or non-essential and tried to keep essential workers working as normally as possible while also forbidding non-essential workers from commuting to work. For example, the Cybersecurity and Infrastructure Security Agency of the U.S. Department of Homeland Security provides guidelines on which jobs are essential for critical infrastructure.

Our version of the real-world lockdown follows Palomino, Rodríguez, and Sebastian (2020), who show which occupations and industries were effectively locked down in Europe.<sup>13</sup> In the context of our problem (1), jobs sent home in a real-world lockdown are a set  $\{i \in \{1, 2, \dots, I\} | x_i = 1\}$ , which in general will differ from the optimal solution. We set  $\psi = 0.473$ , as we did in the previous section, which generates a 30 percent drop in aggregate wages under the actual lockdown.<sup>14</sup> Given  $\psi$ , we find that a real-world lockdown reduces aggregate exposure by 32 percent.

Panel A of Figure 3 shows the non-essential jobs that sent workers home under the actual lockdown (dark-gray circles) and the jobs that sent workers home under the optimal policy (black circles). Both policies achieve the same reduction in aggregate exposure, but there is not much overlap between the two policies in terms of which jobs send workers home. In particular, the actual lockdown sent home many workers in jobs that have relatively low exposure. Panel B shows the jobs that sent workers home under the constrained optimal policy (black circles) (i.e., all health-care-related workers working normally) still reduce aggregate exposure by 32 percent.

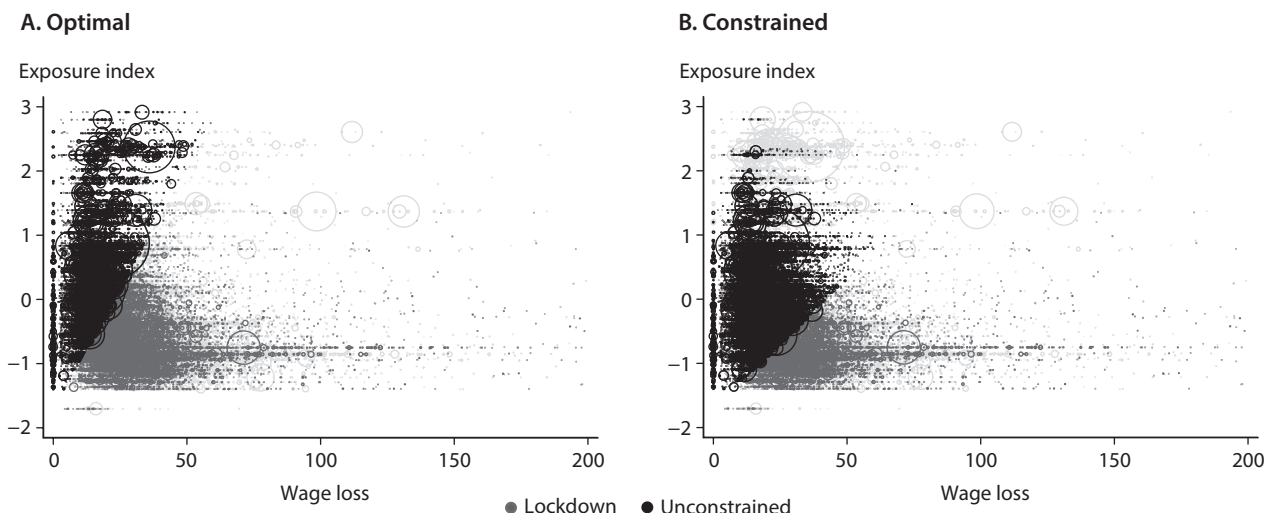
Table 2 shows exactly how this difference manifests in terms of the fractions of workers sent home and the aggregate wage loss. The gains from implementing the optimal policy are substantial: The same reduction in exposure can be attained by sending home one-third fewer workers (36 percent under the lockdown vs. 24 percent under the optimal policy) at half the economic cost (aggregate wage loss of 30 percent under the lockdown vs. 15 percent under the optimal policy).

Even with the constraint that all health-care-related workers continue to work normally, the constrained optimal policy does substantially better than the actual lockdown. It delivers the same reduction in aggregate exposure, with a one-third smaller loss in aggregate wages.<sup>15</sup>

**Distribution of Exposure Reduction and Wage Loss.** Because jobs are different in terms of exposure and WFH ability across occupations and industries, the reductions in exposure and in wage losses from the policies are distributed unequally across workers. In fact, low-wage jobs tend to have high exposure and low WFH ability, while high-wage jobs tend to have low exposure and high WFH ability. As a result, on average, low-wage workers see a large reduction in exposure and higher wage losses under a real-world lockdown, while high-wage workers see only a small reduction

**Figure 3**

**Optimal Policy vs. Lockdown**



NOTE: Each circle is a specific occupation within a specific industry (458 occupations × 254 industries). The size of a circle represents its employment share. Dark-gray circles indicate jobs sent home during the actual lockdown. Black circles are the jobs sent home under the unconstrained optimal policy in Panel A and under the constrained optimal policy in the Panel B.

**Table 2**

**Policy Outcomes**

	Fraction sent home	Wage loss
A. Lockdown	0.359	0.300
B. Optimal	0.244	0.146
C. Constrained optimal	0.308	0.196

NOTE: The first column is the (hours-adjusted) fraction of workers sent home, and the second is the aggregate wage loss. In all three cases, aggregate exposure is reduced by 32 percent.

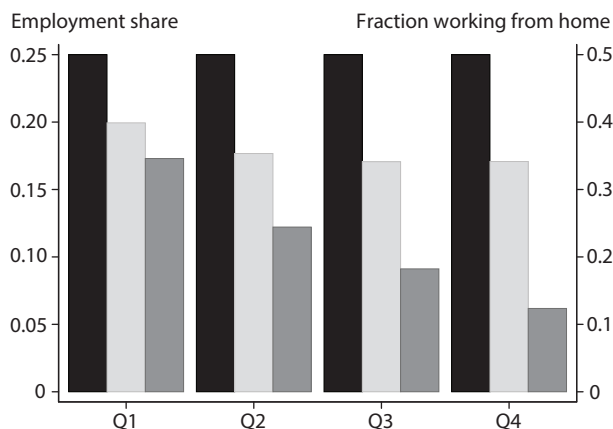
in exposure and lower wage losses.<sup>16</sup> Furthermore, since the real-world lockdown and the optimal policy target different jobs, the two policies have different distributional, as well as aggregate, consequences.

Figure 4 shows the distributional impacts of the actual lockdown and the optimal policy, across wage quartiles constructed from the average wage across the occupation-industry pairs.<sup>17</sup> In Panel A, the black bars are the shares of workers in each quartile, each 25 percent by definition. The other two bars within each quartile show the fractions of workers sent home under the actual lockdown (light-gray bars, right scale) and the optimal policy (dark-gray bars, right scale). The actual lockdown, which follows the essential/non-essential distinction, sends more low-wage workers home than high-wage workers (40 percent of the lowest wage quartile vs. 34 percent of the highest wage quartile). The difference is magnified under the optimal policy (37 percent of the lowest quartile vs.

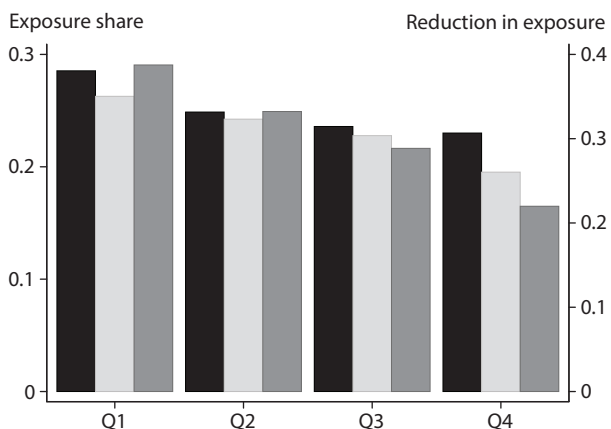
**Figure 4**

**Policy Impact Across Wage Quintiles**

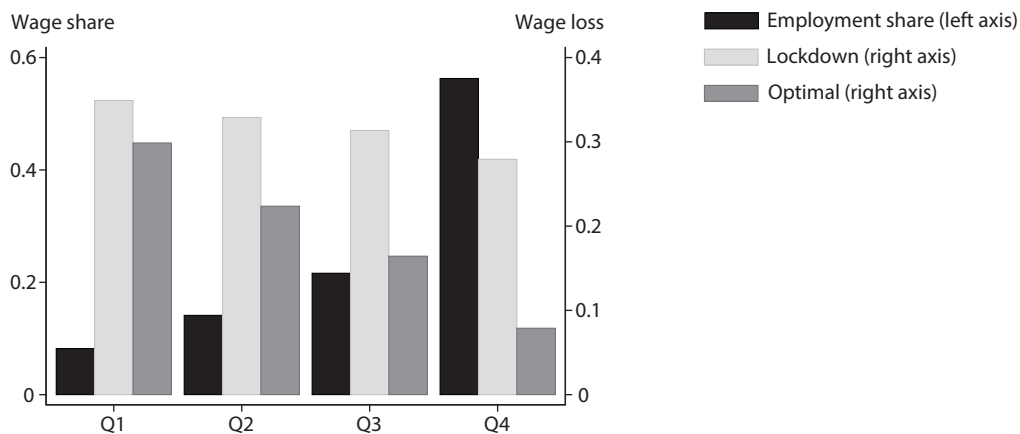
**A. Fraction working from home**



**B. Reduction in exposure**



**C. Wage loss**



NOTE: Each industry-occupation pair is ordered by its average wage and assigned to a quartile. In Panel A, the black bars are each wage quartile’s employment share, which is 0.25 by definition. The light-gray and the dark-gray bars depict the fraction of each quartile sent home due to the actual lockdown and the optimal policy, respectively, on the right scale. In Panel B, the black bars are each wage quartile’s share of aggregate exposure, and the light-gray and the dark-gray bars depict the within-quartile reduction in exposure under the actual lockdown and the optimal policy, respectively. In Panel C, the black bars are each wage quartile’s share of the aggregate wage, and the light-gray and the dark-gray bars depict within-quartile wage losses under each policy.

13 percent of the highest quartile). There are two reasons for this difference. First, high-wage workers tend to have low exposure and are hence less likely to be sent home. Second, because of their high wages, holding other things equal, it is more costly to send home high-wage workers. Nevertheless, the optimal policy sends fewer workers home in all quartiles than the actual lockdown does.

In Panel B, the black bars are each wage quartile’s share of aggregate exposure, confirming that low-wage workers have higher exposure on average. The other two bars show the reduction in exposure due to the actual lockdown (light-gray bars, right scale) and the optimal policy (dark-gray bars, right scale). Note that both policies reduce exposure more for low-wage workers than high-wage

workers. The optimal policy strengthens this pattern, reducing exposure more for low-wage than high-wage workers than the actual lockdown does.<sup>18</sup> More significant, for the lower wage quartiles, the optimal policy achieves a larger reduction in exposure while sending fewer workers home than the actual lockdown does.

In Panel C, the black bars are each wage quartile's share of aggregate labor income, which is higher for higher quartiles, by construction. The other two bars are the wage losses due to the actual lockdown (light-gray bars, right scale) and the optimal policy (dark-gray bars, right scale). Both policies incur larger wage losses for the lower wage quartiles. This is because low-wage jobs are more likely to send workers home under both policies but tend to be harder to perform from home (low WFH ability). The optimal policy leads to especially small wage losses for the top quartile, although it generates smaller wage losses than the actual lockdown does across all wage quartiles, as it sends home fewer workers across the board and also takes into account wage losses.

In summary, low-wage workers are more affected than high-wage workers by both policies: They are more likely to be sent home. As a result, they experience a larger reduction in exposure but also larger wage losses. Comparing the two policies, holding the reduction in aggregate exposure constant, low-wage workers see a larger reduction in exposure under the optimal policy, while high-wage workers see a smaller reduction. However, although the optimal policy results in smaller wage losses for all wage quartiles, high-wage workers lose the least relative to the actual lockdown.

## 5 CONCLUSION

We construct exposure and WFH indexes that vary by both occupation and industry and study their relationship across jobs. WFH ability varies widely even among jobs with similar levels of exposure, indicating that a planner could reduce the economic cost of a workplace lockdown by selectively sending home groups of workers based on the two indexes, rather than using broad essential/non-essential categories. Compared with the actual lockdown, the optimal policy sends home one-third fewer workers and causes only half the losses in aggregate wages while also reducing aggregate exposure by the same magnitude. In addition, under the optimal policy, high-wage workers have the smallest wage losses but low-wage workers have the largest reduction in exposure.

While we abstract from some key dimensions, our work is a blueprint for an easily implementable smarter lockdown of the workplace during a pandemic. In addition, our crosswalk and merging of the ACS/O\*NET and the ATUS can be applied to a wider range of research—beyond the exposure and WFH indexes used in this article. ■

## APPENDIX A: RELATED LITERATURE

A new strand of literature measures the degree to which jobs can be performed from home or are contact intensive. One of the earlier articles is by Dingel and Neiman (2020), who use job characteristics in the O\*NET data to determine which occupations can work from home. Koren and Petó (2020), Hicks, Faulk, and Devaraj (2020), and Leibovici, Santacreau, and Familglietti (2020) use O\*NET data to compute contact intensity. Only a few studies consider differences in such job characteristics across both industries and occupations. Adams-Prassl et al. (2020) collect information on WFH ability from geographically representative U.S. and British surveys and demonstrate large variation in WFH ability across industries even for the same occupation.

Most articles consider only either exposure or WFH ability. One exception is by Mongey, Pilossoph, and Weinberg (2020), who measure both PP (one component of our exposure index) and WFH ability and show that there is a negative correlation between the two. Their WFH index is based on occupational characteristics from O\*NET alone and does not vary by industry, and they find a tighter correlation between PP and WFH ability than what we find using WFH ability based on the ATUS, even when we ignore the industry dimension. Adams-Prassl et al. (2020) also document a negative relationship between their WFH index and the O\*NET PP data.

Few studies explicitly consider the costs of real-world lockdowns. Adams-Prassl et al. (2020) construct an occupation-level remote labor index using O\*NET data and combine it with industry-wide lockdown measures to assess the heterogeneous effect of industry-level supply shocks across occupations. Palomino, Rodríguez, and Sebastian (2020) construct a lockdown working ability (LWA) index by combining a telework index from O\*NET and lockdown measures based on government policies in Italy and Spain. They then simulate the impact of social distancing policies on inequality across European countries. Gottlieb et al. (2020) simulate the economic costs of various lockdown policies in developing countries, exploiting detailed data on each country's demographic and labor market composition. Aum, Lee, and Shin (2021a) do the same for Korea and the United Kingdom but also model the individual choice of voluntarily working from home out of fear of infection.

To the best of our knowledge, we are the first to analyze both WFH ability and exposure by industry and occupation and to go beyond merely documenting their correlation, by using indexes to determine which jobs to send home to minimize the aggregate wage loss for a given reduction in exposure.<sup>19</sup>

## APPENDIX B: ALTERNATIVE POLICIES

### *B1 Exposure Reduction or Wage Loss*

The optimal policy minimizes the aggregate wage loss subject to a given reduction in exposure. We consider alternative programs that minimize exposure or wage losses subject to a given fraction of workers sent home.

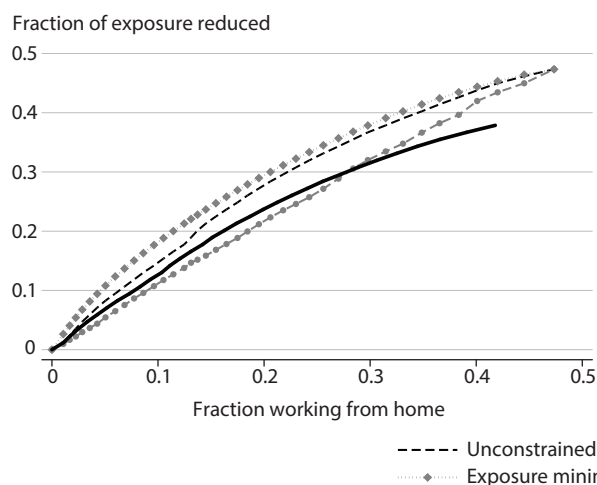
The program that minimizes the aggregate exposure is

$$(B1) \quad \max_{\{x_i \in \{0,1\}\}_{i=1}^I} \psi \sum_{i=1}^I s_i e_i x_i \quad \text{s.t.} \quad \psi \sum_{i=1}^I s_i x_i \leq z,$$

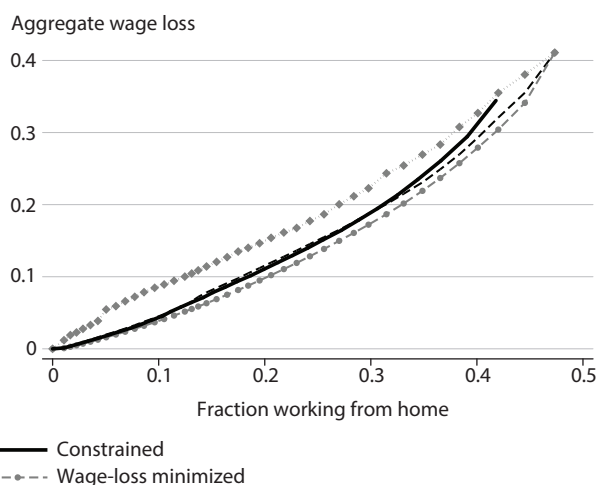
## Figure B1

### Optimal Policy vs. Alternatives

#### A. Exposure reduction



#### B. Wage loss



NOTE: In Panel A, for a given fraction of workers sent home on the horizontal axis, the reduction in aggregate exposure is plotted for the optimal policy (dashed line), the constrained optimal policy (solid line), the policy that minimizes aggregate exposure (line with diamonds), and the policy that minimizes the aggregate wage loss (line with circles). Panel B plots the corresponding aggregate wage loss.

which is in fact a maximization of the reduction in exposure subject to a given fraction of workers sent home. Clearly, the solution is to first send home workers in jobs with the highest exposure, regardless of wage losses. The other program that minimizes the aggregate wage loss subject to a given fraction of workers sent home is

$$(B2) \quad \min_{\{x_i \in [0,1]\}_{i=1}^I} \psi \sum_{i=1}^I s_i w_i (1 - h_i) x_i \quad \text{s.t.} \quad \psi \sum_{i=1}^I s_i x_i \geq z,$$

with the result that only workers with jobs that pay a low wage and/or are easy to perform from home are sent home.

Figure B1 plots the reduction in exposure and the aggregate wage loss from these two alternative policies for all levels of  $z$  (the fraction of workers sent home) with  $\psi = 0.473$ . The outcomes of the optimal policy and the constrained optimal policy are also plotted for comparison. As shown in Panel A, unsurprisingly, the exposure minimization policy (line with diamonds) reduces aggregate exposure by more than any other policy, but at the cost of higher aggregate wage loss, as shown in Panel B. Conversely, the wage-loss minimization policy (line with circles) has the smallest aggregate wage loss of all policies, but at the cost of a smaller reduction in exposure than the other policies. The outcomes of the optimal policy lie between those of the two alternative policies.

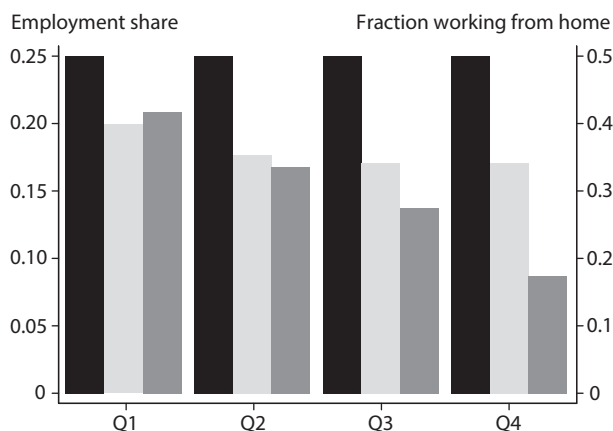
### B2 Impact of Constrained Optimal Policy Across Wage Quartiles

Finally, in Figure B2 we show the counterpart of Figure 4 for the constrained optimal policy that keeps all healthcare-related workers working normally.

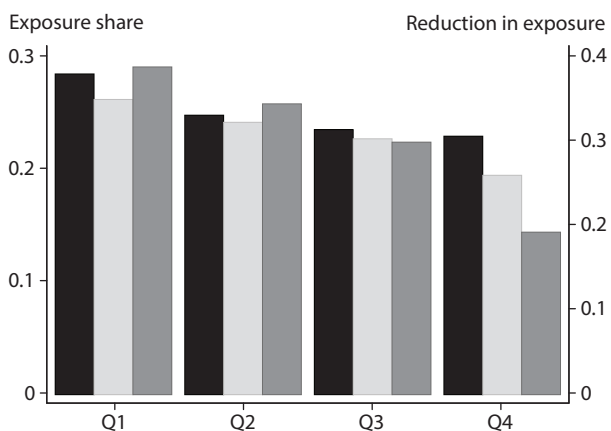
**Figure B2**

**Policy Impact Across Wage Quartiles**

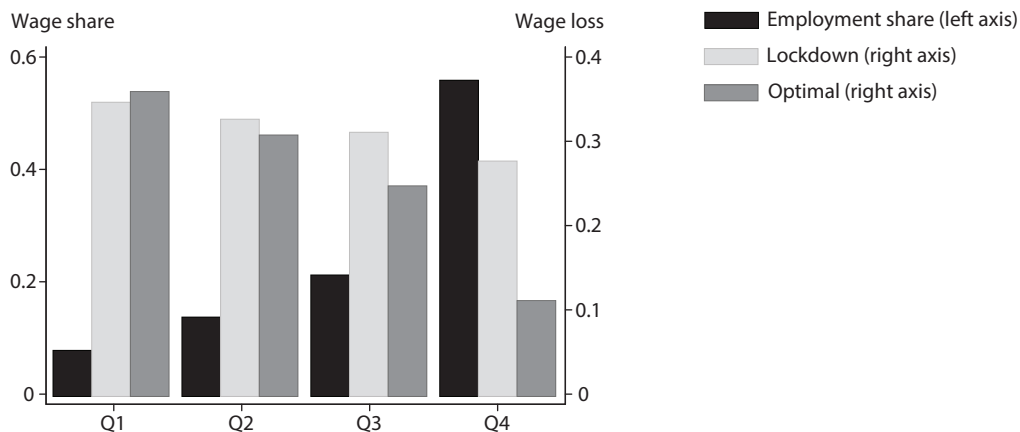
**A. Fraction working from home**



**B. Reduction in exposure**



**C. Wage loss**



NOTE: Each industry-occupation pair is ordered by its average wage and assigned to a quartile. In Panel A, the black bars are each wage quartile’s employment share, which is 0.25 by definition. The light-gray and the dark-gray bars depict the fraction of each quartile sent home due to the actual lockdown and the constrained optimal policy, respectively, on the right scale. In Panel B, the black bars are each wage quartile’s share of the aggregate exposure, and the light-gray and the dark-gray bars are the within-quartile reduction in exposure under the actual lockdown and the constrained optimal policy, respectively. In Panel C, the black bars are each wage quartile’s share of the aggregate wage, and the light-gray and the dark-gray bars depict within-quartile wage losses by each policy.

The patterns across wage quartiles are similar to those in Figure 4, but there is one important difference. Relative to the actual lockdown, the constrained optimal policy sends home more workers in the bottom-wage quartile, resulting in larger wage losses for this group.

**Table C1**  
**Regression Results**

	Occupation		Industry		Occupation-industry pair	
	Exposure	WFH ability	Exposure	WFH ability	Exposure	WFH ability
Females	1.24*** (0.43)	-0.02 (0.02)	2.08*** (0.45)	-0.07* (0.04)	1.11*** (0.22)	-0.00 (0.02)
≥ Bachelor's degree or more education	-0.88* (0.51)	0.12*** (0.04)	-1.45*** (0.51)	0.23*** (0.06)	-0.61* (0.33)	0.13*** (0.03)
Blacks	4.32*** (1.12)	-0.03 (0.08)	0.68 (0.76)	0.11 (0.11)	1.85*** (0.37)	-0.01 (0.04)
Hispanics	-1.48* (0.81)	-0.23*** (0.08)	-1.50** (0.74)	-0.19* (0.11)	-0.67** (0.28)	-0.14*** (0.04)
Young (under 30 years of age)	-2.71*** (0.99)	0.13 (0.08)	-4.71*** (1.08)	0.22 (0.15)	-0.64 (0.41)	0.05 (0.04)
Old (over 50 years of age)	-5.37*** (1.14)	0.15* (0.09)	-5.15*** (1.29)	0.13 (0.17)	-1.74*** (0.36)	0.03 (0.03)
Self-employed	0.81 (0.49)	0.29*** (0.07)	-0.64* (0.39)	0.40*** (0.07)	0.26 (0.24)	0.26*** (0.05)
Average hourly wage (in logs)	0.05 (0.57)	-0.00 (0.04)	-0.45 (0.42)	0.02 (0.06)	0.03 (0.29)	0.01 (0.03)
Constant	1.54 (2.12)	0.01 (0.15)	3.82** (1.63)	-0.12 (0.23)	0.17 (0.99)	0.02 (0.09)
R <sup>2</sup>	0.320	0.448	0.606	0.510	0.222	0.159
Observations	458		254		53,694	

NOTE: Robust standard errors are in parentheses. \*, \*\*, and \*\*\* indicate significance at the 90 percent, 95 percent, 99 percent levels, respectively.

## APPENDIX C: EXPOSURE AND WFH ABILITY BY DEMOGRAPHIC GROUP

How different demographic groups are affected by our index-based optimal policy depends on whether the variations in the indexes are captured by demographics. Thus, we regress the exposure index and the WFH index on demographic variables constructed for each occupation, industry, or industry-occupation pair: the female share, college share, Black and Hispanic shares, young (under 30 years of age) and old (over 50 years of age) shares, and the self-employment share. The excluded groups in the regression are males, those without a four-year college degree, Whites, the middle aged (between 30 and 50 years of age), and employees.

The results are shown in Table C1. Some groups of workers, in particular college graduates, are both less exposed to infection risk and less likely to be sent home. But as expected from Figure 1, low exposure does not necessarily mean high WFH ability. Female, Black, and middle-aged workers tend to be more exposed to infection risk, but it is not related to how easily they can work from home. Hispanics are less likely to work from home but also face relatively lower infection risk. In contrast, the self-employed are more likely to work from home but do not necessarily face lower infection risk.



Demographics are correlated with both the exposure index and the WFH index more by industry and less by occupation. And when the industry-occupation pairs are used, demographics barely predict either index. This finding implies that industry-based lockdowns are at risk of disadvantaging vulnerable demographic groups, while more sophisticated policies that take into consideration both industry and occupation could reduce the likelihood of disproportionately affecting those groups.

## NOTES

- <sup>1</sup> A review of the relevant literature is in the appendix.
- <sup>2</sup> These are questions 4.C.2.a.3 and 4.C.2.c.1.b, respectively. These two descriptors were analyzed in detail by the *New York Times* (Gamio, 2020) for the United States and reproduced for the United Kingdom by the U.K. Office for National Statistics. The *New York Times* and U.K. Office for National Statistics focused on each measure separately and showed that the two are strongly positively correlated. The economics literature has only used physical proximity as a measure of exposure (Leibovici, Santacreau, and Famiglietti, 2020, and Mongey, Pilossoph, and Weinberg, 2020).
- <sup>3</sup> Available at <https://www.onetcenter.org/crosswalks.html>.
- <sup>4</sup> In general, the O\*NET SOC codes are finer than the ACS OCC codes, so each descriptor for a lower-level SOC occupation is averaged and subsumed into a higher-level SOC occupation using the SOC-OCC crosswalk. However, O\*NET does not list descriptors for any lower level of some of the three-digit OCC codes, and OCC codes changed in 2018, necessitating additional manipulation by us.
- <sup>5</sup> PP and EDI are positively correlated. In particular, the highest EDI occupations (mostly healthcare-related occupations) also have high PP. At the same time, there are occupations that have high EDI but low PP (e.g., couriers, refuse collectors, and janitors and building cleaners) and those that have high PP but low EDI (e.g., meeting and event planners, engine and other machine assemblers, and parts salespersons).
- <sup>6</sup> Bick, Blandin, and Mertens (2020) also emphasize the difference between index-based potential home-based work and actual home-based work.
- <sup>7</sup> Adams-Prassl et al. (2020) show that this is also true in their own surveys in the United States and the United Kingdom.
- <sup>8</sup> Appendix C shows that the correlation between our indexes and demographics are also weak.
- <sup>9</sup> Thus, the maximal possible reduction in aggregate exposure is  $\psi$ , which may be due to non-compliance, selective furloughs, or a reduction in working hours (such as a curfew on pubs and restaurants) as opposed to a shutdown order, and so on. Note that the value of  $\psi$  does not affect the optimal solution  $\{x_i\}$  but only the magnitudes of the reductions in wage loss and exposure.
- <sup>10</sup> Problem (1) is the dual problem of minimizing exposure subject to a given level of the aggregate wage loss.
- <sup>11</sup> We are not considering job losses at the extensive margin. One could assume that a WFH index  $h_i$  below a certain threshold  $\bar{h}$  implies workers sent home are too unproductive to be employed and hence lose their jobs. With this assumption, one could work out an alternative problem that minimizes job losses rather than the aggregate wage loss.
- <sup>12</sup>  $\psi = 0.473$  corresponds to a 30 percent drop in aggregate wages under the real-world lockdown, to be shown in Section 4.
- <sup>13</sup> Palomino, Rodríguez, and Sebastian (2020) identify essential jobs by ISCO (two digit) and NACE (one digit) from the lockdowns implemented in Italy and Spain. We match these jobs to the OCC and IND codes in the Census. Their list of essential jobs are broadly consistent with the CISA (Cybersecurity and Infrastructure Security Agency) guidelines.
- <sup>14</sup> A labor income share of 60 percent implies an 18 percent drop in gross domestic product (GDP) due to the lockdown, which is between the GDP loss in the United States (10 percent) and the GDP loss in the United Kingdom (20 percent) from the first to the second quarter of 2020.
- <sup>15</sup> Of course, there are other jobs that are truly essential. Those jobs can be incorporated as additional constraints to the minimization problem in (1). While the exemption of a larger set of truly essential jobs will bring our constrained policy closer to real-world lockdowns, one important distinction is that our constrained policy is at the level of industry-occupation pairs, while the real-world policies were industry based and hence more blunt.
- <sup>16</sup> Aum, Lee, and Shin (2021b) document a similar unequal impact of the pandemic—even without a lockdown.

<sup>17</sup> The distributional impact of the constrained optimal policy is in the appendix.

<sup>18</sup> This is by design, since we solved the optimal policy subject to the same reduction in aggregate exposure as for the lockdown. The total reduction across all groups must be the same for the lockdown (light-gray bars) and the optimal policy (dark-gray bars).

<sup>19</sup> While most theoretical and structural articles that analyze the effect of the pandemic and lockdowns incorporate a trade-off between exposure and WFH ability (e.g., Krueger, Uhlig, and Xie, 2020, and Assenza et al., 2020), few consider the heterogeneity of the trade-off at the micro level. Some exceptions are Alon et al. (2020) and Brotherhood et al. (2020), who consider differences by age, and Aum, Lee, and Shin (2021a), who consider differences by worker occupation and skill level.

## REFERENCES

- Adams-Prassl, A; Boneva, T.; Golin, M. and Rau, C. "Work That Can Be Done from Home: Evidence On Variation Within and Across Occupations and Industries." IZA Discussion Paper 13374, Institute of Labor Economics, June 2020.
- Alon, T.; Kim, M.; Lagakos, D. and VanVuren, M. "How Should Policy Responses to the COVID-19 Pandemic Differ in the Developing World?" NBER Working Paper 27273, National Bureau of Economic Research, May 2020; <https://doi.org/10.3386/w27273>.
- Assenza, T.; Collard, F.; Dupaigne, M.; P. Féve; Hellwig, C.; Kankanamge, S. and Werquin, N. "The Hammer and the Dance: Equilibrium and Optimal Policy during a Pandemic Crisis." CEPR Discussion Paper 14731, Centre for Economic Policy Research, May 2020; [https://cepr.org/active/publications/discussion\\_papers/dp.php?dpno=14731](https://cepr.org/active/publications/discussion_papers/dp.php?dpno=14731).
- Aum, Sangmin; Lee, Sang Yoon (Tim) and Shin, Yongseok. "Inequality of Fear and Self-Quarantine: Is There a Trade-Off Between GDP and Public Health?" *Journal of Public Economics*, February 2021a, Volume 194; <https://doi.org/10.1016/j.jpubeco.2020.104354>.
- Aum, Sangmin; Lee, Sang Yoon (Tim) and Shin, Yongseok. "COVID-19 Doesn't Need Lockdowns to Destroy Jobs: The Effect of Local Outbreaks in Korea." *Labour Economics*, June 2021b, Volume 70; <https://doi.org/10.1016/j.labeco.2021.101993>.
- Autor, D.H. and Dorn, D. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review*, August 2020, 103(5), 2013, pp. 1553-97; <https://doi.org/10.1257/aer.103.5.1553>.
- Bick, A.; Blandin, A. and Mertens, K. "Work from Home after the COVID-19 Outbreak." Unpublished manuscript, 2020.
- Brotherhood, Luiz; Kircher, Philipp; Santos, Cezar and Tertilt, Michèle. "An Economic Model of the Covid-19 Epidemic: The Importance of Testing and Age-Specific Policies." CESifo Working Paper Series 8316, 2020; <http://dx.doi.org/10.2139/ssrn.3618840>.
- del Rio-Chanona, R.M.; Mealy, P.; Pichler, A.; Lafond, F. and Farmer, J.D. "Supply and Demand Shocks in the COVID-19 Pandemic: An Industry and Occupation Perspective." *Oxford Review of Economic Policy*, August 2020, 36(Supplement 1), pp. S94-S137; <https://doi.org/10.1093/oxrep/graa033>.
- Dingel, J.I. and Neiman, B. "How Many Jobs Can Be Done at Home?" *Journal of Public Economics*, September 2020, 189, 104235; <https://doi.org/10.1016/j.jpubeco.2020.104235>.
- Gamio, L. "The Workers Who Face the Greatest Coronavirus Risk." *New York Times*, March 15, 2020.
- Gottlieb, C.; Grobovšek, J; Poschke, M. and Saltiel, F. "Lockdown Accounting." Centre for Economic Policy Research *Covid Economics: Vetted and Real-Time Papers*, June 2020.
- Hicks, M.J.; Faulk, D. and Devaraj, S. "Occupational Exposure to Social Distancing: A Preliminary Analysis Using O\*NET Data." Technical report, Center for Business and Economic Research, Ball State University, March 2020.
- Koren, M. and Petó, R. "Business Disruptions from Social Distancing." *PLOS ONE*, September 18, 2020, 15(9), pp. 1-14; <https://doi.org/10.1371/journal.pone.0239113>.
- Krueger, D.; Uhlig, H. and Xie, T. "Macroeconomic Dynamics and Reallocation in an Epidemic." Centre for Economic Policy Research *Covid Economics: Vetted and Real-Time Papers*, April 2020; <https://cepr.org/sites/default/files/CovidEconomics5.pdf>.

- Lee, S.Y. and Shin, Y. "Horizontal and Vertical Polarization: Task-Specific Technological Change in a Multi-Sector Economy." NBER Working Paper 23283, National Bureau of Economic Research, March 2017; <https://doi.org/10.3386/w23283>.
- Leibovici, F.; Santacreau, A. and Familgietti, M. "Social Distancing and Contact-Intensive Occupations." Federal Reserve Bank of St. Louis *On the Economy Blog*, March 24, 2020; <https://www.stlouisfed.org/on-the-economy/2020/march/social-distancing-contact-intensive-occupations>.
- Mongey, S.; Pilossoph, L. and Weinberg, A. "Which Workers Bear the Burden of Social Distancing Policies?" Centre for Economic Policy Research *Covid Economics: Vetted and Real-Time Papers*, May 2020; <https://cepr.org/sites/default/files/CovidEconomics12.pdf>.
- Palomino, J.C.; Rodríguez, J.G. and Sebastian, R. "Wage Inequality and Poverty Effects of Lockdown and Social Distancing in Europe." *European Economic Review*, 2020, 129, 103564; <https://doi.org/10.1016/j.eurocorev.2020.103564>.

# Failing to Provide Public Goods: Why the Afghan Army Did Not Fight

Rohan Dutta, [David K. Levine](#), and Salvatore Modica

The theory of public goods is mainly about the difficulty in paying for them. Our question here is this: Why might public goods not be provided, even if funding is available? We use the Afghan Army as our case study. We explore this issue using a simple model of a public good that can be provided through collective action and peer pressure, by modeling the self-organization of a group (the Afghan Army) as a mechanism design problem. We consider two kinds of transfer subsidies from an external entity such as the U.S. government. One is a Pigouvian subsidy that simply pays the salaries, rewarding individuals who provide effort. The second is an output/resource multiplier (the provision of military equipment, tactical skill training, and so forth) that amplifies the effort provided through collective action. We show that the introduction of a Pigouvian subsidy can result in less effort being provided than in the absence of a subsidy. By contrast, an output/resource multiplier subsidy, which is useful only if collective action is taken, necessarily increases output via an increase in effort. Our conclusion is that the United States provided the wrong kind of subsidy, which may have been among the reasons why the Afghan Army did not fight. (JEL A1, D7, D9)

Federal Reserve Bank of St. Louis *Review*, Second Quarter 2022, 104(2), pp. 110-19.  
<https://doi.org/10.20955/r.104.110-19>

---

“We paid their salaries...What we could not provide was the will to fight.” **President Joe Biden**<sup>1</sup>

## 1 INTRODUCTION

The speed with which the Afghan Army collapsed and the Taliban took over Afghanistan came as a surprise to many, but not to economists or individuals versed in game theory. By backward induction, if you plan to surrender anyway, then sooner is generally better than later (unless you are indifferent between present versus future payoffs). This article, however, explores a deeper question: Why would a well-equipped army that outnumbered their opponents by three or four to one in manpower and with decades of training plan to surrender to an apparently much inferior opponent?

Rohan Dutta is an associate professor of economics at McGill University. David K. Levine is a professor of economics at the European University Institute, a professor of economics emeritus at Washington University in St. Louis, and a research fellow at the Federal Reserve Bank of St. Louis. Salvatore Modica is a professor of economics at Università di Palermo, Dipartimento SEAS. We thank Andrea Mattozzi. We gratefully acknowledge support from the MIUR PRIN 2017 n. 2017H5KPLL\_01.

© 2022, Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

The quote from President Joe Biden’s speech following the fall of Kabul is revealing: What Biden and many others fail to understand is that there is a causal connection between paying the salaries of the Afghan Army and the fact that they lacked the will to fight. Our goal in this article is to explain why that is so and why it need not have been so.

Insofar as nation building is measured by the national defense, the place to start is to understand that national defense is a public good. Many Afghans would prefer not to be ruled by the Taliban, but most would prefer that someone else do the fighting. This problem of free riding is endemic to public goods problems, and economists and other social scientists have analyzed these problems for over a century. We recognize three ways of overcoming the free-rider problem. The most familiar one involves collective action through formal systems, usually Pigouvian taxes or subsidies, which are widely recommended by economists to achieve, for example, reductions in carbon emissions to combat global warming. A second, less formal means of providing public goods is through voluntary provision: People contribute to a public good either because the personal benefit of the public good is sufficiently great to outweigh the cost of contributing or because they are altruistic and desirous of helping society (e.g., by funding public radio and television in the United States [NPR and PBS]). There is little evidence, however, that voluntary public goods provision can provide public goods on a large scale—for an entire country, for example, as in the case of an army.

We wish to focus on a third means of public good provision: providing incentives informally or “socially” through means such as peer pressure, resulting in various forms of ostracism of those who fail to contribute. Although economists have not studied this to the extent that they have studied taxes and subsidies, we know, particularly from the work of Coase (1960), Ostrom (1990), and Townsend (1994), that these methods work in practice. Indeed, the effectiveness of large-scale lobbying organizations such as farmers show that peer pressure can be effective even on a very large scale. After all, all farmers want the benefits of farm subsidies but prefer that other farmers bear the cost of lobbying.

Of course, while groups and societies can collectively self-organize social norms that induce provision of public goods, they may also choose simply to follow the “law of the jungle” and allow members to go their own way and free ride as they wish. The right choice depends on how valuable the public good is and how costly it is to organize and enforce collective decisions. Our starting observation is that the intervention of outside agencies—be they non-governmental organizations (NGOs) or the United States—changes the trade-offs for collective decisionmaking. Simply put, if the United States pays the salaries of the Afghan Army, then there is little benefit from the Afghans collectively organizing to encourage people to join the army and fight for their country. In practice, if the salary is sufficiently high relative to the outside option, people might join but will not fight when it is time to deliver. Indeed, General Wesley Clark, former NATO supreme allied commander, gives the following description of the motivation of Afghan soldiers: “People signed up with the Afghan military to make money...but they did not sign up to fight to the death, for the most part.”<sup>2</sup> Contrast this with J.R.R. Tolkien’s description of Britain at the start of World War I: “In those days chaps joined up, or were scorned publicly.”<sup>3</sup> We think it is reasonable to assume that such peer pressure to defend the country did not exist in Afghanistan.

Here is the key point: The displacement of self-organization by subsidy can result in less provision of the public good than in the absence of the subsidy. In other words, subsidizing a public good—the Pigouvian approach—can reduce the provision of that good if it displaces

self-organization. The reason is that self-organization is costly, and so the benefit of not organizing can exceed the cost of having less of the public good.

In this article, we examine a simple model of a public good that can be provided through collective action and peer pressure and examine the effect of subsidies. Our model follows Townsend (1994), Levine and Modica (2016), and Dutta, Levine, and Modica (2022) by modeling the self-organization of a group as a mechanism design problem. The group establishes an output quota, it has a noisy monitoring technology for observing whether the quota is followed, and it can punish group members based on these signals. A key feature of the model is that if monitoring and punishment is to be used, it has an associated fixed cost that includes both the physical cost of monitoring and the costs of negotiating and finding an agreement as to what the mechanism will be.

In this setting, we consider two kinds of subsidies. One is a Pigouvian subsidy that simply “pays the salaries,” rewarding individuals who provide effort. We show that this can result in less effort being provided than in the absence of a subsidy. The other is an output multiplier: the provision of training, equipment, and so forth, that amplifies the effort provided through collective action. Because such provision is useful only if collective action is taken, unlike with a Pigouvian subsidy, it necessarily increases output.

In Dutta, Levine, and Modica (2022), we showed more broadly how Pigouvian subsidies can have the perverse effect of undermining existing collective action. We pointed there to the case of NGOs. Bano (2012) did extensive field research in Pakistan. She documented how public goods, particularly welfare, were provided through voluntary efforts with socially provided incentives for contribution. Donor organizations—mostly NGOs—subsequently attempted to increase public good provision through subsidies in the form of salaries to those contributing to the public good. In a series of case studies, she showed how such subsidies led to the unraveling of the provision of social incentives and ultimately to decreased provision of the public good. In one of several cases, she indicates that “the Maternity and Child Welfare Association...almost collapsed with the influx of such aid.”<sup>4</sup>

The evidence now shows that similar considerations can be applied to the Afghan Army. We cannot know how strong the social pressure to self-organize resistance to the Taliban would have been without subsidies; but the fact is that by paying the salaries of soldiers, the incentive for collective action to encourage volunteers to join the army for the common good was reduced so much that provision of the public good—measured not by the number of soldiers, but by the number of soldiers willing to fight—was minimal. Hence, the Taliban, an army recruited through social incentives, predominates and once again rules Afghanistan.

The collapse of Afghanistan is often compared to the collapse of South Vietnam. In this context it is worth pointing out that the United States did not pay soldiers’ salaries in South Vietnam but only provided subsidies in the form of training and equipment. What is less well known is that, as a result, the South Vietnamese Army (ARVN) did fight.<sup>5</sup> The United States withdrew from Vietnam in 1973. In the next year the ARVN largely drove the Vietcong, the North Vietnam irregular army somewhat akin to the Taliban, out of South Vietnam. In 1975 the North invaded with a large regular army of similar strength to the ARVN, including a great many artillery pieces. The fighting lasted about four months, and the casualties on both sides combined were about 45,000 killed and 80,000 wounded. This is greatly different from Afghanistan, where a large superior well-equipped military refused to fight and was defeated in weeks by a small, lightly armed group of irregular fighters.

The bottom line is not entirely negative—either for nation building or for NGOs. It is not that help cannot be provided, but care must be taken that the help provided does not undermine the provision of effort through collective action and social norms. Hence, providing military training and equipment will generally result in greater defense, just as providing computers and training to charitable organizations can do the same.

## 2 THE MODEL

Identical group members  $i \in [0,1]$  engage in production, choosing a real valued level of output  $X \geq x^i \geq 0$  such as fighting effort in the case of Afghanistan. The utility of a member  $i$  depends on a vector-valued state of the world  $\omega \geq 0$ , their own output, and the average output of the group  $x = \int x^i di$  according to  $u(\omega, x, x^i)$ , where  $u(\omega, x, x^i)$  is specified below.

The output of the group  $x$  is a public good—in our case national defense. Because all members benefit from the public good, the group collectively faces a mechanism design problem, and we assume that incentives can be given to group members in the form of individual punishments based on monitoring: The group can set a production quota  $y$  and receives signals of whether or not individual output adheres to the quota. Based on these signals it can impose punishments. Specifically, monitoring generates a noisy signal  $z^i \in \{0,1\}$ , where zero means “good, likely respected the quota” and 1 means “bad, likely produced less than the quota.” The probability of the bad signal is  $\pi > 0$  if  $x^i \geq y$  and  $\pi_B > \pi$  if  $x^i < y$ . When the signal is bad, the group imposes an endogenous utility penalty of  $P$ . This may be in the form of social disapproval or even in the form of monetary penalties.

The social cost of the punishment  $P$  is  $\psi P$ , where  $\psi > 0$  could be greater or less than 1. For example, if the punishment is that group members are prohibited from drinking beer with the culprit, that might be costly to the culprit’s friends as well as the culprit. In this case,  $\psi > 1$ . Or it might be that the punishment is a monetary fine, most of which is shared among the group members. In that case, there would be very little social loss, so we would expect  $\psi < 1$ .

In addition to the social cost of punishment, there is a fixed cost  $F \geq 0$  of choosing  $P > 0$ . There are two reasons we expect  $F$  to be positive. First, there will generally be costs of operating the monitoring system—for example, sending spies to observe output. Second, it is costly to gather group members to negotiate an agreement and form a consensus on what the mechanism will be.

The tools available for mechanism design consist of a quota  $y$  and a punishment level for a bad signal  $P$ . The overall utility of a member  $i$  is  $u(\omega, x, x^i) - \pi P$  if  $x^i \geq y$  and  $u(\omega, x, x^i) - \pi_B P$  if  $x^i < y$ . These utilities define a game for the group members. If the mechanism designer chooses  $(y, P)$ , we denote by  $X(y, P)$  the set of  $x$  such that  $x^i = x$  is a symmetric pure strategy Nash equilibrium of this game. We refer to a triple  $(x, y, P)$  with  $x \in X(y, P)$  as an *incentive compatible social norm*.<sup>6</sup> If an incentive compatible social norm issues no punishments ( $P = 0$ ), we call it *noncooperative*. The mechanism designer is benevolent, and welfare from an incentive compatible social norm  $(x, y, P)$  is given by

$$W(x, y, P) \equiv u(\omega, x, x) - \pi \psi P - F \cdot 1\{P > 0\}.$$

We now specify the utility function and interventions. Each individual has a private cost of output  $(c/2)(x^i)^2$ , and the benefit of the public good is  $x - (1 - c)(1/2)x^2$ . These units are chosen so that the first-best  $x^f$  maximizing  $-(c/2)x^2 + x - (1 - c)(1/2)x^2$  is normalized so that  $x^f = 1$ . We take

the effort limit  $X$  to coincide with the satiation level for the public good gross benefit  $x - (1 - c)(1/2)x^2$ ; so  $X = 1/(1 - c)$ , and we assume that  $0 < c < 1$ .

A novel aspect of the present article—which is what drives its main result—is the exact specification of the state variable. The state  $\omega$  has two components: a Pigouvian subsidy  $\omega_s$ , which may be thought of as contributing to the salary of group members who provide effort, and an output multiplier  $\omega_m$ , which may be thought of as equipment and training that increases the effectiveness of effort provided by group members. Overall individual utility is then given by  $u(\omega, x, x^i) = -(c/2)(x^i)^2 + \omega_s x^i + (1 + \omega_m)x - (1 - c)(1/2)((1 + \omega_m)x)^2$ . We are going to show that the two types of subsidies—in our case provided mainly by the United States—have quite different consequences in terms of effort provision in the organization.

We define the *monitoring difficulty* as  $\theta = \psi\pi/(\pi_B - \pi)$ .

### 3 SUBSIDIES ARE BAD, TRAINING IS GOOD

We are interested in reversals, that is, conditions under which introducing a subsidy reduces the effort level  $\hat{x}(\omega)$  that solves the mechanism design problem. Recalling that the group solves a mechanism design problem, we denote the optimal choice of  $x$  conditional on  $P > 0$  by  $x^M(\omega)$  and the output of the noncooperative social norm (the Nash equilibrium) by  $x^N(\omega)$ . The solution to the design problem  $\hat{x}$  may be either  $x^M$  or  $x^N$ . Formally, there is a *reversal* if  $(1 + \omega_m)\hat{x}(\omega) < \hat{x}(0)$ . By *no reversal* we mean the opposite inequality holds (strictly). We show in the appendix that  $x^N(\omega)$  and  $x^M(\omega)$  are strictly increasing and that for the relevant range of  $\omega$  it is  $x^M(0) > (1 + \omega_m)x^N(\omega)$ . It follows that the only way in which a reversal can occur is if the group prefers to use the mechanism  $M$  with punishment at  $\omega = 0$  but reverts to noncooperation at  $\omega > 0$ . That is, the subsidy, salaries paid by the United States for example, substitutes for the costly use of peer pressure. Our main result states the conditions under which this may or may not occur:

**Theorem 1.** *For each  $\omega_s$  in a range  $0 < \omega_s < \bar{\omega}_s$  and for all sufficiently small  $F > 0$  and  $\omega_m \geq 0$ , there is a reversal. On the contrary, for each  $\omega_m$  in a range  $0 < \omega_m < \bar{\omega}_m$  and for all  $F \geq 0$  and sufficiently small  $\omega_s \geq 0$ , there is no reversal.*

Subsidies, in other words, are bad in the sense that they can reduce output, while training can only increase output. The result is proved in the appendix.

### 4 DISCUSSION

How does the model explain the collapse of the Afghan Army? According to our theory, the payment of salaries by the United States substituted for a peer enforcement system: When the salary payments were withdrawn, we might expect that it would have been replaced by a peer enforcement system—but there was no time for this. Hence, it was best to surrender right away and not, for example, allow the relatively small number of commandos who were ready to fight to do so. We can of course point to other elements: the corruption of the Afghan government, the removal of air support when the United States withdrew, and the fact that those who were called upon to fight—mostly young men—were different from those who benefited significantly from the fighting—mostly women and older men.<sup>2</sup> The point to emphasize is that none of these things are peculiar to Afghanistan. In the case of Vietnam, where the ARVN did fight, it is equally true that the government was corrupt



and that the United States stopped providing air support. Moreover, it is hard to think of any war in which those who fought were not young men—a group that typically has the lowest relative benefit from victory. The stay-at-homes, be they women or older men, tend to benefit the most. More broadly, while support for the Taliban was strong in opium-producing areas, they were opposed in most other areas, and self-organization in the absence of salary subsidies is not implausible.<sup>8</sup>

We conclude by pointing out that the theory that the collapse of social norms can produce the “perverse” effect of reducing public good output has support in other realms. In the introduction, we pointed to the work of Bano (2012) on NGOs in Pakistan. In a field experiment, Gneezy and Rustichini (2000) examined the effect of a fine on the provision of a public bad—picking children up late from a day-care center. They found that the fine resulted in more parents picking up their children late—an increase of the public bad and the opposite of the expected and intended effect. In a quite different context, Dutta, Levine, and Modica (2022) showed that similar considerations explain why in the face of an enormous drop in demand for oil due to COVID-19 the OPEC+ cartel increased their output. Here also, increased output is a public bad and the exogenous state is not a Pigouvian fine but a reduction in demand. Ordinarily we would expect lower demand to reduce output, but in fact it resulted instead in the inability to reach an agreement over quotas and in a noncooperative social norm in which output—the public bad—went up. ■

## APPENDIX

Here we prove Theorem 1. The proof follows from a few lemmas proven below.

Recalling that the group solves a mechanism design problem, we denote the optimal choice of  $x$  conditional on  $P > 0$  by  $x^M(\omega)$  and the output of the noncooperative social norm by  $x^N(\omega)$ . The solution to the design problem  $\hat{x}$  may be either  $x^M$  or  $x^N$ . The corresponding optimal values exclusive of fixed cost are  $u^M(\omega)$  and  $u^N(\omega)$ . We say that  $\omega$  is of *moderate size* if  $(1 + \omega_m)\omega_s \leq c/(1 + \theta c)$ .

Lemma 3 shows that  $x^N(\omega)$  and  $x^M(\omega)$  have strictly positive partial derivatives so are strictly increasing. We also show for moderate  $\omega$  that  $x^M(0) > (1 + \omega_m)x^N(\omega)$ . It follows that the only way in which a reversal can occur— $(1 + \omega_m)\hat{x}(\omega) < \hat{x}(0)$ —is if the group prefers to use the mechanism  $M$  with punishment at  $\omega = 0$  but reverts to noncooperation at  $\omega > 0$ .

The group will only use  $M$  at  $\omega = 0$  if  $F \leq u^M(0) - u^N(0)$  and will only use  $N$  at  $\omega$  if  $F \geq u^M(\omega) - u^N(\omega)$  and has strict preferences when the inequalities are strict. If  $u^M(\omega) - u^N(\omega) > u^M(0) - u^N(0)$ , then there can be no such  $F$ . Conversely, if  $u^M(\omega) - u^N(\omega) < u^M(0) - u^N(0)$ , then there will be a reversal for any  $u^M(\omega) - u^N(\omega) < F < u^M(0) - u^N(0)$ . Hence, we must know if  $u^M(\omega) - u^N(\omega)$  increases or decreases with  $\omega$ .

In Lemma 3 we show that at  $\omega = 0$  the partial derivative of  $u^M(\omega) - u^N(\omega)$  is negative with respect to  $\omega_s$  and positive with respect to  $\omega_m$ .

Take  $\omega_s$  first. Since the partial derivative of  $u^M(\omega) - u^N(\omega)$  is negative at zero, there is a range of  $\omega_s$  for which  $u^M(\omega) - u^N(\omega)$  is strictly decreasing; hence, for any  $\omega_s$  in that range and  $\omega_m = 0$ , we have  $u^M(\omega) - u^N(\omega) < u^M(0) - u^N(0)$ . Since  $u^M(\omega)$  and  $u^N(\omega)$  are continuous by Lemmas 1 and 2, it follows that this remains true for  $\omega_m$  sufficiently small. Hence, we obtain the first result, that there is a range of  $F$ 's for which there is a reversal.

Similar reasoning with respect to  $\omega_m$  shows that there is a range of  $\omega_m$  with  $\omega_s = 0$  for which  $u^M(\omega) - u^N(\omega) > u^M(0) - u^N(0)$ , and again by continuity this continues to hold for sufficiently small  $\omega_s$ , giving the result about no reversal.

The cited lemmas follow.

**Lemma 1.** *The individual optimum is  $x^B(\omega) = \omega_s/c$  with utility  $u(\omega, x, x^B) = \omega_s^2/(2c) + (1 + \omega_m)x - (1 - c)(1/2)((1 + \omega_m)x)^2$ . As the optimum is independent of  $x$ , this is also the noncooperative (Nash) social norm:  $x^N(\omega) = \omega_s/c$ , with corresponding welfare*

$$u^N(\omega) = u(\omega, x^N, x^N) = \omega_s^2 / (2c) + (1 + \omega_m)\omega_s / c - (1 - c)(1/2)((1 + \omega_m)\omega_s / c)^2.$$

*Proof.* The first sentence follows from maximizing the objective,  $u(\omega, x, x^i) = -(c/2)(x^i)^2 + \omega_s x^i + (1 + \omega_m)x - (1 - c)(1/2)((1 + \omega_m)x)^2$  with respect to  $x^i$ .

**Lemma 2.** *The optimal incentive compatible quota  $x^M(\omega)$  and the corresponding utility  $u^M(\omega)$  are given by*

$$x^M(\omega) = \frac{(1 + \theta)\omega_s + 1 + \omega_m}{(1 + \theta)c + (1 - c)(1 + \omega_m)^2}$$

and

$$u^M(\omega) = \frac{1}{2} \frac{((1 + \theta)\omega_s + 1 + \omega_m)^2}{(1 + \theta)c + (1 - c)(1 + \omega_m)^2} - \theta\omega_s^2 / (2c).$$

*Proof.* Given  $\omega$  and a quota  $y = x$ , the incentive constraint is  $u(\omega, x, x) - \pi P \geq u(\omega, x, x^B) - \pi_B P$ , so the quota is made incentive compatible by punishment

$$P = [u(\omega, x, x^B) - u(\omega, x, x)] / (\pi_B - \pi).$$

Then monitoring cost is  $\pi P$ , yielding a social utility exclusive of fixed costs of

$$\begin{aligned} & u(\omega, x, x) - \theta [u(\omega, x, x^B) - u(\omega, x, x)] \\ &= (\omega_s + 1 + \omega_m)x - \left( (1 - c)(1 + \omega_m)^2 + c \right) x^2 / 2 - \theta \left[ \omega_s^2 / (2c) - \omega_s x + cx^2 / 2 \right] \\ &= (\omega_s + 1 + \omega_m + \theta\omega_s)x - \left( (1 - c)(1 + \omega_m)^2 + c \right) x^2 / 2 - \left[ \theta\omega_s^2 / (2c) + \theta cx^2 / 2 \right] \\ &= (\omega_s + 1 + \omega_m + \theta\omega_s)x - \left( (1 - c)(1 + \omega_m)^2 + c + \theta c \right) x^2 / 2 - \theta\omega_s^2 / (2c) \\ &= -(1/2) \left[ (1 - c)(1 + \omega_m)^2 + (1 + \theta)c \right] x^2 + ((1 + \theta)\omega_s + 1 + \omega_m)x - \theta\omega_s^2 / (2c). \end{aligned}$$

Using the fact that maximizing  $Ax - (B/2)(x)^2$  has solution  $x = A/B$  and optimum value,  $A^2/(2B)$  yields the values of  $x^M, u^M$  given in the result.

We say that  $\omega$  is of *moderate size* if  $(1 + \omega_m)\omega_s \leq c/(1 + \theta c)$ .

**Lemma 3.** *If  $\omega$  is of moderate size, then  $x^M(0) > (1 + \omega_m)x^N(\omega)$ ,*

$$\frac{\partial(1 + \omega_m)x^M}{\partial\omega_s}, \quad \frac{\partial(1 + \omega_m)x^M}{\partial\omega_m} > 0$$

and

$$\left. \frac{\partial [u^M - u^N]}{\partial \omega_s} \right|_{\omega=0} < 0, \quad \left. \frac{\partial [u^M - u^N]}{\partial \omega_m} \right|_{\omega=0} > 0.$$

*Proof.* From Lemma 2,

$$x^M(0) = \frac{1}{(1+\theta)c + (1-c)} = \frac{1}{1+\theta c},$$

while from Lemma 1,  $(1 + \omega_m)x^N(\omega) = (1 + \omega_m)\omega_s/c$ . Hence,  $x^M(0) > (1 + \omega_m)x^N(\omega)$  exactly when

$$(1 + \omega_m)\omega_s < \frac{c}{1 + \theta c},$$

which says that  $\omega$  is of moderate size.

Next we assess the partial derivatives of  $(1 + \omega_m)x^M(\omega)$ , where  $x^M(\omega)$  is given in Lemma 2. We have

$$\frac{\partial (1 + \omega_m)x^M}{\partial \omega_s} = \frac{(1 + \omega_m)(1 + \theta)}{(1 + \theta)c + (1 - c)(1 + \omega_m)^2} > 0.$$

A little algebra shows that

$$\frac{\partial (1 + \omega_m)x^M}{\partial \omega_m} = (1 + \theta) \frac{(1 + \theta)c\omega_s + 2c(1 + \omega_m) - (1 - c)\omega_s(1 + \omega_m)^2}{\left( (1 + \theta)c + (1 - c)(1 + \omega_m)^2 \right)^2}.$$

Divide the numerator by  $1 - c$  and observe that

$$\frac{c}{1 - c} [(1 + \theta)\omega_s + 2(1 + \omega_m)] - \omega_s(1 + \omega_m)^2 \geq \frac{c}{1 - c} [2(1 + \omega_m)] - \omega_s(1 + \omega_m)^2,$$

where the right-hand side has the same sign as

$$\frac{2c}{1 - c} - \omega_s(1 + \omega_m).$$

So the derivative is positive if  $\omega_s(1 + \omega_m) < 2c/(1 - c)$ ; and this holds by moderation:  $\omega_s(1 + \omega_m) < c/(1 + \theta c) < 2c/(1 - c)$ .

Finally, we assess the partial derivatives of  $u^M(\omega) - u^N(\omega)$  at  $\omega = 0$ . From Lemmas 1 and 2, the difference  $u^M(\omega) - u^N(\omega)$  is given by

$$\begin{aligned} & \frac{1}{2} \frac{\left( (1 + \theta)\omega_s + 1 + \omega_m \right)^2}{(1 + \theta)c + (1 - c)(1 + \omega_m)^2} - \frac{\theta\omega_s^2}{2c} - \frac{\omega_s^2}{2c} - \frac{(1 + \omega_m)\omega_s}{c} + \frac{1}{2}(1 - c) \left( \frac{(1 + \omega_m)\omega_s}{c} \right)^2 \\ &= \frac{1}{2} \frac{\left( (1 + \theta)\omega_s + 1 + \omega_m \right)^2}{(1 + \theta)c + (1 - c)(1 + \omega_m)^2} - \frac{(1 + \theta)\omega_s^2}{2c} - \frac{(1 + \omega_m)\omega_s}{c} + \frac{1}{2}(1 - c) \frac{(1 + \omega_m)^2\omega_s^2}{c^2}. \end{aligned}$$

From this,

$$\begin{aligned} \left. \frac{\partial [u^M - u^N]}{\partial \omega_s} \right|_{\omega=0} &= \frac{(1+\theta)}{(1+\theta)c + (1-c)} - 1/c = \frac{c(1+\theta) - (1+\theta)c - (1-c)}{c[(1+\theta)c + (1-c)]} \\ &= \frac{-(1-c)}{c[(1+\theta)c + (1-c)]} < 0 \end{aligned}$$

and

$$\begin{aligned} \left. \frac{\partial [u^M - u^N]}{\partial \omega_m} \right|_{\omega=0} &= \frac{1}{2} \cdot \frac{2(1+\omega_m)((1+\theta)c + (1-c)(1+\omega_m)^2) - 2(1+\omega_m)^2((1-c)(1+\omega_m))}{((1+\theta)c + (1-c)(1+\omega_m)^2)^2} \\ &= \frac{(1+\theta)c + (1-c) - (1-c)}{((1+\theta)c + (1-c))^2} = \frac{(1+\theta)c}{((1+\theta)c + (1-c))^2} > 0. \end{aligned}$$

This concludes the proof.

## NOTES

- <sup>1</sup> Biden (2021).
- <sup>2</sup> Clark (2021).
- <sup>3</sup> Tolkien (1981).
- <sup>4</sup> Bano (2012, p. 126).
- <sup>5</sup> The historical facts are not controversial and are discussed in many histories; see, for example, Willbanks (2004).
- <sup>6</sup> In the language of contract theory, it is an enforcement contract with costly state verification.
- <sup>7</sup> Winning a war would require a great benefit to provide a net benefit to those who fight it. Indeed, fighters are generally those from relatively modest backgrounds for whom the type of government is not of crucial importance. In the case of Afghanistan, men are actually likely to do *relatively* well under Taliban rule.
- <sup>8</sup> See Mery-Khosrowshahi (2021).

## REFERENCES

- Bano, M. *Breakdown in Pakistan: How Aid Is Eroding Institutions for Collective Action*. Stanford University Press, 2012.
- Biden, Joe. "Remarks by President Biden on the Economy." White House, September 16, 2021; <https://www.whitehouse.gov/briefing-room/speeches-remarks/2021/09/16/remarks-by-president-biden-on-the-economy-4/>.
- Clark, Wesley. Interview with CNN. "Former NATO Commander Speaks to CNN About the Taliban's Takeover." August 18, 2021; [https://www.cnn.com/world/live-news/afghanistan-taliban-us-news-08-18-21/h\\_10d2c797318aff2a0d10497b762164d5](https://www.cnn.com/world/live-news/afghanistan-taliban-us-news-08-18-21/h_10d2c797318aff2a0d10497b762164d5).
- Coase, R.H. "The Problem of Social Cost." *Journal of Law and Economics*, 1960, 3, pp. 1-44; <https://doi.org/10.1086/466560>.
- Dutta, R.; Levine, D.K. and Modica, S. "Interventions with Sticky Social Norms: A Critique." *Journal of the European Economic Association*, February 2022, 20(1), pp. 39-76; <https://doi.org/10.1093/jeea/jvab015>.
- Gneezy, U. and Rustichini, A. "A Fine Is a Price." *Journal of Legal Studies*, 2000, 29(1), pp. 1-17; <https://doi.org/10.1086/468061>.
- Levine, David and Modica, Salvatore. "Peer Discipline and Incentives Within Groups." *Journal of Economic Behavior and Organization*, 2016, 123, pp. 19-30; <https://doi.org/10.1016/j.jebo.2015.12.006>.

Mery-Khosrowshahi, Aschkan. "The Opium of the People: Essays on Counter-Narcotics Efforts in Afghanistan." PhD Thesis, European University Institute, 2021.

Ostrom, Elinor. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, 1990.

Tolkien, J.R.R. "Letter 43," in H. Carpenter and C. Tolkien, eds., *The Letters of JRR Tolkien*. Houghton Mifflin, 1981.

Townsend, Robert M. "Risk and Insurance in Village India." *Econometrica*, 1994, 62, pp. 539-39;  
<https://doi.org/10.2307/2951659>.

Willibanks, James H. *Abandoning Vietnam: How America Left and South Vietnam Lost Its War*. University Press of Kansas, 2004.

# Venture Capital: A Catalyst for Innovation and Growth

*Jeremy Greenwood, Pengfei Han, and [Juan M. Sánchez](#)*

This article studies the development of the venture capital (VC) industry in the United States and assesses how VC financing affects firm innovation and growth. The results highlight the essential role of VC financing for U.S. innovation and growth and suggest that VC development in other countries could promote their economic growth. (JEL E13, E22, G24, L26, O16, O31, O40)

Federal Reserve Bank of St. Louis *Review*, Second Quarter 2022, 104(2), pp. 120-30.  
<https://doi.org/10.20955/r.104.120-30>

---

## 1 INTRODUCTION

Venture capital (VC) is a particular type of private equity that focuses on investing in young companies with high-growth potential. The companies and products and services VC helped develop are ubiquitous in our daily lives: the Apple iPhone, Google Search, Amazon, Facebook and Twitter, Starbucks, Uber, Tesla electric vehicles, Airbnb, Instacart, and the Moderna COVID-19 vaccine. Although these companies operate in drastically different industries and with dramatically different business models, they share one common and crucial footprint in their corporate histories: All of them received major financing and mentorship support from VC investors in the early stages of their development.

This article outlines the history of VC and characterizes some stylized facts about VC's impact on innovation and growth. In particular, this article empirically evaluates the relationship between VC, firm growth, and innovation.

## 2 THE VC INDUSTRY IN THE UNITED STATES

This section outlines the historical background on the rise of VC firms as limited partnerships and characterizes some stylized facts about the VC industry in the United States.

Jeremy Greenwood is a professor of economics at the University of Pennsylvania. Pengfei Han is an assistant professor of finance at Guanghua School of Management at Peking University. Juan M. Sánchez is a vice president and economist at the Federal Reserve Bank St. Louis. We thank Ana Maria Santacreu for helpful comments.

© 2022, Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

## **2.1 Historical Background: VC Firms as Limited Partnerships**

Financing cutting-edge technologies has always been challenging.<sup>1</sup> It is difficult to know whether new ideas are viable, if they will be salable, and how best to bring them to market. Also, it is important to ensure that entrepreneurs' and investors' incentives are aligned. Traditional financial institutions, such as banks and equity/securities markets, are not well suited to engage in this sort of underwriting. Historically, the introduction of new technologies was privately financed by wealthy individuals. Investors were plugged into networks of inventive activity in which they learned about new ideas, vetted them, and drew on the expertise needed to operationalize them. These financiers are similar to today's "angel investors."

The Brush Electric Company provided such a network for inventors and investors in Cleveland around the turn of the twentieth century. The use of electricity rapidly expanded during the Second Industrial Revolution. Individuals linked with the Brush Electric Company network spawned ideas for arc lighting, liquefying air, smelting ores electrically, and electric cars and trolleys, among other things. The shops at Brush were a meeting place for inventors; they could develop and debug new ideas with help from others. Investors connected with the Brush network learned about promising new ideas from the scuttlebutt at the shops. They became partners/owners in the firms that they financed. Interestingly, in the Midwest at the time, prolific inventors (those with more than 15 patents) who were principals in companies were much more likely than other investors to keep their patents or assign them to the companies where they were principals; other investors typically sold their patents to businesses where they had no concern. These practices aligned the incentives of innovators and investors.

World War II and the start of the Cold War ushered in new technologies, such as jets, nuclear weapons, radars, and rockets, along with a splurge of spending by the U.S. Department of Defense. A handful of VC firms were formed to leverage the commercialization of scientific advances. American Research and Development (ARD), founded by General Georges Doriot and others, was one of these. ARD pulled in money from mutual funds, insurance companies, and an initial public stock offering. The founders knew that it was important for venture capitalists to provide advice to the fledging enterprises in which they were investing. In 1956, ARD invested \$70,000 in Digital Equipment Corporation (DEC) in exchange for a 70 percent equity stake. ARD's share was worth \$38.5 million when DEC went public in 1966, which represented an annual return of 100 percent. While this investment was incredibly successful, the organizational form of ARD did not come to dominate the industry. The compensation structure of ARD made it difficult for the company to retain the VC professionals needed to evaluate startups and provide the guidance necessary for success.

An alternative organizational form came to emblemize the industry: the limited partnership. This form is exemplified by the formation of Davis and Rock in 1961. These partnerships allowed VC professionals to share in the gains from startups along with the entrepreneurs and investors. Limited partnerships served to align venture capitalists' interests with those of entrepreneurs, investors, and key employees. Money was put in only at the beginning of the partnership. The general partners received management fees as a salary plus a share of the capital gains from the investments, say 40 percent, with the limited partners earning 60 percent. The limited partners had no say in the decisions of the general partners. The partnerships were structured for a limited length of time, say seven to ten years. The returns from the partnership were paid out to the investors only

when the partnership was dissolved—there were no dividends, interest payments, etc. Therefore, the returns upon dissolution were subject only to capital gains taxation at the investor level. The VC industry also used stock options to reward founders, CEOs, and key employees. Thus, these recipients too were subject to capital gains taxation rather than taxation on labor income. The short time horizon created pressure to ensure a venture’s rapid success.

Banks and other financial institutions are not well suited to invest in cutting-edge new ventures. While banks are good at evaluating systematic lending risk, they have limited ability to judge the skill of entrepreneurs or the worth of new technologies and limited expertise to help commercialize them. Additionally, financial institutions have faced roadblocks to investing in ventures. The Glass-Steagall Banking Act of 1933, which was later repealed in 1999, prohibited banks from taking equity positions in industrial firms. The Allstate Insurance Company created a private placements program in the 1960s to undertake VC-type investments; however, it abandoned the program because it could not compensate the VC professionals enough to retain them. The Employee Retirement Income Security Act of 1974 prevented pension funds (and dissuaded other traditional fiduciaries) from investing in high-risk ventures. The act has been reinterpreted since 1979 to allow pension funds to invest in VC-operating companies, which provided a fillip for the VC industry.

## **2.2 Stylized Facts About the VC Industry**

Venture capitalists provide funding to startup companies in exchange for a share of company equity. Apart from money, venture capitalists also provide mentorship services to foster the growth of startups. Since the life span of a VC fund is typically 10 years, venture capitalists are incentivized to target deals where a small amount of investment can generate a large financial return within a short period. Hence, VC investment tends to focus on high-growth companies in the high-tech sector. To illustrate, Figure 1 shows the share of VC investment received by each industry in 2016.<sup>2</sup> Attracting almost one-half of total VC investment, software companies are the top choice of VC investors. Pharmaceutical and biotech companies rank second, accounting for about one-eighth of total VC investment. Together, these two top industries comprise 60 percent of total VC investment. Other major industries receiving VC investment include healthcare devices and supplies (6 percent), healthcare services and systems (5 percent), commercial services (5 percent), IT hardware (4 percent), consumer goods and recreation (3 percent), energy (2 percent), and media (2 percent).<sup>3</sup> As the figure shows, venture capitalists are active investors in virtually all cutting-edge technologies.

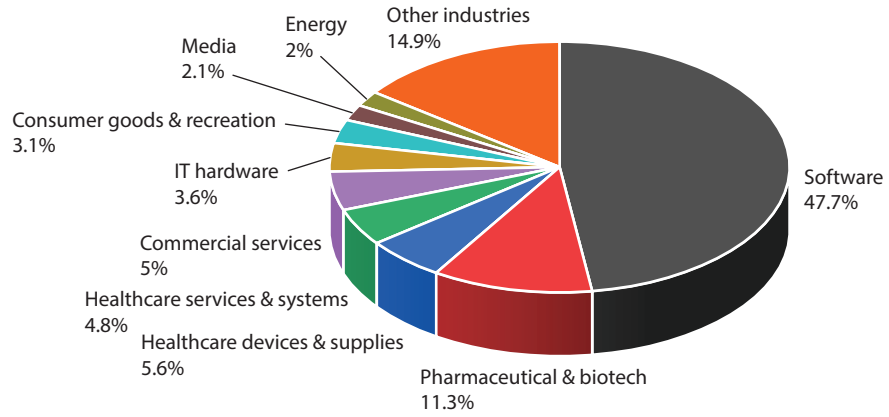
There are 898 VC firms in existence in the United States as of 2016.<sup>4</sup> These VC firms are managing 1,562 venture funds with a total amount of assets under management (AUM) of \$333 billion. The distribution of VC firms by AUM is shown in Figure 2. Many VC firms are rather small in terms of AUM. One-sixth of VC firms have an AUM of less than \$10 million, and the majority of them (92 percent) have an AUM below \$1 billion. In fact, only 68 VC firms (8 percent) have an AUM above \$1 billion. As revealed by Figure 2, VC is a fairly competitive industry populated with many small players and only a few large ones.

To track the evolution of the VC industry in the United States, Figure 3 plots the time series of VC investment. Numerous prominent VC firms were created in the 1970s, including the renowned industry leader Sequoia Capital, and Kleiner Perkins Caufield & Byers. In the 1980s, VC firms financed a series of successful companies, including Apple, Microsoft, and Cisco in the IT industry; Genentech in the biotech industry; and FedEx in the courier industry. Thanks to the “gold rush” in



**Figure 1**

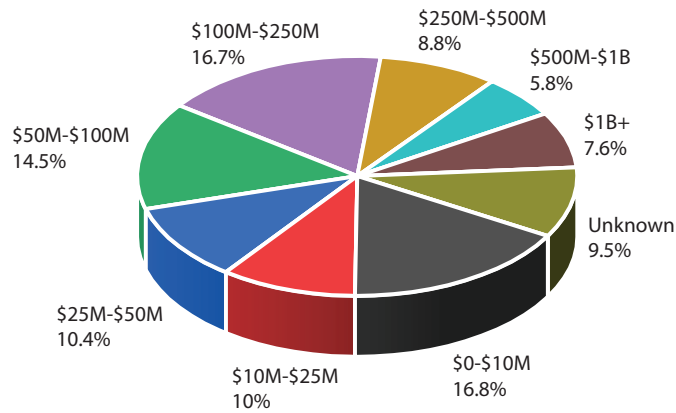
**Share of VC Investment Received By Each Industry**



SOURCE: NVCA (2016).

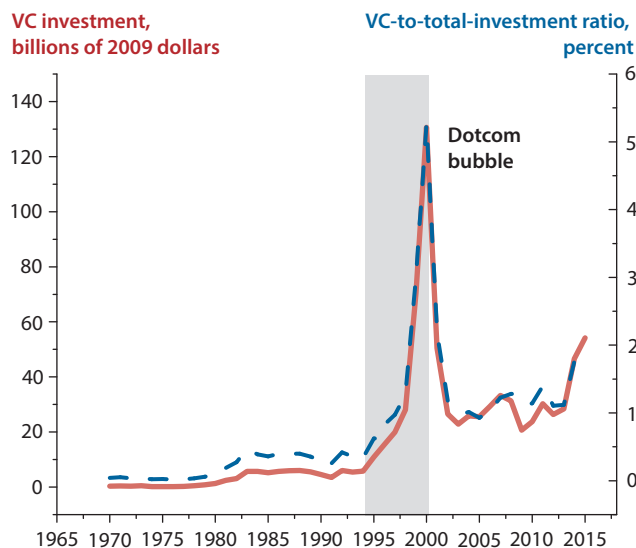
**Figure 2**

**Distribution of VC Firms by Assets Under Management**



NOTE: M, million; B, billion.

SOURCE: NVCA (2016).

**Figure 3****Investment by VC**

NOTE: The gray shaded area indicates recession as determined by the National Bureau of Economic Research.

SOURCE: Greenwood, Han, and Sánchez (forthcoming).

the internet sector during the dot.com bubble, the amount of VC investment soared during the 1990s. Many internet-related companies received VC financing during this era, including Amazon, eBay, Netscape, Sun Microsystems, and Yahoo!. The bursting of the dot.com bubble in 2000, however, triggered an unprecedented collapse of VC investment in the early 2000s. Nevertheless, the amount of VC investment in the post-bubble era was still well above its pre-bubble level and has returned to its long-run trend. Despite suffering from a decline during the Great Recession, the VC industry quickly recovered and continued to grow over time.

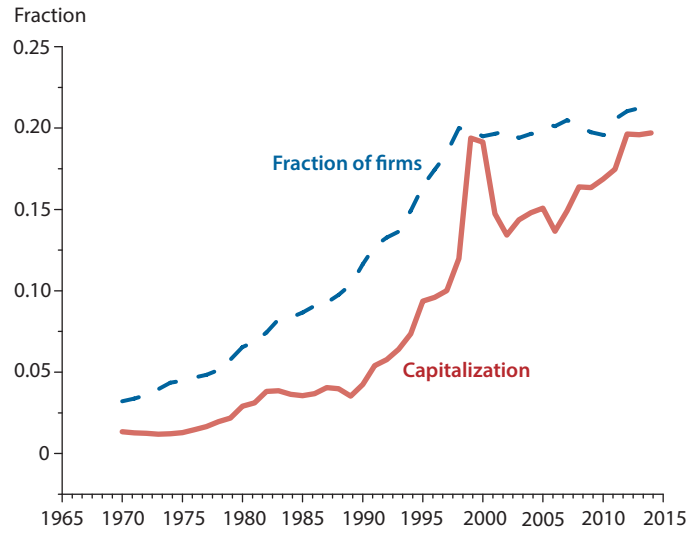
Though the amount of VC investment only accounts for a small share of aggregate U.S. investment, VC-backed companies (i.e., the ones financed by VC before going public) are playing an increasingly critical role in the aggregate economy, as demonstrated in Figures 4, 5, and 6. The fraction of VC-backed companies in all publicly traded firms is shown in Figure 4. The fraction of VC-backed companies in terms of market capitalization surged from 4 percent in 1970 to 20 percent by 2015.

Figure 5 reports the employment and R&D shares of VC-backed companies in all publicly traded firms. As indicated by the booming share of VC-backed companies, such companies are increasingly important for job creation and technological innovation. Analogously, Figure 6 displays the shares of patents and patents adjusted by the quality (proxied by citations) of VC-backed companies. It confirms the increasing importance of VC-backed companies for innovation.

The VC industry has exhibited remarkable resiliency despite the COVID-19 pandemic. The race for a COVID-19 vaccine has been a boon for startups in the pharmaceutical and biotech sector. Social distancing requirements have spurred VC investment in e-commerce, delivery, and work-

**Figure 4**

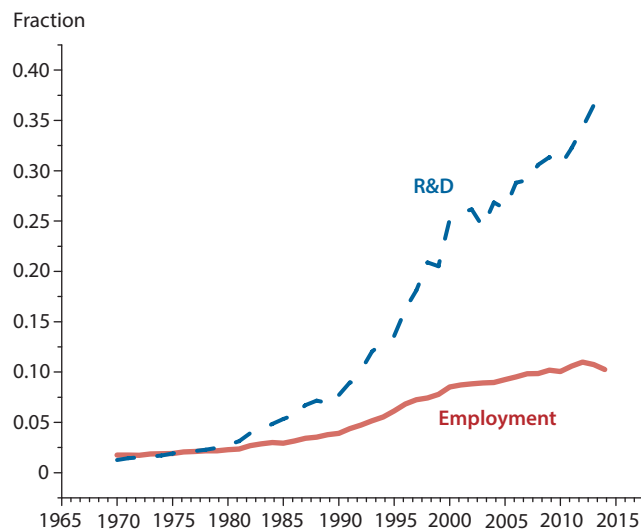
**Share of VC-Backed Companies Among Publicly Traded Firms**



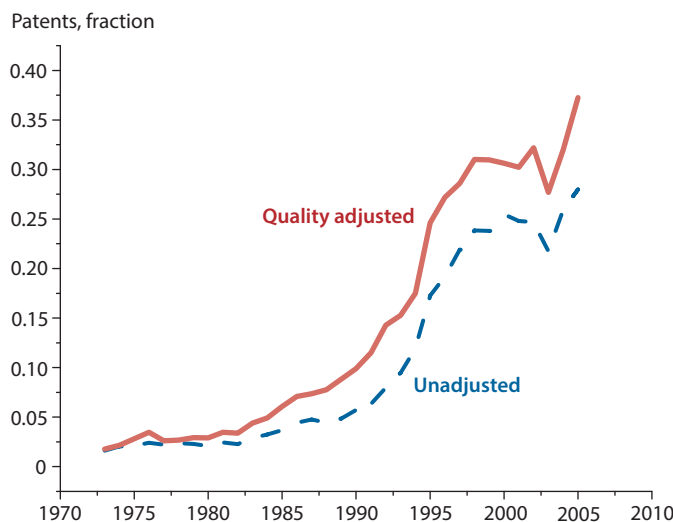
SOURCE: Greenwood, Han, and Sánchez (forthcoming).

**Figure 5**

**Shares of VC-Backed Public Companies in Employment and R&D Spending Among All Publicly Traded Firms**



SOURCE: Greenwood, Han, and Sánchez (forthcoming).

**Figure 6****Shares of VC-Backed Public Companies in Innovation Among All Publicly Traded Firms**

SOURCE: Greenwood, Han, and Sánchez (forthcoming).

from-home technologies. For instance, DoorDash (a VC-backed online food ordering and delivery company) succeeded in going public in December 2020, raising \$3.37 billion from the capital market.

### 3 THE IMPACT OF VC ON FIRM INNOVATION AND FIRM GROWTH

Building on the descriptive statistics in the last section, some regression evidence is presented in this section to disentangle the relationship between VC, firm growth, and innovation.

#### 3.1 VC and Firm Growth

Regression analysis is now conducted to evaluate the performance of VC-backed and non-VC-backed firms along four dimensions for the years following an initial public offering (IPO) of stock: the R&D-to-sales ratio, growth rate of employment, growth rate of sales, and firm market value (in natural logarithm). The results are presented in Table 1. The regressions are based on U.S. public companies between 1970 and 2014. To compare VC-backed companies with their non-VC-backed counterparts, a VC dummy is entered as an independent variable that takes the value of 1 if the company is funded by VC before its IPO. In all regressions, industry dummies, year dummies, and a year dummy for the IPO are included. In addition, a cross term is added between the VC dummy and the number of years since the firm's IPO.

As shown in the first row of Table 1, VC-backed companies are more R&D intensive and grow faster than their non-VC-backed counterparts. On average the R&D-to-sales ratio of a public VC-backed company is higher than its non-VC-backed counterpart by 5.2 percentage points, and it grows faster—by 4.9 percentage points in terms of employment and 7.0 percentage points in terms of sales. These superior performances translate into higher market values: VC-backed companies

**Table 1****VC-Backed vs. Non-VC-Backed Public Companies**

	R&D/Sales	Employment growth	Sales growth	ln(firm value)
VC-backed	0.0521*** (0.0017)	0.0490*** (0.0021)	0.0696*** (0.0027)	0.373*** (0.0141)
VC-backed × years since IPO	-0.000780*** (0.0001)	-0.00304*** (0.0002)	-0.00406*** (0.0002)	-0.0110*** (0.0011)
ln(employment)	-0.0133*** (0.0002)	-0.00567*** (0.0002)	-0.00641*** (0.0003)	0.851*** (0.0017)
Observations	84,116	148,834	149,672	168,549
$R^2$	0.383	0.084	0.108	0.737

NOTE: All specifications include year dummies, industry dummies (4-digit SIC codes), and a year dummy for the IPO. Standard errors are in parentheses. \*\*\* denotes significance at the 1 percent level.

are valued 37.3 percent higher than their non-VC-backed counterparts. The difference in performance, however, gradually dwindles over time, as shown by the negative estimates of the regression coefficients in the second row. As a consequence, the performances of VC- and non-VC-backed public companies tend to converge in the long run, though the speed of convergence is fairly low, as revealed by the magnitudes of the estimates in the second row.

### 3.2 VC and Innovation

The role of VC in encouraging technological innovation is now gauged at an annual periodicity; specifically, the impact of VC funding on patenting performance is evaluated at the firm level, and the impact of VC on employment and sales growth is assessed at the industry level. The data contains all companies funded by venture capitalists between 1970 and 2015. These VC-funded patentees are identified by matching firm names in VentureXpert and PatentsView.

**3.2.1 Firm-Level Regressions.** In the firm-level regressions, the primary independent variable is (the natural logarithm of) annual VC funding, while the dependent variable is a measure of patenting performance three years after the firm receives the funding. The primary independent variable may suffer from both measurement error and selection issues.<sup>5</sup> So, an instrumental variable (IV) is used in some of the regressions. The IV is based on the deregulation of pension funds since 1979, as highlighted in Section 2.1. The deregulation of pension funds reduced the fundraising costs of VC and led to increasing VC investment in all industries. In addition, industries that relied more on external finance enjoyed a stronger boost of VC funding.<sup>6</sup> Hence, a cross term between a “deregulation dummy” and a variable reflecting the industry’s (i.e., the industry in which the firm operates) dependence on external finance is introduced as an IV. The deregulation dummy takes the value of 1 after 1979. The dependence on external finance is a Rajan-Zingales-type measure (Rajan and Zingales, 1998) that reflects the extent to which outside funds are used in the industry for expendi-

**Table 2****VC Funding and Patenting****Panel A: Firm-level regressions, extensive margin analysis**

	1{Patent > 0}		1{"Breakthrough patent" > 0}	
	Probit (1)	IV (2)	Probit (3)	IV (4)
ln(firm VC funding)	0.126*** (0.0123)	0.610*** (0.0932)	0.125*** (0.0118)	0.525*** (0.123)
Observations	7,589	7,589	7,589	7,589

**Panel B: Firm-level regressions, intensive margin analysis**

	ln(patent)		ln(patent, quality adjusted)	
	OLS (5)	IV (6)	OLS (7)	IV (8)
ln(firm VC funding)	0.137*** (0.0107)	0.792*** (0.233)	0.182*** (0.0173)	0.748** (0.369)
Observations	5,538	5,538	4,958	4,958
$R^2$	0.244		0.135	

NOTE: See the main text for a description of the dependent and independent variables. The control variables are the number of patents held by the firm at the beginning of the year, the age of the firm, the total amount of privately and federally funded R&D of the industry in which the firm operates, and a cluster dummy for a firm headquartered in California or Massachusetts. All regressions include year fixed effects and industry fixed effects. Standard errors are in parentheses. \*\*\* denotes significance at the 1 percent level, \*\* at the 5 percent level.

tures on property, plant and equipment, R&D, advertising, and employee training. In all of the regressions, controls are added for the number of patents held by the firm at the beginning of the year, the age of the firm, and the total amount of privately and federally funded R&D of the industry in which the firm operates. Additionally, both a year dummy and an industry dummy (a 2-digit Standard Industrial Classification [SIC] code) are entered. Last, since both innovation and VC activities are remarkably clustered in California and Massachusetts, a “cluster dummy” for a firm headquartered in California or Massachusetts is included.

The results of the regression analysis are reported in Table 2. Panel A of Table 2 conducts the analysis along the extensive margin, that is, based on whether the firm obtains any patents three years after receiving VC funding. In regressions (1) and (2), the dependent variable is a dummy that takes the value of 1 if the firm files any successful patent applications at the U.S. Patents and Trademark Office within three years following funding. Regressions (3) and (4) focus on “breakthrough” patents, a measure pioneered by Kerr (2010). Breakthrough patents refer to those in the right tail of the citation distribution. Here the dependent variable in regressions (3) and (4) is a dummy variable that takes the value of 1 if the firm files any patents in the top 10 percent of the citation distribution in its cohort (i.e., patents with the same technological class and same application year) within three years following funding. Panel B of Table 2 turns to the intensive margin. In regressions (5) and (6), the dependent variable is (the natural logarithm of) the number of patents filed within three years following funding. The (natural logarithm of the) number of patents is

**Table 3****VC Funding and Industry Growth**

	Employment growth		Sales growth	
	OLS (1)	IV (2)	OLS (3)	IV (4)
ln(industry VC funding)	0.00338*** (0.000748)	0.00608*** (0.00178)	0.00495*** (0.000958)	0.00898*** (0.00228)
ln(employment)	-0.00646*** (0.00161)	-0.00817*** (0.00189)	-0.00476** (0.00207)	-0.00730*** (0.00243)
Observations	1,909	1,909	1,909	1,909
R <sup>2</sup>	0.285		0.334	

NOTE: See the main text for a description of the dependent and independent variables. The control variable is logged employment in each industry, and all regressions include year fixed effects and industry fixed effects. Standard errors are in parentheses. \*\*\* denotes significance at the 1 percent level, \*\* at the 5 percent level.

weighted by citations in regressions (7) and (8).

As shown by the positive estimates for VC funding in Panel A, larger VC funding increases the likelihood of a firm filing a patent. Larger funding also increases the likelihood of a firm coming up with a breakthrough patent, although the impact of VC funding is somewhat smaller in spurring breakthrough patents than ordinary patents. According to the IV estimates in regressions (6) and (8), a 10 percent increase in VC funding will induce in the three years subsequent to that funding a 7.9 percent boost in patenting and 7.5 percent boost in quality-adjusted patenting.

**3.2.2 Impact of VC on Industry Growth.** Attention is now turned to evaluating the impact of VC funding on growth at the industry level between 1970 and 2011. The main explanatory variable is the (natural logarithm of the) amount of VC funding each industry receives in each year. The dependent variables are the average annual growth rates of employment and sales for the three-year period after an industry receives VC funding.<sup>2</sup> In all the regressions, controls are added for logged employment in each industry, year dummies, and industry dummies (2-digit SIC codes). An IV is applied to address the issues of measurement errors and selection bias in the ordinary-least squares regressions. As detailed earlier, the IV is a cross term between the deregulation dummy and a variable reflecting the industry's dependence on external finance.

As demonstrated in Table 3, increasing VC funding in an industry in a given year is associated with a higher growth rate of employment and sales in the subsequent three years. According to IV regressions (2) and (4), a one-standard-deviation increase in logged industry-level VC funding is associated with increases of 1.3 percentage points and 1.9 percentage points in annual employment and sales growth, respectively, following funding.

## 4 CONCLUSION

The empirical evidence presented in this article suggests that VC financing is positively associated with firm innovation and growth in the United States. A structural model of VC is developed in Greenwood, Han, and Sánchez (forthcoming). The model is calibrated to fit stylized facts about VC in the United States. Through the lens of this model, the effects of VC and its taxation are examined. One of the crucial questions left unanswered is the role of VC financing in accounting for differences in development across countries. Although more research is needed in this area, the findings in Cole, Greenwood, and Sánchez (2016) and Greenwood, Han, and Sánchez (forthcoming) suggest that differences in the cost of enforcing contracts, the efficiency of financial intermediation in which VC plays a role, and the taxation of successful startups may be behind such cross-country differences. ■

## NOTES

- <sup>1</sup> This section draws *heavily* on Lamoreaux, Levenstein, and Sokoloff (2007) for the period prior to World War II and on Kenney (2011) for the period after.
- <sup>2</sup> The data source is the *2016 National Venture Capital Association Yearbook* (NVCA, 2016).
- <sup>3</sup> The category of “other industries” in this figure comprises the following industries: commercial products, commercial transportation, other business products and services, consumer durables, consumer non-durables services (non-financial), transportation, other consumer products and services, utilities, other energy, capital markets/institutions, commercial banks, insurance, other financial services, other healthcare, IT services, other information technology, agriculture, chemicals and gases, construction (non-wood), containers and packaging, forestry, metals, minerals and mining, textiles, and other materials. The data source is NVCA (2016).
- <sup>4</sup> The number of VC firms in existence is defined as a rolling count of firms that have raised a fund in the last eight years. The data source for all statistics in this paragraph is NVCA (2016).
- <sup>5</sup> The selection issue refers to the possibility that the positive relationship between VC investment and firm performance may be attributed to the ability of VCs to select promising companies to invest in.
- <sup>6</sup> This is revealed by the first-stage results of the IV regressions. The first-stage results are not presented due to space limitations.
- <sup>7</sup> The employment and sales information is based on the NBER-CES Manufacturing Industry Database available at <https://www.nber.org/nberces/>.

## REFERENCES

- Cole, Harold L. and Greenwood, Jeremy and Sánchez, Juan M. “Why Doesn’t Technology Flow from Rich to Poor Countries?” *Econometrica*, July 2016, 84(4), pp. 1477-521; <https://doi.org/10.3982/ECTA11150>.
- Greenwood, Jeremy; Han, Pengfei and Sánchez, Juan M. “Financing Ventures.” *International Economic Review*, forthcoming 2022.
- Kenney, Martin. “How Venture Capital Became a Component of the U.S. National System of Innovation.” *Industrial and Corporate Change*, 2011, 20(6), pp. 1677-723; <https://doi.org/10.1093/icc/dtr061>.
- Kerr, William R. “Breakthrough Inventions and Migrating Clusters of Innovation.” *Journal of Urban Economics*, January 2010, 67(1), pp. 46-60; <https://doi.org/10.1016/j.jue.2009.09.006>.
- Lamoreaux, Naomi; Levenstein, Margaret and Sokoloff, Kenneth L. “Financing Invention During the Second Industrial Revolution: Cleveland, Ohio, 1870-1920,” in Naomi Lamoreaux and Kenneth L. Sokoloff, eds., *Financing Innovation in the United States: 1870 to the Present*. MIT Press, 2007; <https://doi.org/10.7551/mitpress/9780262122894.003.0002>.
- National Venture Capital Association. *2016 National Venture Capital Association Yearbook*. Thomas Reuters, 2016.
- Rajan, Raghuram G. and Zingales, Luigi. “Financial Dependence and Growth.” *American Economic Review*, 1998, 88(3), pp. 559-86.



# On the Relative Performance of Inflation Forecasts

Julie K. Bennett and [Michael T. Owyang](#)

Inflation expectations constitute important components of macroeconomic models and monetary policy rules. We investigate the relative performance of consumer, professional, market-based, and model-based inflation forecasts. Consistent with the previous literature, professional forecasts most accurately predict one-year-ahead year-over-year inflation. Both consumers and professionals overestimate inflation over their respective sample periods. Market-based forecasts as measured by the swap market breakeven inflation rates significantly overestimate actual inflation; Treasury Inflation-Protected Securities market breakeven inflation rates exhibit no significant bias. We find that none of the forecasts can be considered rationalizable under symmetric loss. We also find that each forecast has predictive information that is not encompassed within that of another. (JEL E31, E37)

Federal Reserve Bank of St. Louis *Review*, Second Quarter 2022, 104(2), pp. 131-48.  
<https://doi.org/10.20955/r.104.131-48>

---

## 1 INTRODUCTION

Pandemic-related product and labor supply shortages have led to an uptick in inflation that represents some of the fastest price growth since the beginning of the Great Recession. This uptick has renewed interest in forecasts of future inflation. Moreover, in some macroeconomic models, expectations of inflation (or, alternatively, forecasts of inflation) can be nearly as important as realizations of inflation. For example, in models with a short-run Phillips curve, unemployment and expectations of inflation are assumed to have a negative relationship. In some monetary policy rules, the policymaker sets the interest rate, in part, as a function of the deviation between inflation expectations and the inflation target.

Long periods of relatively low and steady price growth had apparently made inflation forecasting easier, where simple random walk models' performance belied their computational ease (Atkeson and Ohanian, 2001, and Stock and Watson, 2007). Still, during that low-inflation period, professional forecasters seemingly still held an advantage over pure model-based forecasts (Faust and Wright, 2013). Here, we investigate the relative performance of consumer, professional, market-based, and model-based forecasts of inflation over a few different forecasting horizons.

Julie K. Bennett is a research associate and Michael T. Owyang is an assistant vice president and economist at the Federal Reserve Bank of St. Louis.

© 2022, Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

Macroeconomists use various measures of inflation expectations. In the absence of a sophisticated econometric model, consumers could construct forecasts based on information from outside sources with simple models, heuristics, personal experience, or some combination of these. Professional survey consensus forecasts such as the Blue Chip Economic Indicators (hereafter Blue Chip) are an amalgam of individual forecasts. Market-based forecasts are breakeven inflation rates (BEIs) derived from the Treasury Inflation-Protected Securities (TIPS) and inflation swap markets. We include a few simple model-based forecasts as a baseline for comparison.

How well do agents forecast inflation? That question lies at the root of a vast literature comparing consumer, professional, financial market, and model-based forecasts (Gramlich, 1983; Thomas, 1999; Mehra, 2002; Ang, Bekaert, and Wei, 2007; Gil-Alana, Moreno, and Pérez de Gracia, 2012; Wright, 2009; Faust and Wright, 2013; and Trehan, 2015, among others). Inflation forecasts are often evaluated for (i) accuracy, (ii) bias (whether the forecaster exhibits a tendency to overestimate or underestimate actual inflation), (iii) rationality (whether forecasts were made using relevant information known to the forecaster), and (iv) encompassing (whether one forecast has additional predictive information relative to another).

We evaluate forecasts for one-year-ahead year-over-year inflation (consumer, professional, and model-based) and five-year average inflation (market-based and model-based) with regard to these four metrics. We define five-year average inflation as the average year-over-year inflation rate over a five-year span, consistent with the inflation rate that five-year TIPS and swaps BEIs are interpreted to forecast. In terms of accuracy, professional forecasts most accurately predict actual one-year-ahead year-over-year inflation, while market-based forecasts and simple model-based forecasts perform similarly well in forecasting actual five-year average inflation. In terms of bias, both professional and consumer forecasts significantly overestimate one-year-ahead year-over-year inflation, while model-based forecasts do not exhibit any significant bias. Market-based forecasts of five-year average inflation as measured by the swaps BEI significantly overestimate inflation, while those measured by the TIPS BEI do not exhibit significant bias. This difference likely arises because of inflation and risk premia embedded in the rates. For rationality, none of the forecasts evaluated can be considered rational under the assumption of a symmetric loss function and an information set consisting of the unemployment rate (UR), federal funds rate (FFR), and lagged inflation rate available at the forecasting origin. Last, for encompassing, all forecasts have predictive information that is not fully contained within that of another forecast.

## 2 INFLATION FORECASTS

We consider a few examples of four types of forecasts: (i) consumer forecasts; (ii) professional forecasts; (iii) financial market implied forecasts; and (iv) econometric model-based forecasts, each described in more detail below. Consumer forecasts are taken from the University of Michigan Surveys of Consumers (Michigan Surveys). Professional forecasts are from the Blue Chip consensus forecasts. While forecasts from individual firms are available, these forecasts vary in their samples, have missing observations, and may not always be trying to minimize the squared forecast error. Thus, we consider only consensus-level forecasts. Survey of Professional Forecasters (SPF) forecasts are also commonly cited professional forecasts of inflation; however, we do not use these in our analysis, as SPF forecasts are published on a quarterly basis and thus more difficult to compare with monthly forecasts. Financial market implied forecasts are breakeven inflation rates computed from

the TIPS and inflation swap markets. We employ two simple types of model-based forecasts: (i) the Atkeson-Ohanian-type (hereafter AO) random walk forecast (Atkeson and Ohanian, 2001) and (ii) a simple vector autoregression- (VAR-) based forecast. We evaluate these forecasts in reference to inflation rates calculated using the seasonally adjusted consumer price index (CPI) from the Bureau of Labor Statistics (BLS).

## **2.1 Consumer Expectations**

To examine the performance and formation of consumer inflation forecasts, researchers often employ the consumer inflation expectations data series from the Michigan Surveys. These monthly surveys ask regular people about their outlook on current and future economic conditions. The results of the surveys help assess how the average consumer processes economic data in forming expectations of future outcomes. Questions in the surveys include the following:

- During the next 12 months, do you think that prices in general will go up, or go down, or stay where they are now?
- By about what percent do you expect prices to go (up/down) on the average, during the next 12 months?

Using responses to these questions, the Michigan Surveys constructs a data series of consumers' median expectations of year-over-year inflation.<sup>1</sup> This data series is available going back to January 1978 and is available in the Federal Reserve Bank of St. Louis FRED® database.<sup>2</sup>

When taken on their own, consumer inflation forecasts are often evaluated in the context of determining what factors influence the formation of inflation expectations. Ehrmann, Pfajar, and Santoro (2015) use the Michigan Surveys microdata from 1980 to 2011 and find that consumers with current or expected financial difficulties, pessimistic attitudes about major purchases, or expectations that income will go down tend to have a stronger upward bias compared with other households. Consumer inflation expectations have also been found to be responsive to media reporting (Carroll, 2003) and influenced by price changes in frequently purchased items such as gasoline (Coibion and Gorodnichenko, 2015).

## **2.2 Professional Forecasts**

When assessing professional inflation forecasts, researchers have used a variety of metrics, including forecasts from the SPF, Livingston Survey, and Blue Chip. The Blue Chip, for example, surveys some of America's top business economists each month and collects their forecasts of various macroeconomic indicators, including inflation. The Blue Chip publishes forecasts from each surveyed economist as well as an average—or consensus—of their forecasts for each variable. These forecasts are often cited by media outlets and used by corporate and government decisionmakers. We use the Blue Chip consensus forecast for one-year-ahead inflation, constructed by Haver Analytics, which is available going back to January 1986.<sup>3</sup>

Numerous earlier studies have evaluated the bias and rationality of professional inflation forecasts. Brown and Maital (1981) find that the average Livingston Survey forecasts of inflation from 1961 to 1977 were not biased, but that information on monetary growth was often underutilized in the forecasts. Figlewski and Watchel (1981) use individual Livingston Survey inflation forecasts from 1947 to 1975 and find that they exhibit a significant downward bias and that forecast errors are positively serially correlated. Keane and Runkle (1989, 1990) use individual price forecasts

from the SPF (at the time conducted jointly by the American Statistical Association and National Bureau of Economic Research) and conclude that the forecasts are rational, a contrast to most other studies evaluating professional inflation forecast rationality. Croushore (2010) evaluates the performance of the Livingston Survey and SPF forecasts from 1971 to 2006 in comparison to inflation as measured by the gross domestic product deflator (as opposed to CPI or personal consumption expenditure inflation used in other studies) and finds that there are only short periods where professional forecasters showed persistent and significant bias. Using individual forecasts from the Blue Chip from 1976 to 1986, Batchelor and Dua (1991) find that inflation and unemployment forecasts show the most cases of deviation from rationality, and real gross national product and interest rate forecasts the least; the authors posit, however, that it may not be optimal for individual commercial forecasters to change their forecasts to be rational, as they are attempting to differentiate their products from consensus forecasts.

### 2.3 Financial Market Forecasts

We use two measures of financial market inflation forecasts: BEIs as measured by the TIPS market and those as measured by the inflation swaps market.

Traders in the TIPS market make implicit forecasts of inflation rates based on the difference between the yield of a nominal bond and the yield of an inflation-linked bond of the same maturity. These forecasts are called *breakeven* inflation rates (BEIs), as an investor would receive the same yield on an inflation or non-inflation-indexed security if inflation averages that level over the course of the maturity. We focus on the five-year TIPS BEI, which is (i) calculated by subtracting the five-year Treasury Inflation-Indexed Constant Maturity Securities yield from the five-year Treasury Constant Maturity Securities yield and (ii) is often interpreted as what market participants expect the year-over-year inflation to be over the next five years, on average. We use monthly averages of daily observations for the five-year TIPS BEI; these data are available going back to January 2003 and are obtained from the Board of Governors of the Federal Reserve System and Haver Analytics.

In the inflation swaps market, two parties negotiate a contract under which inflation risk is transferred through an exchange of fixed cash flows. As both parties are trying to negotiate a fair price under the contract, the inflation swap rate can be seen as the expected BEI over the length of the contract. We focus on the five-year swaps BEI, which is often interpreted as what market participants expect the year-over-year inflation rate to be over the next five years, on average. We use monthly averages of daily observations for the five-year swaps BEI; these data are available going back to August 2004 and are obtained from Bloomberg.

Bauer (2014) finds that both TIPS and swaps BEIs are closely related to movements in the nominal interest rates and are sensitive to macroeconomic data surprises. The TIPS and swaps BEIs are not necessarily straightforward forecasts of inflation, though, as they reflect market participants' inflation expectations as well as inflation and liquidity risk premia. A number of articles decompose the TIPS and swaps BEIs into these components using different methods and find that both the TIPS and swaps BEIs embed nontrivial inflation and liquidity risk premia (Gurkaynak, Sack, and Wright, 2010; Zeng, 2013; Abrahams et al., 2016; D'Amico, Kim, and Wei, 2018; Haubrich, Pennacchi, and Ritchken, 2012; and Casiraghi and Miccoli, 2019; for a full review, see Kupfer, 2018).

Zeng (2013) finds that TIPS BEIs tend to underestimate inflation due to the liquidity risk and that model-implied inflation expectations at various horizons outperform the SPF and Michigan

Surveys forecasts. For TIPS BEIs, Abrahams et al. (2016) find that the estimated inflation risk premium is highly correlated with observable macroeconomic and financial variables, such as disagreement about future inflation among professional forecasters. They also find that the BEIs adjusted for risk and liquidity premia outperform unadjusted BEIs and a random walk in predicting realized inflation in sample and out of sample. Casiraghi and Miccoli (2019) find that swaps BEIs incorporate sizable inflation risk premia and that these premia increase with maturity. They also find that the inflation risk premium was positive before the financial crisis, became negative at the end of 2008 after the bankruptcy of Lehman Brothers, and recovered shortly afterward.

For simplicity, we use BEIs unadjusted for liquidity and inflation risk premia as the financial market forecasts in our analysis.

## 2.4 Model-Based Forecasts

As a baseline for comparison, we use two model-based forecasts: (i) the AO random walk forecast (Atkeson and Ohanian, 2001) and (ii) a VAR-based forecast.

The AO forecast uses the inflation rate over the previous four quarters as the forecast for what inflation will be over the next four quarters. To fit with our other monthly forecast series, we adapt the AO forecast using the inflation rate over the previous 12 months as the forecast for inflation over the next 12 months.<sup>4</sup> We use a similar model to forecast the five-year average inflation rate to compare with the five-year BEI. In this case, the average year-over-year inflation rate over the previous five years is used to forecast the average year-over-year inflation rate over the next five years.

We also produce simple VAR-based forecasts for both one-year-ahead inflation and five-year average inflation. We estimate three different VAR forecasts for one-year-ahead CPI inflation: (i) a VAR including lags of year-over-year CPI inflation and lags of year-over-year industrial production (IP) growth; (ii) a VAR including lags of year-over-year inflation and lags of the UR; and (iii) a VAR including lags of year-over-year inflation, lags of IP growth, and lags of the UR.<sup>5</sup> We use the same three VARs to forecast five-year average inflation using lags of five-year average inflation rather than lags of year-over-year inflation.

For the one-year-ahead inflation model-based forecasts, the first forecast origin is March 1960, corresponding to the first available vintage of UR data that can be used to estimate the VAR. For the five-year average inflation model-based forecasts, the first forecast origin is June 1966, corresponding to the first forecast origin whose VAR model can be estimated with at least 100 observations given the available data.

## 3 FORECAST EVALUATION

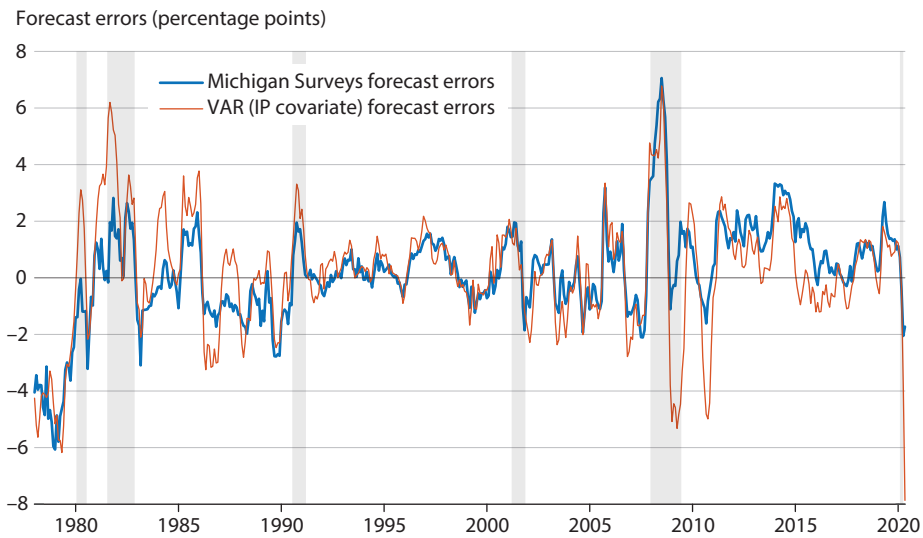
Figure 1 shows forecast errors corresponding to a few of the forecasts we discuss.<sup>6</sup> Panel A shows forecast errors for one-year-ahead year-over-year inflation forecasts. Panel B shows forecast errors for five-year average inflation forecasts. For each panel, the date on the  $x$ -axis corresponds to the date the forecast was made. The forecast errors (and, thus, the forecasts themselves) generally move together, with the one-year-ahead year-over-year inflation forecasts tracking each other more closely than the five-year average inflation forecasts.

Some notation will facilitate discussion of the properties of the forecasts. At a forecast origin  $t$ , suppose the forecaster has information  $X_t$  to construct the  $h$ -period-ahead forecast, denoted  $Y_{t+h|t}$ . Each forecast produces an error relative to the realization,  $Y_{t+h}$ , defined as

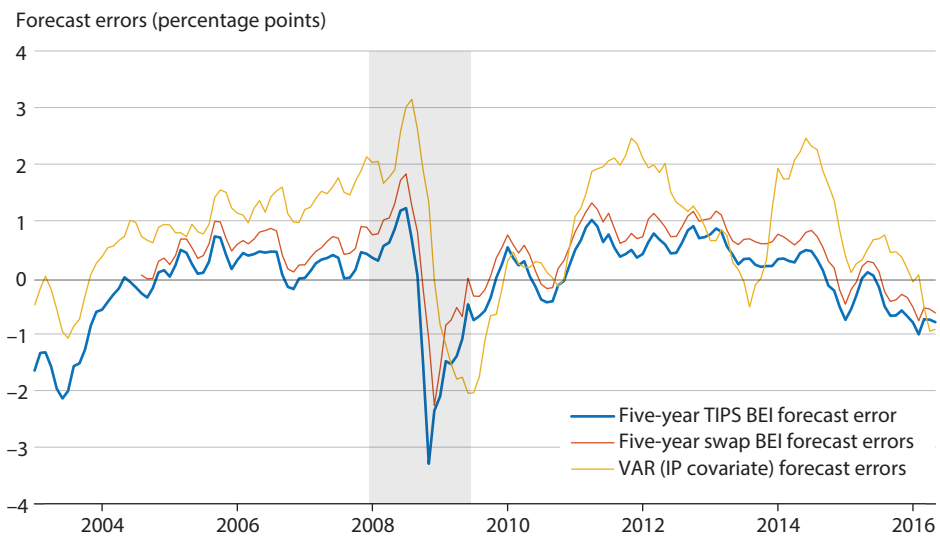
**Figure 1**

**Inflation Forecast Errors**

**A. One-year-ahead year-over-year inflation forecast errors**



**B. Five-year average inflation forecast errors**



NOTE: This figure shows the forecast errors for both one-year-ahead year-over-year inflation forecast metrics (Panel A) and five-year average inflation forecast metrics (Panel B). The date on the x-axis corresponds to the forecast origin. Gray shaded regions indicate NBER recessions.

$$(1) \quad \epsilon_{t+h} = Y_{t+h} - Y_{t+h|t}.$$

We can evaluate these forecasts using a number of measures described in detail below.

### 3.1 Accuracy

One way to evaluate forecasts is to examine their relative accuracy. We first compute the root mean squared error (RMSE) as

$$(2) \quad RMSE = \sqrt{\frac{1}{P} \sum_{t=1}^P \epsilon_{t+h}^2},$$

where  $P$  is the number of forecasts being evaluated. Squaring the error focuses on the magnitude of the forecast's deviation from the realized value rather than on whether the forecast is underestimated or overestimated. We compare the relative accuracy of two forecasts by computing the ratio of one forecast's RMSE to another's. A relative RMSE less (greater) than 1 indicates that the forecast represented in the numerator (denominator) is relatively more accurate over the sample period.

While the relative RMSEs provide some suggestion of which of the two forecasts is more accurate, statistical tests are required to determine whether we can assert that one forecast is significantly "better" than another. These tests (e.g., Diebold and Mariano, 1995) are typically formulated with the null of equal predictive ability and require assumptions about the data generating process and how the forecasts were constructed. We forgo formal statistical testing of equal predictive ability here and refer the reader to the literature.

### 3.2 Bias

A forecast is considered biased if the average forecast error over the sample period is statistically different from zero. Thus, we compute the average forecast error over all available forecast origins for each inflation forecast metric and test whether this average is statistically different from zero using a standard  $z$ -statistic.

Forecasts might be biased for a number of reasons. A forecaster may prefer to underestimate the variable they are forecasting (rather than overestimate it). For example, say a firm is forecasting demand for its product and there is a high carrying cost for extra inventory. In this case, the forecaster might prefer to underestimate demand, leading to underproduction so that no goods need to be stored in inventory.

### 3.3 Rationality

Keane and Runkle (1989) define rational (or rationalizable) forecasts as those whose forecast errors are unpredictable given what the forecaster knew at the time they made the forecast. Rationality is evaluated conditional on both a loss function and a fixed information set. The loss function implicit in Keane and Runkle (1989) is quadratic loss, where the loss is an increasing function of the forecast error, regardless of the direction of the error. The information set consists of potentially relevant variables known to the forecaster at the forecast origin. If not taken into account (i.e., incorporated into the forecast), a variable could explain the realization but not the forecast, making the forecast error a function of the variable.

We can test the rationalizability of inflation forecasts by determining whether the forecast errors can be predicted by factors known to the forecaster at the forecast origin,  $t$ . To do this, we regress the forecast error on a constant, the forecaster's prediction ( $Y_{t+h|t}$ ), and a set of variable(s),  $X_t$ , in the forecaster's information set:

$$(3) \quad Y_{t+h} - Y_{t+h|t} = \alpha_0 + \alpha_1 Y_{t+h|t} + \alpha_2 X_t + u_{t+h}.$$

For the forecasts to be rationalizable given information  $X_t$ , the values of  $\alpha_1$  and  $\alpha_2$  should not be significantly different from zero; otherwise, the forecast error could be predicted by  $X_t$ . Thus, we consider a forecast rationalizable if we cannot reject the null hypothesis that  $\alpha_1$  and  $\alpha_2$  are jointly zero using an  $F$ -test.<sup>7</sup>

We test rationality in this way for each inflation forecast series, using all available forecasting origins for each series. The information set,  $X_t$ , for each rationality test consists of past data that are common in macroeconomic models of the inflation rate: the UR, effective FFR, and most recent value of inflation (either year-over-year or five-year average inflation).<sup>8</sup> The composition of  $X_t$  varies slightly across rationality tests for different forecasts. The tests for the VAR forecasts and the AO forecasts do not include the most recent value of inflation in  $X_t$ , as that lagged inflation value is inherently part of the forecast for both of those models. The VAR forecasts that include the UR as a covariate also do not include the UR in  $X_t$ , as that value is incorporated into the forecast for those models.

### 3.4 Encompassing

Comparing the relative RMSEs of two forecasts may determine which is more accurate on its own. However, a substantial literature has shown that combining forecasts can lead to improved accuracy. Tests for forecast encompassing determine whether one forecast has additional predictive information relative to another. Another way of interpreting this statement asks whether one forecast would have any weight in a forecast combination with the other. For example, previous studies have examined the hypothesis that the Fed's forecasts encompass those made by the private sector.

To conduct an encompassing test, we follow similar methodology to that used by Fair and Shiller (1989) and Romer and Romer (2000). We regress the realized values,  $Y_{t+h}$ , on a constant and the two forecasts,  $Y_{1,t+h|t}$  (Forecast 1) and  $Y_{2,t+h|t}$  (Forecast 2):

$$(4) \quad Y_{t+h} = \alpha + \beta_1 Y_{1,t+h|t} + \beta_2 Y_{2,t+h|t} + \eta_{t+h},$$

and we test the null hypothesis that  $(\alpha, \beta_1, \beta_2)' = (0, 1, 0)'$ . The null assumes that Forecast 2 has no significant predictive information not already contained in Forecast 1. Previous studies have found that forecasts of the inflation rate and output growth rates are biased (see Caunedo et al., 2020); thus, a test could reject this joint null hypothesis because  $\alpha \neq 0$ . Therefore, we also conduct the encompassing test, omitting the unbiasedness condition, and test only the null  $(\beta_1, \beta_2)' = (1, 0)'$ .

We conduct encompassing tests for all pairs of inflation forecast metrics that are forecasting the same inflation metric (i.e., one-year-ahead year-over-year inflation or five-year average inflation), using a Wald test to test the null hypotheses posited in the preceding paragraph. The encompassing tests use data from the forecasting origins that are available for both Forecasts 1 and 2.



**Table 1****Relative RMSEs****A. One-year-ahead inflation**

	Michigan Surveys	Blue Chip	AO	VAR (IP covariate)	VAR (UR covariate)	VAR (IP and UR covariates)
Michigan Surveys	1.000	1.342	0.869	0.824	0.771	0.780
Blue Chip		1.000	0.697	0.634	0.590	0.606
AO			1.000	0.908	0.859	0.847
VAR (IP covariate)				1.000	0.946	0.933
VAR (UR covariate)					1.000	0.986
VAR (IP and UR covariates)						1.000

**B. Five-year average inflation**

	Five-year TIPS BEI	Five-year swaps BEI	AO	VAR (IP covariate)	VAR (UR covariate)	VAR (IP and UR covariates)
Five-year TIPS BEI	1.000	0.908	1.027	0.577	0.608	0.662
Five-year swaps BEI		1.000	0.980	0.533	0.563	0.619
AO			1.000	0.630	0.597	0.610
VAR (IP covariate)				1.000	0.948	0.969
VAR (UR covariate)					1.000	1.022
VAR (IP and UR covariates)						1.000

NOTE: This table displays the RMSEs for each combination of relevant forecasts. Each relative RMSE is the RMSE of the forecast indicated in the row divided by the RMSE of the forecast indicated in the column. For each entry, the two RMSEs are calculated using the date range of whichever forecast of the two has the smallest available date range. A relative RMSE < 1 (> 1) indicates that the row forecast is more (less) accurate than the column forecast over the evaluated date range. A relative RMSE > 1 indicates that the column forecast is more accurate than the row forecast over the evaluated date range.

**4 RESULTS****4.1 Accuracy**

Table 1 compares the forecasts' accuracy, reporting the RMSE of the forecast indicated in the row relative to the RMSE of the forecast indicated in the column. For each entry, the two RMSEs are calculated using the common date range of the two forecasts. A relative RMSE < 1 (> 1) indicates that the row forecast is more (less) accurate than the column forecast over the evaluated date range.

Panel A of Table 1 presents results for one-year-ahead inflation forecasts. In terms of accuracy, the Blue Chip outperforms all other evaluated metrics. The Michigan Surveys forecasts outperform all metrics except those for the Blue Chip. The AO model does better than any of the VARs, and the VARs all perform similarly well. The result that survey-based forecasts outperform simple time-series models is consistent with previous findings (Thomas, 1999; Ang, Bekaert, and Wei, 2007; and Gil-Alana, Moreno, and Pérez de Gracia, 2012).

Panel B of Table 1 presents results for five-year average inflation forecasts. The five-year TIPS BEI outperforms the five-year swaps BEI and the VARs, but the AO model narrowly outperforms

**Table 2****Forecast Bias and Rationality Results**

	Bias		Rationality			
			Restrict $\alpha = 0$		Allow $\alpha \neq 0$	
	Mean error	<i>p</i> -Value	<i>F</i> -statistic	<i>p</i> -Value	<i>F</i> -statistic	<i>p</i> -Value
<b>A. One-year-ahead inflation</b>						
Michigan Surveys	-0.187	0.013	15.993	0.000	18.239	0.000
Blue Chip	-0.130	0.016	7.754	0.000	8.140	0.000
AO	0.005	0.950	29.717	0.000	39.621	0.000
VAR (IP covariate)	0.247	0.003	25.848	0.000	31.043	0.000
VAR (UR covariate)	-0.038	0.660	49.379	0.000	73.952	0.000
VAR (IP and UR covariates)	0.098	0.268	55.444	0.000	82.414	0.000
<b>B. Five-year average inflation</b>						
Five-year TIPS BEI	0.056	0.352	53.860	0.000	66.752	0.000
Five-year swaps BEI	-0.424	0.000	157.654	0.000	121.366	0.000
AO	-0.147	0.100	102.665	0.000	135.380	0.000
VAR (IP covariate)	-0.490	0.000	247.684	0.000	319.729	0.000
VAR (UR covariate)	-0.351	0.018	362.431	0.000	535.812	0.000
VAR (IP and UR covariates)	-0.179	0.221	351.027	0.000	524.483	0.000

NOTE: This table presents results for tests of bias and rationality for each of the forecast metrics, separated by whether the metric is forecasting one-year-ahead inflation or five-year average inflation. The two leftmost columns for the rationality results report the *F*-statistics and associated *p*-values corresponding to testing the joint null hypothesis that  $(\alpha_0, \alpha_1, \alpha_2) = (0, 0, 0)$ . The two rightmost columns for the rationality results report the *F*-statistics and associated *p*-values corresponding to testing the joint null hypothesis that  $(\alpha_1, \alpha_2) = (0, 0)$ .

the TIPS BEI. The five-year swaps BEI outperforms all metrics except for the five-year TIPS BEI. The TIPS BEI outperforming the swaps BEI, the swaps BEI outperforming the AO model, and the AO model outperforming the TIPS BEI each arise because of the differences in the timespans of data available for the TIPS BEI and the swaps BEI. The AO model does better than any of the VARs, and the VARs all perform similarly well.

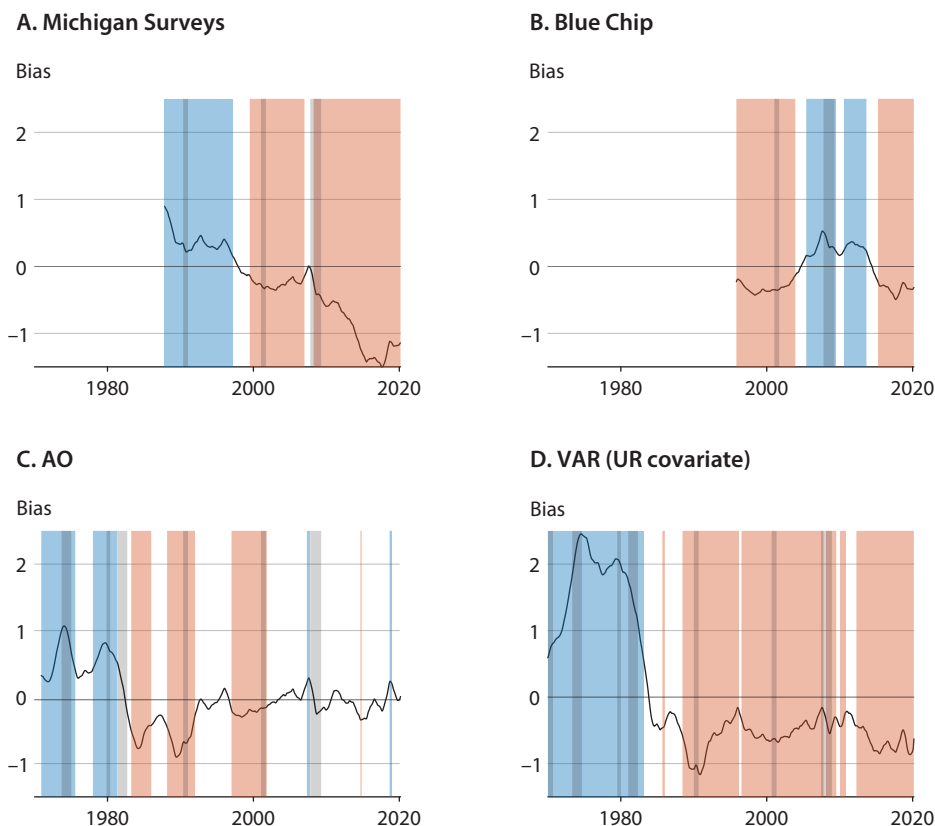
#### 4.2 Bias

The first two columns of Table 2 present the results of evaluating forecast bias over the entirety of each forecast's sample period. A positive (negative) bias value indicates that the forecast systematically underpredicts (overpredicts) actual inflation. For one-year-ahead inflation metrics, both the Michigan Surveys forecasts and the Blue Chip forecasts significantly overestimate actual inflation. Meanwhile, the AO model is not significantly biased. Results for the VARs vary, though they typically do not show significant bias.

The five-year TIPS BEI does not exhibit any significant bias in forecasting actual inflation, while the five-year swaps BEI significantly overestimates it. While they are both financial market forecasts, the distinction may arise from the differences in liquidity and inflation risk premia in the two markets.

**Figure 2**

**Ten-Year Rolling Bias for One-Year-Ahead Inflation Forecasts**



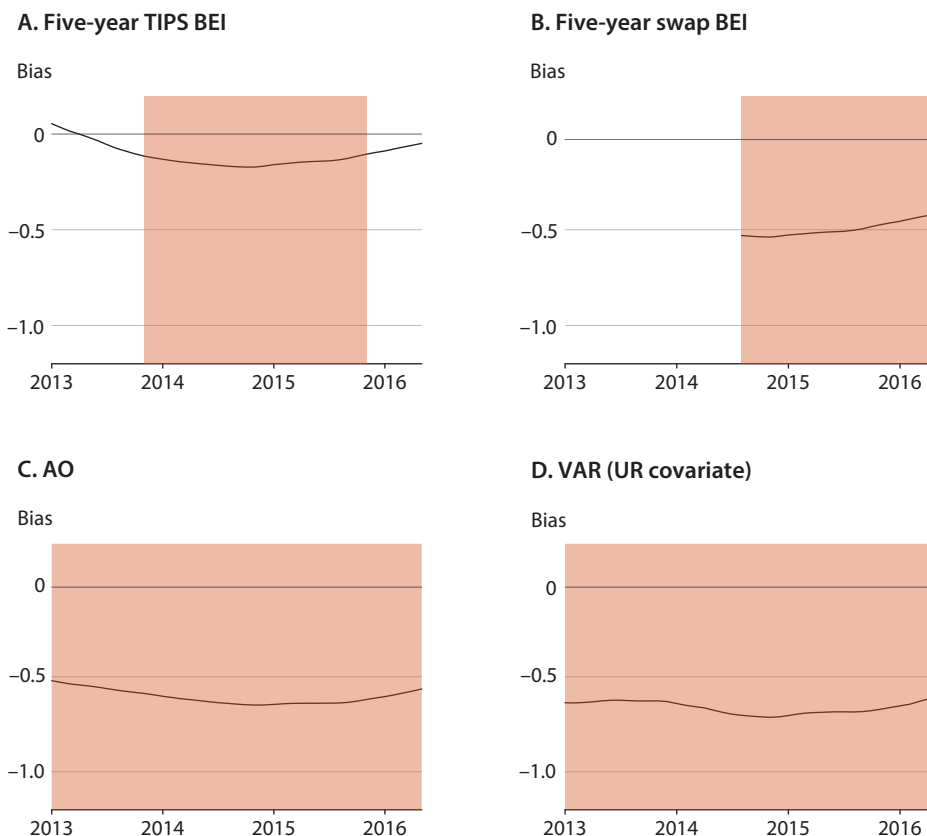
NOTE: The figure shows the average bias of a given one-year-ahead inflation forecast for the 10-year period that ends on the date indicated by the *x*-axis. Orange shaded regions indicate subsamples in which the forecast bias is negative (overestimate) and statistically significant at the 10 percent level. Blue shaded regions indicate subsamples in which the forecast bias is positive (underestimate) and statistically significant at the 10 percent level. Gray shaded regions indicate NBER recessions.

The five-year swaps BEI incorporates an inflation risk premium that is, on average, positive, resulting in overestimation of actual inflation (Casiraghi and Miccoli, 2019). On the other hand, previous research has argued that the TIPS BEI is subject to countercyclical risk premia (Abrahams et al., 2016, and Andreasen, Christensen, and Riddell, 2020). Because these premia are time varying, this variance leads to the overestimation of actual inflation at some points and underestimation at others. Results for the VARs vary.

Previous work suggests that forecast bias results may be sensitive to the time period chosen for evaluation (Croushore, 2010). Therefore, we also look at each forecast’s bias over a 10-year rolling window. Figures 2 and 3 display these results for one-year-ahead year-over-year inflation forecasts and five-year average inflation forecasts, respectively. Each line in Figures 2 and 3 shows the average bias of a given forecast for the 10-year period that ends on the date indicated by the *x*-axis. Orange shaded regions indicate subsamples in which the forecast bias is an overestimate and statistically significant at the 10 percent level. Blue shaded regions indicate subsamples in which the forecast

**Figure 3**

**Ten-Year Rolling Bias for Five-Year Average Inflation Forecasts**



NOTE: This figure shows the average bias of a given five-year average inflation forecast for the 10-year period that ends on the date indicated by the x-axis. Orange shaded regions indicate subsamples in which the forecast bias is negative (overestimate) and statistically significant at the 10 percent level.

bias is an underestimate and statistically significant at the 10 percent level. Gray shaded regions indicate National Bureau of Economic Research (NBER) recessions.

For one-year-ahead inflation forecasts, the sign and significance of the forecast bias varies across the sample periods. For example, at the beginning of the Michigan Surveys forecast sample period, consumers tended to underestimate inflation; however, beginning in the mid-1990s, they tend to overestimate it. Consumers’ forecast bias is significant across a majority of the 10-year rolling windows, though there are two short periods in which it is not. On the other hand, Blue Chip forecasters overestimated inflation in the beginning of their sample, underestimated it during the mid-2000s to mid-2010s, and more recently have overestimated it. Similar to that of consumers, professionals’ forecast bias is significant across a majority of the 10-year rolling windows, though there are three short periods in which it is not.

For five-year average inflation forecasts, the 10-year rolling-window forecasts overestimate actual inflation, save for a short window at the beginning of the TIPS BEI sample period that is insignificant. Looking at the time period comparable to the BEI forecasts (which is displayed in

Figure 3), the swaps BEI, AO model, and VAR significantly overestimate actual inflation over all 10-year windows, while the TIPS BEI significantly overestimates inflation only for a specific period of windows (window with end date November 2013 to window with end date November 2015). The AO model and VAR model, however, do have periods preceding 2013 in which they do not exhibit significant bias.

### 4.3 Rationality

The four rightmost columns in Table 2 presents the results of the forecast rationality tests described in Section 3.3. Columns 3 and 4 report the  $F$ -statistics and associated  $p$ -values corresponding to testing the joint null hypothesis that  $(\alpha_0, \alpha_1, \alpha_2) = (0, 0, 0)$  (i.e., restricting the forecast to be unbiased); Columns 5 and 6 report the  $F$ -statistics and associated  $p$ -values corresponding to testing the joint null hypothesis that  $(\alpha_1, \alpha_2) = (0, 0)$  (i.e., allowing the forecast to be biased). In either case, failure to reject the null indicates that the forecast is rationalizable. Rejection of the null indicates that the forecast error can be predicted by relevant macroeconomic variables known to the forecaster at the forecast origin; thus, the forecaster did not incorporate all relevant, known information into their forecast.

Based on our results, we reject rationality for all forecasts, regardless of whether the forecast is allowed to be biased. These results fall in line with previous work that rejects rationalizability for consumer, professional, financial market, and model-based forecasts (Brown and Maital, 1981; Batchelor and Dua, 1991; and Thomas 1999, among others).

No particular economic indicator appears unaccounted for across forecasts. Instead, the rejection of the null hypothesis appears to be driven by different factors across types of forecasts. For example, consumers appear to appropriately incorporate the most recent value of inflation into their forecasts, but leave information from the UR and FFR on the table. Meanwhile, five-year TIPS BEIs appropriately incorporate information from the FFR into their forecasts but leave information from the UR and most recent value of inflation underutilized. Professional forecasters underutilize information from all considered regressors. These results suggest that agents make inflation forecasts based on different information sets.

These results on rationalizability are predicated on the selected information set and assumed loss function (quadratic). Previous work finds that forecasts, particularly professional forecasts, are often rationalizable under the assumption of an asymmetric loss function (Elliot, Komunjer, and Timmerman, 2008; and Capistran and Timmerman, 2009; among others).

### 4.4 Encompassing

Table 3 presents the results of the encompassing tests for one-year-ahead inflation forecast metrics. Panel A displays the  $p$ -values associated with testing the joint null hypothesis  $(\alpha, \beta_1, \beta_2) = (0, 1, 0)$  based on the regression described in Section 3.4, and Panel B displays the  $p$ -values associated with testing the joint null hypothesis  $(\beta_1, \beta_2) = (1, 0)$ . In either case, a failure to reject the null indicates that the predictive information contained in Forecast 1 encompasses that which is contained in Forecast 2; a rejection of the null indicates that Forecast 2 contains predictive information that is not contained in Forecast 1.

In each of the pairwise combinations evaluated in Table 3—regardless of whether or not the forecasts are allowed to be biased—the null hypothesis is rejected at the 10 percent significance level.

**Table 3****One-Year Ahead Inflation Encompassing Results**

		Forecast 2					
		Michigan Surveys	Blue Chip	AO	VAR (IP covariate)	VAR (UR covariate)	VAR (IP and UR covariates)
<b>A. Restrict <math>\alpha = 0</math></b>							
Forecast 1	Michigan Surveys		0.000	0.000	0.000	0.000	0.000
	Blue Chip	0.000		0.002	0.000	0.000	0.000
	AO	0.000	0.000		0.000	0.000	0.000
	VAR (IP covariate)	0.000	0.000	0.000		0.000	0.000
	VAR (UR covariate)	0.000	0.000	0.000	0.000		0.000
	VAR (IP and UR covariates)	0.000	0.000	0.000	0.000	0.000	
<b>B. Allow <math>\alpha \neq 0</math></b>							
Forecast 1	Michigan Surveys		0.000	0.000	0.000	0.000	0.000
	Blue Chip	0.000		0.013	0.000	0.000	0.000
	AO	0.000	0.000		0.000	0.000	0.000
	VAR (IP covariate)	0.000	0.000	0.000		0.000	0.000
	VAR (UR covariate)	0.000	0.000	0.000	0.000		0.000
	VAR (IP and UR covariates)	0.000	0.000	0.000	0.000	0.000	

NOTE: This table presents results of encompassing tests that test whether Forecast 1 encompasses Forecast 2 for a given row/column pair. Panel A presents results for encompassing tests where the regression constant is tested to be zero in the joint null hypothesis  $(\alpha, \beta_1, \beta_2)' = (0, 1, 0)'$ . Panel B presents results for encompassing tests where the regression constant is not tested under the joint null hypothesis,  $(\beta_1, \beta_2)' = (1, 0)'$ . The values reported in both panels are the  $p$ -values associated with the  $F$ -statistic testing the relevant joint null hypothesis.

Thus, each of the forecasts has predictive content that is not fully contained within that of another. When the forecast is allowed to be biased (Panel B), we could conclude that Blue Chip forecasts encompass the AO model forecasts if evaluating the results using a 1 percent significance level.

Because professional forecasters likely take the most recent value of inflation into consideration, Blue Chip forecasts nearly encompass the AO model forecasts. One might expect Blue Chip forecasts to nearly encompass the Michigan Surveys forecasts, assuming that professional forecasters have more and better information than the average consumer. That consumer forecasts have predictive information not contained in professional forecasts, however, is consistent with previous findings that households and professionals focus on different factors when making inflation forecasts (Berge, 2018, and Palardy and Ovaska, 2015). Moreover, consumer forecasts have outperformed professional forecasts over some sample periods (Gramlich, 1983, and Mehra, 2002).

Table 4 presents the results of the encompassing tests for five-year average inflation metrics. Analogous to Table 3, Panel A displays the  $p$ -values associated with restricting Forecast 1 to be unbiased, and Panel B displays the  $p$ -values associated with allowing Forecast 1 to be biased. As before, the results in Table 4 indicate that, regardless of whether Forecast 1 is allowed to be biased, the null hypothesis is rejected at the 10 percent significance level; thus, each of the forecasts has a distinct information set that is not fully contained within that of another.

**Table 4****Five-Year Average Inflation Encompassing Results**

		Forecast 2					
		Five-year TIPS BEI	Five-year swaps BEI	AO	VAR (IP covariate)	VAR (UR covariate)	VAR (IP and UR covariates)
<b>A. Restrict <math>\alpha = 0</math></b>							
Forecast 1	Five-year TIPS BEI		0.000	0.000	0.000	0.000	0.000
	Five-year swaps BEI	0.000		0.000	0.000	0.000	0.000
	AO	0.000	0.000		0.000	0.000	0.000
	VAR (IP covariate)	0.000	0.000	0.000		0.000	0.000
	VAR (UR covariate)	0.000	0.000	0.000	0.000		0.000
	VAR (IP and UR covariates)	0.000	0.000	0.000	0.000	0.000	
<b>B. Allow <math>\alpha \neq 0</math></b>							
Forecast 1	Five-year TIPS BEI		0.000	0.000	0.000	0.000	0.000
	Five-year swaps BEI	0.000		0.000	0.000	0.000	0.000
	AO	0.000	0.000		0.000	0.000	0.000
	VAR (IP covariate)	0.000	0.000	0.000		0.000	0.000
	VAR (UR covariate)	0.000	0.000	0.000	0.000		0.000
	VAR (IP and UR covariates)	0.000	0.000	0.000	0.000	0.000	

NOTE: This table presents results of encompassing tests that test whether Forecast 1 encompasses Forecast 2 for a given row/column pair. Panel A presents results for encompassing tests where the regression constant is tested to be zero in the joint null hypothesis  $(\alpha, \beta_1, \beta_2)' = (0, 1, 0)'$ . Panel B presents results for encompassing tests where the regression constant is not tested under the joint null hypothesis,  $(\beta_1, \beta_2)' = (1, 0)'$ . The values reported in both panels are the  $p$ -values associated with the  $F$ -statistic testing the relevant joint null hypothesis.

**4.5 Interpretation**

Based on these results and the results in previous studies, no single forecast (or forecaster) appears to dominate any other. What does this imply for the comparison of the consumer and professional survey forecasts, both relative to each other and relative to the model-based forecasts?

First, professional forecasters are generally better than consumers and simple models at predicting inflation. This finding aligns with what one might expect, as professionals may have information and resources to make more informed forecasts than the average consumer.

Second, consumers are not bad at predicting inflation. Although they are not generally as accurate as professional forecasters, they outperform simple econometric models. This finding suggests that the average consumer understands inflation trends. Energy prices and food prices are often correlated with headline inflation (energy and headline correlation: 0.68; food and headline correlation: 0.71).<sup>9</sup> Even if consumers rely on heuristics that overemphasize energy and food prices, they may predict the direction of price growth through casual observation of prices in their everyday experiences.

Third, the BEIs cannot be interpreted as straightforward forecasts of inflation. These rates are extracted from financial market instruments, and they embed time-varying liquidity and inflation

risk premia that mask market participants' actual inflation expectations. Taken at face value as inflation forecasts, they may be misleading.

Fourth, there appears to be exploitable predictive information in all forecasts. Moreover, it appears that consumers, forecasting professionals, and financial market analysts consider different information sets when constructing forecasts. As none of the forecasts evaluated here are rationalizable under the assumption of a symmetric loss function, it may be relevant to consider whether agents form inflation forecasts based on asymmetric preferences.

## 5 CONCLUSION

We evaluate the performance of consumer, professional, market-based, and model-based inflation forecasts with regard to accuracy, bias, rationality, and encompassing. Consistent with previous studies, survey-based forecasts are more accurate than model-based forecasts. Consumer, professional, and swap market forecasts all tend to overestimate inflation on average, but the sign of the bias varies across the sample period for consumer, professional, and TIPS market forecasts. Results of rationality tests indicate that none of the forecasts can be considered rationalizable under the assumption of symmetric loss, and economic agents use different information sets when forming their inflation forecasts. Encompassing tests indicate that each forecast contains predictive information that is not encompassed by another forecast. Future work could consider additional survey-based inflation forecast metrics, such as the consumer inflation expectations series from the Federal Reserve Bank of New York, as well as market-based forecasts adjusted for inflation and risk premia. Further tests of equal predictive ability or rationality under asymmetric loss could also provide additional information on the relative performance of various inflation forecasts. ■

## NOTES

- <sup>1</sup> More information regarding the survey questionnaire and the construction of the consumers' median inflation expectation series can be found at <https://data.sca.isr.umich.edu/>.
- <sup>2</sup> The Federal Reserve Bank of New York also produces consumer inflation expectations series at one-year and three-year horizons. These data are only available for a relatively short sample (from 2012 to the present); thus, we omit them from our analysis. These series could be used in future work.
- <sup>3</sup> This series is constructed using the Blue Chip one-year-ahead consumer price index (CPI) consensus forecast four quarters ahead as the monthly value. The quarterly series are aggregated from the monthly series. For example, the one-year-ahead inflation forecast value made in January 2020 is constructed using the CPI forecast for 2021:Q1 in the January 2020 survey. The one-year-ahead inflation forecast value made in February 2020 is constructed using the CPI forecast for 2021:Q1 in the February 2020 survey.
- <sup>4</sup> For example, at the forecasting origin January 2000, one would use the year-over-year value of inflation measured in December 1999 as the forecast for year-over-year inflation measured in January 2001, as the December 1999 value would be the most recent year-over-year inflation value available to the forecaster in January 2000.
- <sup>5</sup> The seasonally adjusted CPI from the Bureau of Labor Statistics (BLS) is used to calculate inflation rates; the UR data are produced by the BLS; and the IP data are produced by the Board of Governors of the Federal Reserve System. We use vintages of both IP and UR data obtained from the Federal Reserve Bank of St. Louis ALFRED® database: <https://alfred.stlouisfed.org/>. These series are subject to frequent and substantial revisions. We do not use CPI vintages to calculate year-over-year inflation rates, as CPI revisions are typically insubstantial.
- <sup>6</sup> Blue Chip forecasts are proprietary and cannot be depicted here.



- <sup>7</sup> Rationality tests often include  $\alpha_0 \neq 0$  in the null. In this case, if the forecasts are biased, rationality is rejected. We include results for tests that include or exclude  $\alpha_0 \neq 0$  in the null.
- <sup>8</sup> The information at month  $t$  depends on when the forecast is made within a given month. The University of Michigan typically conducts the Surveys of Consumers in the last week of  $t-1$  through approximately the third week of month  $t$ . We assume the most recent value of the UR, the FFR, and inflation available to consumers is  $t-2$ . Blue Chip surveys are conducted in the first week of  $t$ . We assume the most recent value of the UR and inflation available to professionals is  $t-2$  and the most recent value of the FFR is  $t-1$ . Both the TIPS and swap market BEIs are monthly averages of daily rates. The information set available to a market forecaster on the first day of the month consists of UR and inflation values from  $t-2$  and the FFR from  $t-1$ . The model-based forecasts all use  $t-1$  lagged inflation in their forecast calculation. Those forecasts are made on the CPI release date of  $t$ ; thus, all regressors are from  $t-1$ .
- <sup>9</sup> These correlations are calculated using CPI inflation data from January 1978 to May 2021, the dates corresponding with the Michigan Surveys inflation forecasts.

## REFERENCES

- Abrahams, Michael; Adrian, Tobias; Crump, Richard K.; Moench, Emanuel and Yu, Rui. "Decomposing Real and Nominal Yield Curves." *Journal of Monetary Economics*, December 2016, 84, pp. 182-200; <https://doi.org/10.1016/j.jmoneco.2016.10.006>.
- Andreasen, Martin M.; Christensen, Jens H.E. and Riddell, Simon. "The TIPS Liquidity Premium." Working Paper 2017-11, Federal Reserve Bank of San Francisco, 2020; <https://doi.org/10.24148/wp2017-11>.
- Ang, Andrew; Bekaert, Geert and Wei, Min. "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?" *Journal of Monetary Economics*, May 2007, 54(4), pp. 1163-212; <https://doi.org/10.1016/j.jmoneco.2006.04.006>.
- Atkeson, Andrew and Ohanian, Lee E. "Are Phillips Curves Useful for Forecasting Inflation?" Federal Reserve Bank of Minneapolis *Quarterly Review*, Winter 2001, 25(1), pp. 2-11; <https://doi.org/10.21034/qv.2511>.
- Batchelor, Roy and Dua, Pami. "Blue Chip Rationality Tests." *Journal of Money, Credit and Banking*, 1991, 23(4), 1991, pp. 692-705; <https://doi.org/10.2307/1992704>.
- Bauer, Michael. "Inflation Expectations and the News." *International Journal of Central Banking*, 2014; <http://dx.doi.org/10.2139/ssrn.2417046>.
- Berge, Travis J. "Understanding Survey-Based Inflation Expectations." *International Journal of Forecasting*, 2018, 34(4), pp. 788-801; <https://doi.org/10.1016/j.ijforecast.2018.07.003>.
- Brown, Bryan W. and Maital, Shlomo. "What Do Economists Know? An Empirical Study of Experts' Expectations." *Econometrica*, March 1981, 49(2), pp. 491-504; <https://doi.org/10.2307/1913322>.
- Capistrán, Carlos and Timmermann, Allan. "Disagreement and Biases in Inflation Expectations." *Journal of Money, Credit, and Banking*, 2009, 41(2-3), pp. 365-96; <https://doi.org/10.1111/j.1538-4616.2009.00209.x>.
- Carroll, Christopher D. "Macroeconomic Expectations of Households and Professional Forecasters." *Quarterly Journal of Economics*, February 2003, 118(1), pp. 269-98; <https://doi.org/10.1162/00335530360535207>.
- Casiraghi, Marco and Miccoli, Marcello. "Inflation Risk Premia and Risk-Adjusted Expectations of Inflation." *Economics Letters*, February 2019, 175, pp. 36-39; <https://doi.org/10.1016/j.econlet.2018.12.002>.
- Caunedo, Julieta; DiCecio, Riccardo; Komunjer, Ivana and Owyang, Michael T. "Asymmetry, Complementarities, and State Dependence in Federal Reserve Forecasts." *Journal of Money, Credit, and Banking*, 2020, 52(1), pp. 205-28; <https://doi.org/10.1111/jmcb.12590>.
- Coibion, Olivier and Gorodnichenko, Yuriy. "Is the Phillips Curve Alive and Well After All? Inflation Expectations and the Missing Disinflation." *American Economic Journal: Macroeconomics*, 2015, 7(1), pp. 197-232; <https://doi.org/10.1257/mac.20130306>.
- Croushore, Dean. "An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data." *B.E. Journal of Macroeconomics*, 2010, 10(1); <https://doi.org/10.2202/1935-1690.1677>.
- D'Amico, Stefania; Kim, Don H. and Wei, Min. "Tips from TIPS: The Informational Content of Treasury Inflation-Protected Security Prices." *Journal of Financial and Quantitative Analysis*, February 2018, 53(1), pp. 395-436; <https://doi.org/10.1017/S0022109017000916>.

- Diebold, Francis X. and Mariano, Roberto S. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, 1995, 13(3), pp. 253-63; <https://doi.org/10.2307/1392185>.
- Ehrmann, Michael; Pfajfar, Damjan and Santoro, Emiliano. "Consumers' Attitudes and Their Inflation Expectations." FEDS Working Paper No. 2015-015, Board of Governors of the Federal Reserve System, March 2015; <http://dx.doi.org/10.17016/FEDS.2015.015>.
- Elliott, Graham; Komunjer, Ivana and Timmermann, Allan. "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?" *Journal of the European Economic Association*, 2008, 6(1), pp. 122-57; <https://doi.org/10.1162/JEEA.2008.6.1.122>.
- Fair, Ray C. and Shiller, Robert J. "The Informational Content of Ex Ante Forecasts." *Review of Economics and Statistics*, May 1989, 71(2), pp. 325-31; <https://doi.org/10.2307/1926979>.
- Faust, Jon and Wright, Jonathan H. "Forecasting Inflation," in Graham Elliott and Allan Timmermann, eds., *Handbook of Economic Forecasting*. 2013, Volume 2, pp. 2-56; <https://doi.org/10.1016%2Fb978-0-444-53683-9.00001-3>.
- Figlewski, Stephen and Watchel, Paul. "The Formation of Inflationary Expectations." *Review of Economics and Statistics*, February 1981, 63(1), pp. 1-10; <https://doi.org/10.2307/1924211>.
- Gil-Alana, Luis; Moreno, Antonio and Pérez de Gracia, Fernando. "Exploring Survey-Based Inflation Forecasts." *Journal of Forecasting*, September 2012, 31(6), pp. 524-39; <https://doi.org/10.1002/for.1235>.
- Gramlich, Edward M. "Models of Inflation Expectations Formation: A Comparison of Household and Economist Forecasts." *Journal of Money, Credit and Banking*, May 1983, 15(2), pp. 155-73; <https://doi.org/10.2307/1992397>.
- Gürkaynak, Refet S.; Sack, Brian and Wright, Jonathan H. "The TIPS Yield Curve and Inflation Compensation." *American Economic Journal: Macroeconomics*, 2010, 2(1), pp. 70-92; <https://doi.org/10.1257%2Fmac.2.1.70>.
- Haubrich, Joseph; Pennacchi, George, and Ritchken, Peter. "Inflation Expectations, Real Rates and Risk Premia: Evidence from Inflation Swaps." *Review of Financial Studies*, May 2012, 25(5), pp. 1588-629; <https://doi.org/10.1093/rfs/hhs003>.
- Keane, Michael P. and Runkle, David E. "Are Economic Forecasts Rational?" *Federal Reserve Bank of Minneapolis Quarterly Review*, Spring 1989, 13(2), pp. 26-33; <https://doi.org/10.21034/qv.1323>.
- Keane, Michael P. and Runkle, David E. "Testing the Rationality of Price Forecasts: New Evidence from Panel Data." *American Economic Review*, September 1990, 80(4), pp. 714-35; <https://www.jstor.org/stable/2006704>.
- Kupfer, Alexander. "Estimating Inflation Risk Premia Using Inflation-Linked Bonds: A Review." *Journal of Economic Surveys*, 2018, 32(5), pp. 1326-54; <https://doi.org/10.1111%2Fjoes.12265>.
- Mehra, Yash P. "Survey Measures of Expected Inflation: Revisiting the Issues of Predictive Content and Rationality." *Federal Reserve Bank of Richmond Economic Quarterly*, Summer 2002, 88(3), pp. 17-36; [https://www.richmondfed.org/publications/research/economic\\_quarterly/2002/summer/mehra](https://www.richmondfed.org/publications/research/economic_quarterly/2002/summer/mehra).
- Palardy, Joseph and Ovaska, Tomi. "Decomposing Household, Professional and Market Forecasts on Inflation: A Dynamic Factor Model Analysis." *Applied Economics*, 2015, 47(20), pp. 2092-101; <https://doi.org/10.1080/00036846.2014.1002889>.
- Romer, Christina D. and Romer, David H. "Federal Reserve information and the Behavior of Interest Rates." *American Economic Review*, 2000, 90(3), pp. 429-57; <https://doi.org/10.1257/aer.90.3.429>.
- Stock, James H. and Watson, Mark W. "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit and Banking*, 2007, 39, pp. 3-33; <https://doi.org/10.1111%2Fj.1538-4616.2007.00014.x>.
- Thomas, Lloyd, B. "Survey Measures of Expected U.S. Inflation." *Journal of Economic Perspectives*, 1999, 13(4), pp. 125-44; <https://doi.org/10.1257/jep.13.4.125>.
- Trehan, Bharat. "Survey Measures of Expected Inflation and the Inflation Process." *Journal of Money, Credit, and Banking*, 2015, 47(1), pp. 207-22; <https://doi.org/10.1111/jmcb.12174>.
- Wright, Jonathan H. "Forecasting U.S. Inflation by Bayesian Model Averaging." *Journal of Forecasting*, 2009, 28(2), pp. 131-44; <https://doi.org/10.1002%2Ffor.1088>.
- Zeng, Zheng. "New Tips from TIPS: Identifying Inflation Expectations and the Risk Premia of Breakeven Inflation." *Quarterly Review of Economics and Finance*, May 2013, 53(2), pp. 125-39; <https://doi.org/10.1016/j.qref.2013.02.005>.