BOCCONI ECONOMIC JOURNAL

#1
2025

IN THIS ISSUE

Egea

BOCCONI ECONOMIC JOURNAL

Progetto grafico della copertina: Cristina Bernasconi – Milano
Impaginazione: Compomat - Configni (RI)
e-mail: egea.edizioni@unibocconi.it – www.egeaeditore.it
ISBN 978-88-238-6342-2

BOCCONI ECONOMIC JOURNAL

#1
2025

IN THIS ISSUE

Egea

Table of Contents

LETTER FROM THE EDITOR

Roberto Fani

DOI: 10.82029/2025011

Dear Readers,

It is with great pleasure and enthusiasm that I welcome you to the inaugural issue of the Bocconi Economic Journal (BEJ), the student-edited economics journal of the Ettore Bocconi Department of Economics.

The Bocconi Economic Journal was founded in 2024 as a student initiative with the ambition of providing an accessible yet rigorous platform for students' research in economics and related fields. Establishing such an outlet was a challenging endeavor. First, our editorial team had to learn from the ground up what it truly means to edit a manuscript. We take pride in the substantial improvement achieved between initial submissions and the final published versions. Second, this initiative had no precedent in Continental Europe and represents one of the rare cases in which a student-led journal benefits from the invaluable support of faculty members during the final review stage prior to publication. Finally, designing a scientifically robust yet efficient review process required considerable effort. We are satisfied with the outcome, while remaining fully aware of how much work lies ahead.

This achievement was made possible through the generous support of the Department of Economics, which hosts the initiative, provides financial and scientific backing, and offers continuous guidance. We owe special gratitude to the Scientific Director, Professor Jérôme Adda, for his invaluable advice, his sharing of best editorial practices, and his tireless coordination of the final phase of the editorial process, which ensures the faculty's scientific approval of all manuscripts. We are deeply grateful to the members of the faculty who have devoted their time and expertise to validating the editorial work carried out by our student team. Being hosted by a department at the frontier of economic research provides extraordinary learning opportunities and makes us acutely aware of the opportunity cost of the resources allocated to us, a fact that renders our gratitude even deeper.

I also wish to acknowledge the diligent and committed work of our co-editors and associate editors, whose rigorous handling of manuscripts has set a high standard of quality that we are determined to uphold throughout the life of BEJ. We also thank the four external, anonymous referees who kindly contributed their expertise. Our publisher, Egea, has been instrumental in bringing BEJ to publication. Lastly, we extend our sincere thanks to all authors who entrusted BEJ with their research; your confidence in our work is profoundly appreciated.

On November 6, 2024, we held our Inaugural Event in Aula Franceschi on a crisp Milanese evening. For the occasion, we had the immense honor of hosting Nobel Laureate Professor Michael Spence, who delivered an inspiring lecture titled "The Global Economy in the Next Decade." The event also provided an opportunity to reflect on the objective function that a student-led journal should seek to maximize, an intellectually stimulating question that has guided our editorial philosophy.

Economics is a social science. While it relies on rigorous mathematical formalization and sophisticated analytical tools, its ultimate concern lies with social phenomena. It is therefore unsurprising that the most compelling research questions often arise from direct observation of the world around us, and that the most pivotal insights are sometimes those that appear disarmingly simple—so much so that one wonders how they had not been formulated earlier. This guiding principle has informed every decision made by our editorial team. As we bring our modest contribution to the academic debate through this inaugural issue, we are conscious of the extraordinary privilege of studying economics within a university and department at the frontier of research, an environment that not only advances our understanding of the world but also shapes policy and generates real-world impact. Above all, we are continually inspired by the excellence and intellectual vitality of our academic community.

Beyond its purely editorial dimension, we see BEJ as playing a broader role within the Bocconi academic community. We value its open-access, online format, which makes it a small public good—intended not only as a modest contribution to the scientific community, but also as a source of motivation for younger students who aspire to further studies and a career in economics, by illustrating possible paths toward research. We view the journal as having a pedagogical purpose as well, offering students direct exposure to the process of peer review, one of the fundamental building blocks of modern science. In this sense, BEJ is not only an editorial initiative, but also an epistemological exercise: an opportunity to engage with the principles that underpin how scientific knowledge is produced, evaluated, and refined.

This issue features four manuscripts spanning topics from labor economics to decision theory, from fiscal macroeconomics to economic growth. We thoroughly enjoyed engaging with these contributions and are proud to present them in our inaugural volume.

Our lead article, Luigi Lorenzoni's "What Caused U.S. Post-Pandemic Inflation? A Fiscal Extension of Bernanke and Blanchard (2025)," integrates fiscal policy into the recent inflation debate, demonstrating that accommodating fiscal expansions can explain a significant share of the price dynamics previously attributed to supply-side shocks. Gabriele Nespoli's "Regional Minimum Wages and Labor Misallocation in Italy" employs firm-level microdata to estimate optimal national and region-specific minimum wages, showing that regional differentiation could reduce spatial distortions in employment and better align pay with productivity. Matteo Pozzi's "Certainty Crises, Ambiguity Aversion and Self-Confirming Equilibrium" develops a formal framework linking decision theory and game theory to explore how doubt and perceived ambiguity influence strategic behavior, revealing that heightened awareness of uncertainty may trap agents in inefficient equilibria even under rational expectations. Finally, Sharmista Lakkineni's "Innovations and Work: Assessing the Impact of Automation on Labor Outcomes Through a Cross-Country and Cross-Industry Analysis" revisits the labor market consequences of technological change, extending Autor and Salomons' framework to recent decades and uncovering how novel forms of automation may weaken the compensatory mechanisms that once preserved aggregate employment.

Looking ahead, we recognize that much remains to be accomplished. Yet with this first issue, we are proud to present a selection of some of the finest students' research produced at our alma mater. To future authors who may consider BEJ as an outlet for their work, we extend our sincere thanks in advance and reaffirm our commitment to a fair, transparent, and rigorous review process.

Thank you for your attention, and we hope you enjoy the read!

WHAT CAUSED U.S. POSTPANDEMIC INFLATION? A FISCAL EXTENSION OF BERNANKE AND BLANCHARD (2025)

Luigi Lorenzoni[†]

DOI: 10.82029/2025012

Abstract. In this work, I make two main contributions. Firstly, I propose an extension of the empirical model introduced by Bernanke and Blanchard (2025) to account for fiscal policy. Determining endogenously the state of supply chains I show that its pure contribution to inflation is small and that it mediates a demand component attributable to accommodating fiscal policy. Through an historical decomposition I estimate this fiscal component to sum to 2.4 percentage points of price inflation over the period 2020Q1:2025Q1. I interpret this result within a standard aggregate supply-aggregate demand (AS AD) framework with supply constraints. Secondly, I provide an R replication package for all the results of Bernanke and Blanchard (2025). I integrated the possibility to derive confidence intervals for the responses of price inflation implied by the standard model through bootstrapping. I show that these are generally not statistically different from zero.

JEL classification: C30, C50, E12, E31, E62,

Keywords: Inflation, Fiscal Policy, Supply Chains

1. INTRODUCTION

What were the drivers of recent inflation? Was it supply or demand or both? Did fiscal policy play a role? Some recent works, chiefly Bernanke and Blanchard (2025), stress the importance of supply side explanations pointing to relative commodity price shocks as the main driver of inflation. These works treat commodity prices as exogenous and, while acknowledging the role of demand forces, including them in their frameworks remains an open question. Some other works, such as Giannone and Primiceri (2024), point instead to demand forces as the main driver of inflation, showing that the contribution of supply shocks has a demand component that can be identified when these supply factors become endogenous. In this work, I pick up the open issue of Bernanke and Blanchard (2025) (henceforth BB) and I show that in their framework the contribution of shortages–a variable proxied through Google Trend that captures the saliency of supply constraints–is not to be interpreted entirely as supply-side. Endogenously determining shortages I am able to capture its fiscal-driven component, which leads me to estimate a dynamic effect of accommodating fiscal policy of 2.4 percentage points of price inflation. Indeed, the shortages component is directly displaced by the fiscal one. I interpret this result through the lenses of a standard Keynesian mechanism, much in the spirit of Fornaro (2024). In an economy sensitive to supply constraints, fiscal policy stimulates aggregate demand and takes the economy to regions where these constraints bind, that is where the supply curve is steeper–even vertical–ultimately resulting in inflation. This happens because when firms try to increase production beyond their supply constraints, marginal costs increase and so do prices. Production under binding supply constraints is akin to an increase in markup, where inflation spikes with constant wages. This interpretation offers also a suitable solution to the disconnect between price and wage inflation, which raises some puzzles in a classical

[†]Lorenzoni: luigi.lorenzoni@studbocconi.it. I thank Carlo Favero, Francesco Giavazzi, Luisa Pomarici, Paolo Baudino, Anita Sammarini, the editor, and two anonymous referees for helpful comments and discussions.

New-Keynesian framework such as Galí (2015). These demand forces thus make their way into inflation not through labor markets–whose contribution to inflation I estimate being small–but rather through supply chains. This gives a richer picture of inflation. It generally stems from a scarcity on the supply side relative to demand but, since there are both labor and non-labor inputs, the push on prices may well come mainly from non-labor inputs. Nevertheless, although it is commonly intuitive not to think so, this is still a form of demand-driven inflation. I call this mechanism–drawing from Lorenzoni and Werning (2023)–a "supply-constrained demand shock," which well captures the dynamics of the story I outline and highlights how the canonical intuitive dichotomy between demand and supply driven inflation is likely more nuanced. There may be other mechanisms at work, such as those referring to the Fiscal Theory of the Price Level (FTPL), which I do not incorporate in my model.

In estimating these results, I developed a smooth R replication package for Bernanke and Blanchard (2025), which is available on GitHub. [1] I also implemented a version of the paper with confidence intervals to test robustness. I find that, in general, the response to various shocks of inflation is not statistically different from zero and that the parameters are quite robust to changes in sample periods.

This paper is structured as follows: in Section 2, I briefly trace an outline of the events I analyze; in Section 3, I review the recent literature on the issue; in Sections 4 and 5, I delve into the technical details of my empirical approach, examining the model, data, and specifications I employ; in Section 6, I show my replication results; and lastly, in Section 7, I present an extension of the BB model and provide its empirical implications.

2. HISTORICAL BACKGROUND

The aim of this brief section is to quickly trace an outline of the events of the last 4 years concerning the pandemic, inflation, and the associated policy responses. In February 2020, the COVID-19 pandemic broke out, requiring strict lockdown measures that led to extreme consequences, including an all-time-high unemployment rate of 14% and falling economic activity. In the attempt to balance these worrying statistics and ensure appropriate response, the Trump administration passed the CARES Act, the largest fiscal stimulus in American history, amounting to US$2.2 trillion. It included, among others, immediate paychecks and increased unemployment benefits. This stimulus was followed by other unusual measures: the Consolidated Appropriation Act and the American Rescue Plan Act (ARPA), but also the Inflation Reduction Act (IRA). The fiscal effort observed in the last 4 years is unprecedented in recent American history, as shown in Figure 1. Outlays and deficit spiked up on impulse around the burst of the pandemic but are now also plateauing at an all-time-high level.

As pandemic pressures eased, the inflation rate started to increase steadily, along with headline components such as energy and food. The pressure on supply chain, which can be proxied through Google Trend, experienced a spike as well. The inflation rate gained momentum reaching its peak at 9% in June 2022 and then returned on a descending trend toward the 2% target. In the summer of 2022, monetary policy, both in the US and in the EA, started to respond with a rapid tightening, bringing the policy rates quickly to levels well above the zero-lower-bound where they have been stuck for a decade. Central banks have now started to cut them.

3. LITERATURE REVIEW

I aim to dialogue with two main strands of literature. The first one is a recent series of works accounting for postpandemic inflation, both theoretically and empirically. The other is the strand of literature investigating fiscal sources of inflation.

[1] https://github.com/luigilor/bernanke-blanchard23

Figure 1: Outlays as a percentage of GDP, by category (source: CBO)

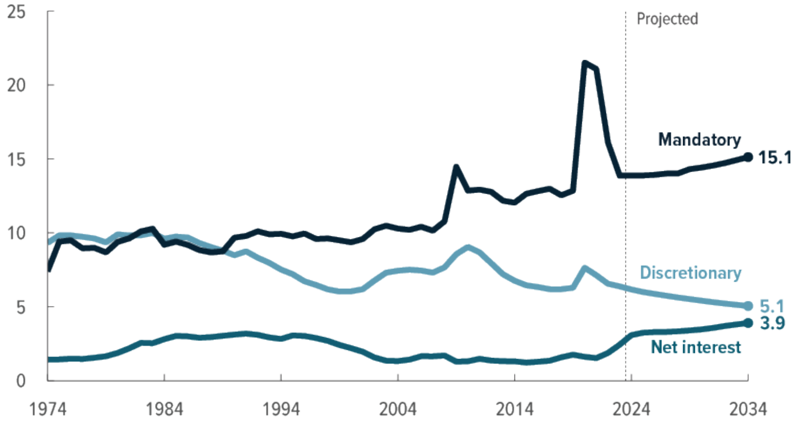
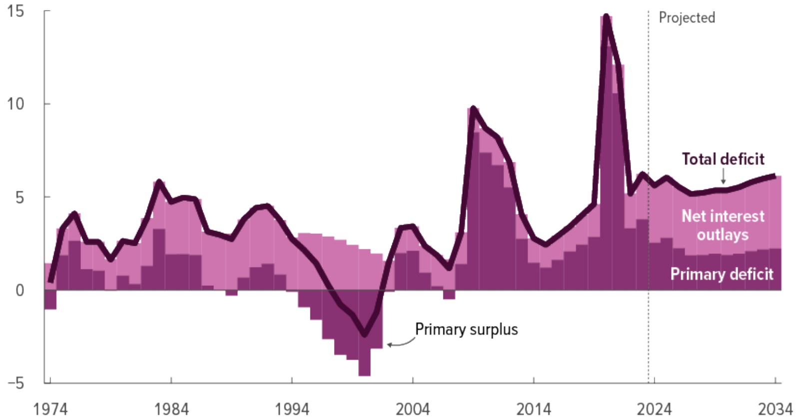


Figure 2: Deficit as a percentage of GDP, by category (source: CBO)

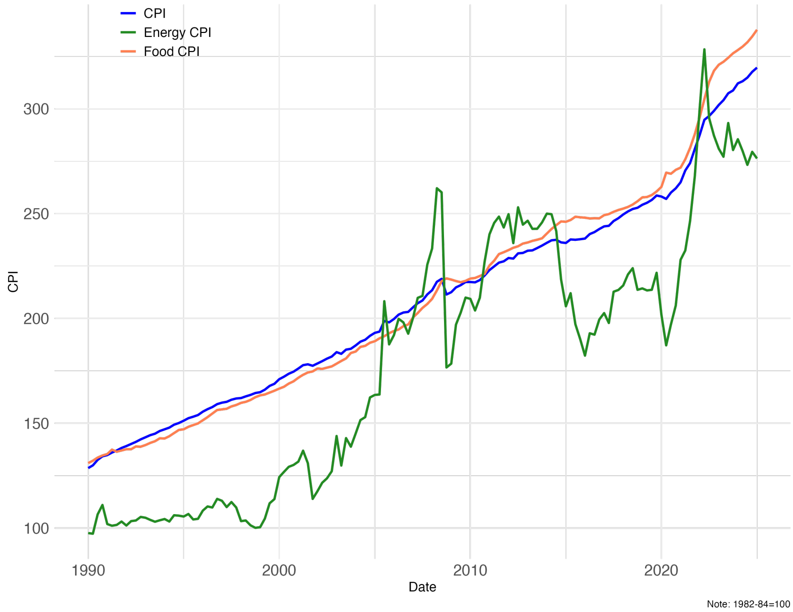


This literature review necessarily starts from the answer provided by Bernanke and Blanchard (2025) to the question opening this work. They develop a simple wage-price determination model (see Section 4), whose empirical counterpart–an approximation of a SVAR–has two main takeaways, both for the United States and for several other economies. Firstly, the source of the observed inflation was primarily the result of the relative price of inputs shocks (mainly energy) and supply chains disruptions.

These shocks did not alter expectations, which were–in contrast to the 1970s–well anchored. Thanks to this condition, the shocks did not immediately trigger a wage-price spiral. Secondly, the model yields a quite persistent response of prices to a labor market tightness shock (8d), though not statistically different from zero. Their result that postpandemic inflation was mainly supply-driven lies on the fundamental assumption of exogenous commodity prices, which is debatable. I discuss in greater detail this assumption in Section 7. Such assumption was indeed challenged in, for instance, Giannone and Primiceri (2024) and Lorenzoni and Werning (2023).

Gagliardone and Gertler (2023) show that the combination of FED's commitment to high employment and oil price shocks well accounts for the observed surge in inflation. Their explanation is partly in the spirit of BB when emphasizing the role of commodity prices, but partly highlights the role of monetary policy, which does not enter the model I treat here. Several other empirical papers have recently come out investigating the role of supply shocks. I mention here two papers with whom I share a particular emphasis on supply chain disruptions. Firstly, Bai et al. (2024) test

FIGURE 3: Consumer Prices Indexes: Energy prices (green), food prices (orange), aggregate index (blue) (source: FRED)



the relationship between the state of supply chains–proxied with a novel variable they constructed–and inflation, finding a link interpretable as causal. They think of supply chain disruptions as something "breaking the link" between production and consumption, which leads to a scenario of contemporaneous spare capacity and supply shortage. I instead interpret the shocks to the supply chain variable not merely as influencing transportation costs but more broadly as informative of the severity of supply constraints. Secondly, Comin et al. (2023) build a New-Keynesian model in which inflation stems from supply constraints. They are also able to disentangle supply-side versus demand-side explanations for binding supply constraints, finding that *"no single set of shocks can explain the inflation takeoff. Rather, shocks that tightened capacity set the stage for demand shocks–most importantly, monetary policy shocks–to trigger binding constraints and accelerate inflation in 2021"* (Comin et al., 2023). Such intuition is very much in the spirit of this work, though the frameworks we employ are different, and they have a greater emphasis on monetary policy, which does not enter my extension.

Benigno and Eggertsson (2023) impute the surge in price inflation to a non-linear Phillips curve (PC). They show that a non linear PC employing the vacancy-unemployment ratio–as BB do–well fits postpandemic data. This has important policy consequences, and suggests that monetary tightening may have a role when the economy is operating in non-linear regions. Non-linearities in supply, reflecting constraints, are crucial to the mechanism I look at.

Guerrieri et al. (2021) propose a model where inflation is characterized as stemming from structural reallocation between sectors. In such a setting, a preference shock (arguably the one that hit the economy during COVID) is equivalent to a cost-push shock generating inflation and unemployment–the former in the "growing" sector and the latter in the other. Reallocation is determined by two incentives–namely job-finding probability and real wage. Expansionary monetary policy stifles reallocation through lower unemployment in the dwindling sector but incentivizes it through wage inflation in the rising sector. The model, depending on which of the two effects dominates, prescribes opposite stances of monetary policy. Guerrieri et al. (2023) emphasize the role of supply, particularly energy, shocks in generating inflation. Their focus is on the uneven nature of

FIGURE 4: Policy rates for U.S. (blue) and EA (purple) (source: Haver Analytics)

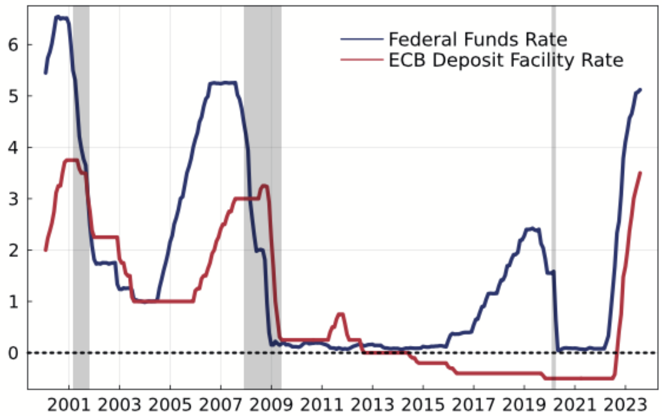
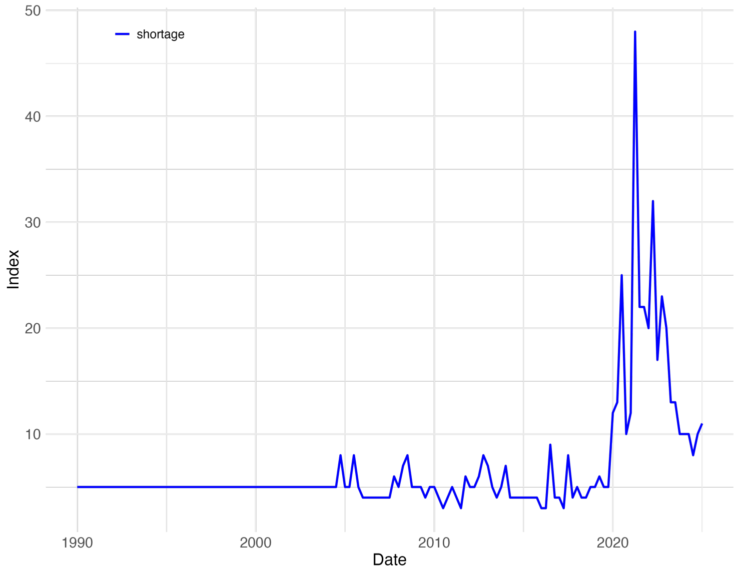


FIGURE 5: Saliency of supply constraints, proxied through Google Trend (source: Google Trend)



an energy shock and the consequent inflation heterogeneity across sectors. Such heterogeneity–in the spirit of Guerrieri et al. (2021)–gives empirical support to the argument against full tightening, as inflation is an adjustment mechanism favoring efficient allocation of resources within an economy. The focus of the paper is though mainly on the supply side, which, as I show in Section 7, is not entirely representative of postpandemic inflation drivers.

Another answer that gained momentum in the public debate goes under the name of *greedflation*, namely inflation being driven by increasing profits. Colonna et al. (2023) show that for Germany and Italy *markups* have been fairly constant or even decreasing in some sectors. Alvarez et al. (2024) show the same trend for the US. This does not invalidate increasing profits in the data, but puts them in perspective as–with constant markups–the increase in profits is mechanical with higher input costs. It shall though be noted that the ability to preserve markups vis-à-vis real wages is not trivial and is a relevant fact. This channel is accounted for in the Bernanke and Blanchard model through the variable *shortages*, as–in the authors' words–*"we are skeptical that* *firms systematically increased markups in non-shortage situations"* (English et al. (2024), p.297), and markup increases should happen only in shortages scenarios, with vertical supplies.

Theories of inflation vary widely. The most simple and emblematic understanding of inflation is summarized in Milton Friedman's words: "*Inflation is always and everywhere a monetary phenomenon*" (1963). This may though not be over-arching–as the papers in the previous paragraphs show–and was indeed challenged by Sargent et al. (1981) from a "fiscal" perspective. Sargent, in 2013, reformulated Friedman's mantra substituting *monetary* with *fiscal*. Sargent's seminal work opened a new strand of literature investigating the relationship between inflation and fiscal policy, featuring different approaches.

Cochrane (2021) shortly summarizes a long-standing effort of theorisation of the so-called Fiscal Theory of the Price Level (FTPL), which–as briefly as it could be said in the author's words–simply states "*the price level adjusts so that the real value of government debt equals the present value of real primary surpluses*"(Cochrane, 2021).[2] More intuitively, this is just another formulation of a No-Ponzi-Game condition, stating that the outstanding stock of public debt should be covered by a flow of expected future primary surpluses. This theorisation, however, encounters some controversies, treated for example by Angeletos et al. (2024). A simpler–yet effective–story, is the "Keynesian" one: with non-Ricardian agents, an increase in public spending stimulates aggregate demand; if this shift in aggregate demand (AD) takes the economy above its natural level, inflation follows. This is the mechanism underlying the textbook IS-LM framework and the Heterogeneous Agents New-Keynesian Model (HANK) literature. This was the mechanism at which Blanchard (2021) looked, and this is the mechanism I look at in Section 7. Angeletos et al. (2024) propose a model with the aim of bridging the gap between FTPL and the Keynesian mechanism within a HANK framework.

This theoretical effort was accompanied by an empirical one, which had mixed findings. Jørgensen and Ravn (2022) provide a thorough survey of the literature highlighting how time-series methods yield mixed findings on the effects of fiscal policy on inflation, as shown in Figure 6. One possible issue related to this literature–as highlighted by Hazell and Hobler (2024)–is that they do not incorporate monetary policy, whose response is crucial in "taking away the punchball" and preventing inflation or–when "late to the party"–letting it rise. My work still does not incorporate monetary policy. This, though, does not undermine the relevance of the "historical" claim I

Figure 6: From Jørgensen and Ravn (2022), survey of empirical estimates of inflation responses to fiscal shocks from papers employing a SVAR methodology.

Empirical estimates of inflation response.

Response of prices/inflation	Fiscal policy study
Inflationary	Edelberg et al. (1999), Caldara and Kamps (2008), Ben Zeev and Pappa (2017) and Ferrara et al. (2021).
Disinflationary or insignificant	Fatas and Mihov (2001a), Fatas and Mihov (2001b), Canzoneri et al. (2002), Perotti (2005), Mountford and Uhlig (2009), Nakamura and Steinsson (2014), Dupor and Li (2015), Ricco et al. (2016) and D'Alessandro et al. (2019).
Not reported	Blanchard and Perotti (2002), Burnside et al. (2004), Galí et al. (2007) and Ramey (2011)

Notes: All studies use U.S. data, though Perotti (2005) also considers other OECD countries. Canova and Pappa (2007) report evidence from the US and from Euro Area countries and obtain mixed price responses. See also Table A.1 in Appendix A.

[2] A bit more formally:

$$\frac{B_t}{P_t} = \sum_{i=0}^{\infty} \frac{(T_{t+i} - G_{t+i})}{(1+r)^i}$$

where B is nominal market value of outstanding public debt, P is the price level, T are taxes, G is government spending and r is a discount rate.

make, which highlights how–conditional on realized monetary policy–fiscal policy may be a driver of inflation. A possible extension of this paper would be to include monetary policy.

As the observed surge in inflation came just after the COVID pandemic, which saw the largest fiscal efforts in American history–about 4 trillions in around a year–the debate on the effects fiscal policy has on inflation recently naturally reignited. Sims (2024), in the spirit of FTPL, relates fiscal and inflation history of the United States highlighting qualitatively how fiscal policy may be a driver of inflation.

Barro and Bianchi (2023) test the FTPL mechanism on the postpandemic period for an OECD countries sample. They argue the mechanism was at work, and about half of excess spending was paid for with additional inflation. FTPL has two quite counterintuitive corollaries to the baseline mechanism. Firstly, a smaller debt-to-GDP ratio implies, everything else equal, a larger inflation rate; that is because a smaller stock of debt requires a *larger* inflation rate in order to balance the increase in primary deficits. Furthermore, a longer debt maturity T implies a smaller increase in the inflation rate. The reason lies in the fact that, with a longer maturity, and with the increase in Gand the necessary inflation rate across periods held constant, for longer T, the required increase in each period is smaller, that is, the devaluation is more evenly spread. Since these two mechanisms are empirically relevant, Barro and Bianchi (2023) interpret them as a robustness element in favor of FTPL.

Hazell and Hobler (2024) test the causal relationship between fiscal deficits and inflation exploiting the elections on January 5th, 2021 in Georgia awarding two seats in the Senate. These two seats would determine the possibility for Democrats to pass fiscal stimulus. In the paper, they find that an exogenous variations in the set of information on future deficits–namely the gain of a majority in Senate for Democrats–had a positive and sizeable effect on markets' inflation expectations. This result comes both from an event study, which estimates that the Georgia Shock statistically significantly increased inflation expectations at various time horizons, and from an IV strategy that uses winning probability to instrument likelihood of future fiscal stimulus, yielding an estimate of 0.65 p.p. price level increase over 1 year for a 50 p.p. increase in probability of win by the Democrats. The paper does not define a specific mechanism behind this result, be it Keynesian or FTPL.

Two recent papers find a significant role of fiscal policy in the recent inflation episode, coherent with the one I highlight in this paper. Mori (2025) shows in a VAR that fiscal shocks account for a large part of inflation not only in the postpandemic episode. My estimated fiscal contribution is coherent with its estimated trajectory of inflation under fiscal shocks, as both peak in 2021Q1. The estimates though quantitatively differ because of the different inflation measures we use and because of the different magnitude of fiscal shocks. Additionally, he interprets its findings within a New-Keynesian model à la Angeletos et al. (2024). My findings are instead more in the spirit of the model proposed by Fornaro (2024), where the aggregate supply curve has a shifting kink and fiscal policy takes the economy on the steeper region of the curve, thereby generating inflation.

4. THE BERNAKE–BLANCHARD EMPIRICAL MODEL

The empirical model proposed in Bernanke and Blanchard (2025) consists of four equations, which read as follows.

The wage equation

$$\begin{aligned}\Delta w_t = &\sum_{i=1}^{4} \beta_{1,i} \Delta w_{t-i} + \sum_{i=1}^{4} \beta_{2,i} (v/u)_{t-i} + \sum_{i=1}^{4} \beta_{3,i} \Delta p^E_{t-i} \\ &+ \sum_{i=1}^{4} \beta_{4,i} \left(\Delta^4 p_t - \Delta^4 p^E_{t-4}\right) + \beta_5 gpty_{t-1} + u_{1,t}\end{aligned} \tag{1}$$

where the following restriction is imposed $1-\sum_{i=1}^{4}\beta_{1,i}=\sum_{i=1}^{4}\beta_{3,i}$. This restriction can be thought of as imposing that–in the long run–there is no trade-off between inflation and employment, or that the economy operates at the natural rate of employment.

The price equation

$$\begin{aligned}\Delta p_t = &\sum_{i=1}^{4}\gamma_{1,i}\Delta p_{t-i}+\sum_{i=0}^{4}\gamma_{2,i}\Delta w_{t-i}+\sum_{i=0}^{4}\gamma_{3,i}grpe_{t-i}\\ &+\sum_{i=0}^{4}\gamma_{4,i}grpf_{t}+\sum_{i=0}^{4}\gamma_{5,i}\text{ shortages }_{t}+\gamma_6\text{ gpty }_{t-1}+u_{2,t}\end{aligned} \tag{2}$$

where the following restriction is imposed $1-\sum_{i=1}^{4}\gamma_{1,i}=\sum_{i=1}^{4}\gamma_{2,i}$, which as above posits a long-run vertical Phillips curve.

The inflation expectations equations

$$\Delta^4 p_t^E=\sum_{i=1}^{4}\delta_{1,i}\Delta^4 p_{t-i}^E+\sum_{i=0}^{4}\delta_{2,i}\Delta^{40}p_{t-i}^E+\sum_{i=0}^{4}\delta_{3,i}\Delta p_{t-i}+u_{3,t} \tag{3}$$

$$\Delta^{40}p_t^E=\sum_{i=1}^{4}\delta_{5,i}\Delta^{40}p_{t-i}+\sum_{i=0}^{4}\delta_{6,i}\Delta p_{t-i}+u_{4,t} \tag{4}$$

where the following restrictions are imposed $\left(1-\sum_{i=1}^{4}\delta_{1,i}=\sum_{i=0}^{4}\delta_{2,i}+\sum_{i=0}^{4}\delta_{3,i}\right)$; $(1-\sum_{i=1}^{4}\delta_{5,i}=\sum_{i=0}^{4}\delta_{6,i}$. More intuitively, they impose that, in the long run, expected inflation converges to realized inflation.

5. DATA AND METHODOLOGY

5.1. **Methodology**

The empirical approach taken by the authors is–in their own words–"*a hybrid approach that approximates the results of a structural vector autoregression with added exogenous variables*." Indeed, it would approximate a SVAR-X. The model though has contemporaneous variables. Thus, it can simply be described as a system of equations whose coefficients are estimated via ordinary least squares and then simulated.

5.1.1. *Impulse Response Functions* To derive responses of inflation to exogenous shocks, I run dynamic simulations for the four (then five) equations both in a baseline scenario and adding five different types of shocks:

1. Energy price shock
2. Food price shock
3. Shortages index shock
4. Labor market tightness (V/U) shock
5. Fiscal deficit shock (see Section 7)

The difference between the "shocked" and the baseline simulations yields the IRFs provided below. That is, more formally, I am using–as authors of the paper do–Hamilton (1994) conditional forecast difference formulation of an IRF:

$$\text{IRF}\,(t,h,\boldsymbol{a}_i)=\mathbb{E}(\boldsymbol{y}_{t+h}\mid\boldsymbol{u}_t=\boldsymbol{a}_i)-\mathbb{E}(\boldsymbol{y}_{t+h}\mid\boldsymbol{u}_t=\boldsymbol{0})\quad h=0,1,2,\ldots\mathbf{T} \tag{5}$$

where $\mathbf{T}$ is the length of the simulation (21 in my case), and $\boldsymbol{a}_i$ changes with the properties of the shocks.

5.1.2. *Historical Decomposition* The most interesting contributions of Bernanke and Blanchard (2025) and Bernanke and Blanchard (2024) is the historical decomposition of price inflation for the United States and 10 other economies. In this setting, the contribution each variables gives to inflation is derived as the difference between the "baseline" dynamic simulation–which has all the observed values for exogenous variables–with a dynamic simulation, which has one component kept fixed at an arbitrary value. A bit more formally, we have the following:

Note that, with x_j being an exogenous regressor, the system in each period may be written as

$$\boldsymbol{y}_t = \boldsymbol{\beta}_0 + \sum_{k=1}^{k=4} \boldsymbol{\beta}_1 \boldsymbol{y}_{t-k} + \sum_{k=0}^{k=4} \boldsymbol{\beta}_2 x_{1t-k} + \ldots + \sum_{k=0}^{k=4} \boldsymbol{\beta}_j x_{jt-k} + \boldsymbol{u}_t$$

Then, the dynamic contribution χ of the j-th variable to the system in each period is

$$\chi_{j;t+h} = \mathbb{E}(\boldsymbol{y}_{t+h} \mid x_{j;t+h} = x_{j;t+h}) - \mathbb{E}(\boldsymbol{y}_{t+h} \mid x_{j;t+h} = \overline{x}_j) \quad h = 0, 1, 2, \ldots \mathbf{T}$$

where $\mathbf{T}$ is the length of the simulation (21 in my case) and $\overline{x}_j$ is an arbitrary value as in Table 1. Initial conditions are derived in the same way, but with the difference that *all* variables are *simultaneously* set at the arbitrary value. It is worth to note that the contribution of the j-th variable is derived exactly as an impulse response (see 5), that is, through a difference of two conditional simulations. This gives a deeper understanding of the functioning of a historical decomposition: it tracks the contribution in each period of a single shock in driving a variable away from its initial conditions.

TABLE 1: Arbitrary values for exogenous variables employed in the historical decomposition.

Variable	**Value**
Energy prices (growth)	0
Food prices (growth)	0
Shortages index	Pre-COVID mean
Productivity (growth)	Pre-COVID mean
V/U ratio	Pre-COVID mean
Fiscal surplus (deficit)	Pre-COVID mean

5.2. **Data**

The data I use in this work are the same as those in Bernanke and Blanchard (2025), updated to 2025Q1. In particular,

- Price inflation (Δp) is derived from the Consumer Price Index regarding all items in US cities, taken from FRED.
- Wage inflation (Δw) is derived from the Employment Cost Index (ECI), taken from the Bureau of Labor Statistics (BLS).
- Inflation expectations (Δp^E), at 1 year and at 10 years, are taken from the Cleveland FED.
- The catch-up term is derived as the difference between expected and realized inflation.
- Energy inflation ($grpe$) is derived from an energy CPI, taken from FRED. Notably, the model is estimated using relative prices of energy. I thus normalize this CPI by the ECI.

- Food inflation (*grpf*) is derived from a food CPI, taken from FRED. Even food prices are normalized by ECI.
- The vacancy/unemployment (v/u) ratio is taken from the BLS job openings and labor turnover survey (JOLTS) and the BLS Employment Report. As the series is truncated, it is reconstructed using data from Barnichon (2010). While V/U generally comoves strictly with unemployment during COVID, the two diverged. V/U was chosen as it better captures–for a given number of unemployed people–the increase in workers' bargaining power through increased search efforts and the consequent upward pressure on wages. This was a choice made also in Benigno and Eggertsson (2023) in a non-linear Phillips Curve framework and is generally standard in models with search.
- The state of supply chain (*shortages*) is proxied with Google Trend. In practice, this approach simply consists in using Google estimated saliency of some words on the search engine. In particular, here, the word is "shortages" and is used to proxy the stress of supply chains.
- Trend productivity growth (*gpty*), measured by the change in the eight-quarter moving average of nonfarm business value added divided by nonfarm employee hours, from the BLS.
- Fiscal surpluses (*fiscal*) are normalized to GDP and de-seasoned using a rolling mean adjustment. The source for both the numerator and the denominator is FRED.

6. REPLICATION

To answer the question opening this work, I have developed a R script that fully replicates the results of Bernanke and Blanchard (2025). The advantage of this package is that it works in a single

TABLE 2: Wage equation.

	Wage equation
Constant	-0.27 (0.21)
gw_l1	0.16 (0.10)
gw_l2	0.12 (0.10)
gw_l3	0.17 (0.10)*
gw_l4	0.00 (0.10)
vu_l1	3.75 (1.97)*
vu_l2	-1.84 (3.61)
vu_l3	-3.56 (3.63)
vu_l4	2.33 (2.09)
exp1_l1	0.34 (0.14)**
exp1_l2	-0.03 (0.15)
exp1_l3	0.21 (0.16)
exp1_l4	0.03 (0.16)
catch_up_l1	-0.01 (0.08)
catch_up_l2	0.01 (0.08)
catch_up_l3	-0.00 (0.08)
catch_up_l4	-0.02 (0.07)
gpty_l1	0.03 (0.06)
eq1: R^2	0.56
eq1: Adj. R^2	0.49
Num. obs. (total)	120

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

NOTES. Wage equation. Sample period is 1990Q1:2019Q4. An homogeneity restriction was imposed to ensure that, in the long run, wage Phillips curve is vertical. Standard errors in parentheses.

TABLE 3: Price equation.

	Price equation
Constant	-0.12 (0.22)
gp_l1	0.06 (0.11)
gp_l2	0.22 (0.10)**
gp_l3	0.27 (0.11)**
gp_l4	-0.11 (0.09)
gw	0.35 (0.10)***
gw_l1	0.17 (0.11)
gw_l2	0.02 (0.11)
gw_l3	0.01 (0.11)
gw_l4	-0.03 (0.06)
grpe	0.09 (0.00)***
grpe_l1	-0.00 (0.01)
grpe_l2	-0.02 (0.01)*
grpe_l3	-0.02 (0.01)**
grpe_l4	0.01 (0.01)
grpf	0.11 (0.04)***
grpf_l1	-0.01 (0.04)
grpf_l2	-0.00 (0.04)
grpf_l3	-0.01 (0.04)
grpf_l4	0.03 (0.04)
gpty_l1	-0.07 (0.06)
shortage	0.10 (0.02)***
shortage_l1	-0.03 (0.02)
shortage_l2	-0.00 (0.02)
shortage_l3	-0.03 (0.02)
shortage_l4	-0.02 (0.02)
eq1: R^2	0.94
eq1: Adj. R^2	0.93
Num. obs. (total)	133

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

NOTES. Price equation. Sample period is 1990Q1:2023Q1. Notably, the price equation is estimated on a longer sample than the other equations, permitting to capture the effects of "shortages," which had low variability in the pre-pandemic sample. This justifies the different number of observations. A homogeneity restriction was imposed, ensuring that coefficients on past price inflation and on wage inflation sum to 1. Standard errors in parentheses.

environment and it allows for smoother replication of the results, as well as faster modifications of the model. Moreover, I have also built a version of the model-implied impulse responses with confidence intervals, which I derive through bootstrapping. In this Section, I show the results of these two exercises, along with some other robustness checks. In addition to the tables provided below, in Appendix 1, I provide plots of my replicated coefficients against those of the paper as a robustness check. These plots refer to the same sample used in BB. Notably, coefficients match (almost) perfectly except for some minor differences in the price equation estimates. The tables also permit to verify the validity of the restrictions I imposed in Section 4, by simply adding the coefficients.

With the estimated coefficients in hand, I run dynamic simulations for the four equations both in a baseline scenario and adding four exogenous shocks:

1. Energy price shock

2. Food price shock

Table 4: Short-run inflation expectations

	1y exp equation
exp1_l1	0.30 (0.09)***
exp1_l2	−0.23 (0.10)**
exp1_l3	0.18 (0.10)*
exp1_l4	0.12 (0.07)
exp10	1.20 (0.14)***
exp10_l1	−0.47 (0.23)**
exp10_l2	0.02 (0.23)
exp10_l3	0.14 (0.23)
exp10_l4	−0.38 (0.17)**
gp	0.04 (0.01)***
gp_l1	0.10 (0.01)***
gp_l2	0.00 (0.02)
gp_l3	0.01 (0.02)
gp_l4	−0.02 (0.02)
eq1: R^2	0.90
eq1: Adj. R^2	0.89
Num. obs. (total)	120

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes. Short-run (1y) expectations equation. Sample period is 1990Q1:2019Q4. A homogeneity restriction was imposed to ensure that, in the long run, expected and realized price inflation converge. Standard errors in parentheses.

Table 5: Long-run inflation expectations

	10y exp equation
gp	0.03 (0.01)***
gp_l1	0.01 (0.01)
gp_l2	−0.01 (0.01)
gp_l3	−0.00 (0.01)
gp_l4	−0.00 (0.01)
exp10_l1	0.85 (0.09)***
exp10_l2	−0.05 (0.12)
exp10_l3	0.20 (0.12)
exp10_l4	−0.03 (0.09)
eq1: R^2	0.93
eq1: Adj. R^2	0.92
Num. obs. (total)	120

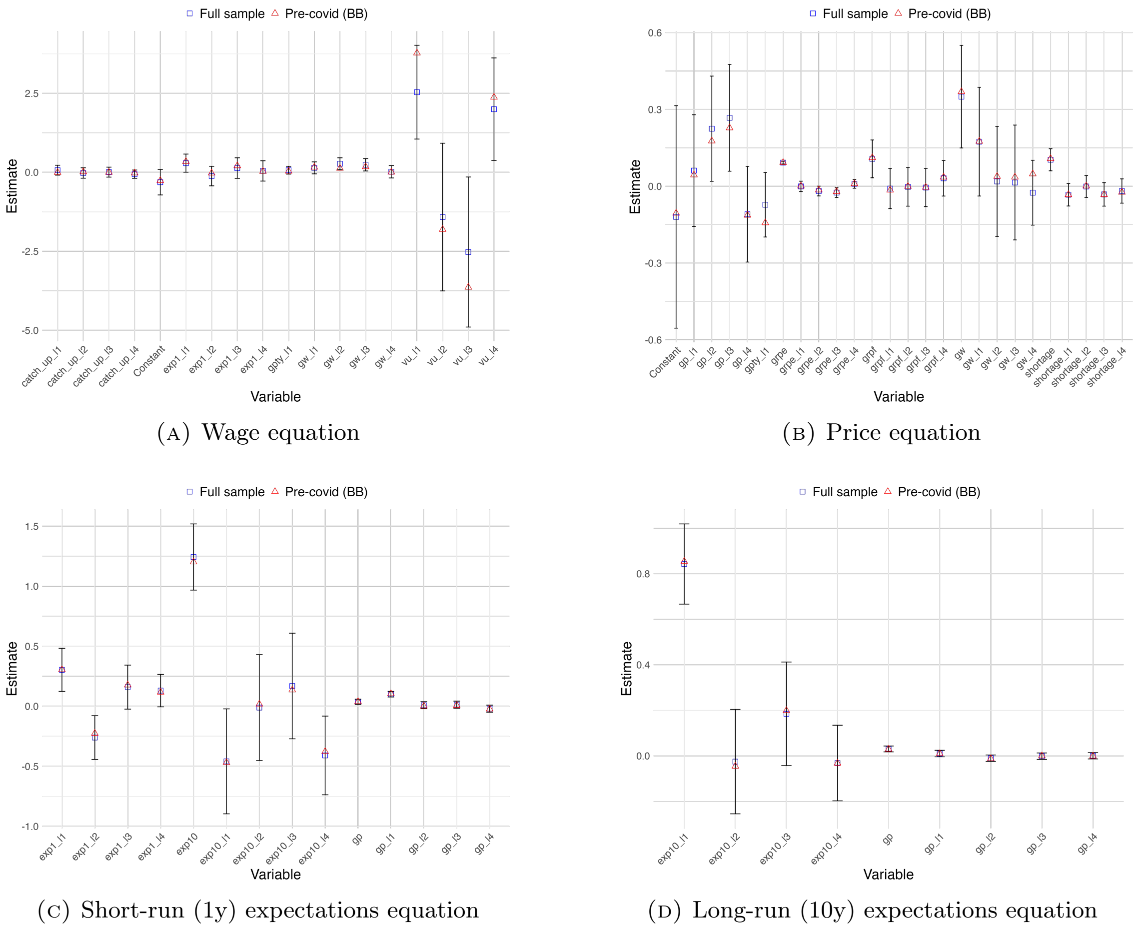
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes. Long-run (10y) expectations equation. Sample period is 1990Q1:2019Q4. A homogeneity restriction was imposed to ensure that, in the long run, expected and realised price inflation converge. Standard errors in parentheses.

3. Shortage index shock
4. Labor market tightness (V/U) shock

As the paper estimates three of the four equations on a truncated sample (up to 2019:Q4), with only the price equation being estimated on the full sample, I investigate how this choice affected the results estimating parameters on the full sample and plotting the two estimates one against the other. It turns out that this had minor effects, as can be seen in Figure 7. The only effect seems to be a bias toward zero of the labor-market variable in the full sample. Confidence intervals refer to my estimates.

FIGURE 7: Confrontation of estimated coefficients on full (up to 2023:Q1) v. Pre-COVID (up to 2019:Q4) sample.



(A) Wage equation

(B) Price equation

(C) Short-run (1y) expectations equation

(D) Long-run (10y) expectations equation

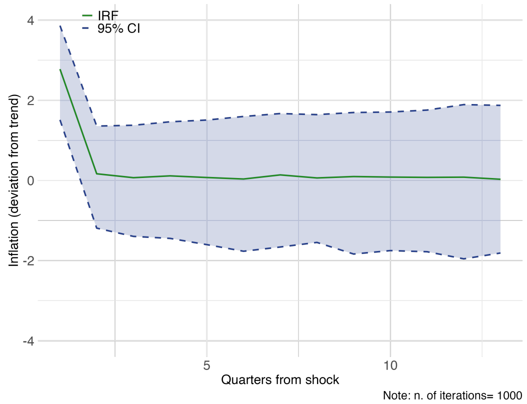
In Figure 8, adding to the results of the paper, I build confidence intervals by bootstrapping. I draw with replacement *rows* of the residual matrix to preserve the row correlation among the four equations, and simulate the equations using resampled residuals. I run n=1000 iterations and select the estimate and the upper and lower bound through quantiles. A convenient feature of bootstrapping is the fact that it allows inference without having to make assumptions on the underlying distributions. Appendix 3 shows the implied residuals distributions for the four equations. As one can see, the response of inflation to the shocks is generally not statistically different from zero. An exception is the first three quarters response induced by a one standard deviation shock to energy prices.

Figure 9 replicates the historical decomposition of price and wage inflation provided in Bernanke and Blanchard (2025) using the methodology described in Section 5.1.2. The replication of these exercises is the basis for extensions of the model, as the one I propose in the next Section.

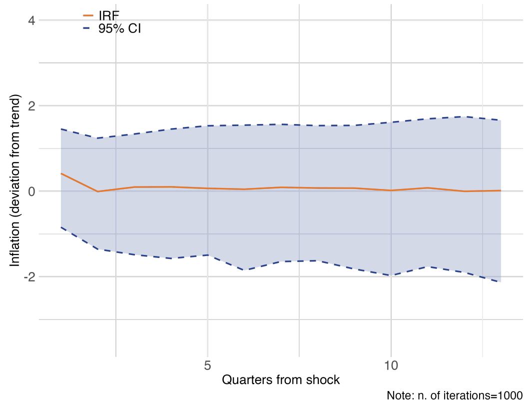
7. AN EXTENSION

In relation with the literature presented in Section 3, I propose an extension of the simple wage-price determination model introduced in Bernanke and Blanchard (2025) that accounts for fiscal policy.

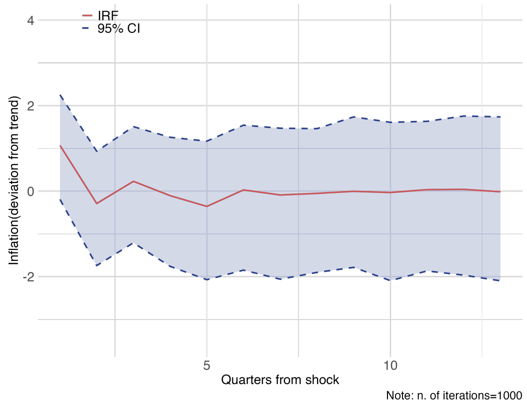
FIGURE 8: Impulse responses of price inflation to four exogenous shocks with confidence intervals. Shaded areas represent 95% confidence intervals derived through bootstrapping.



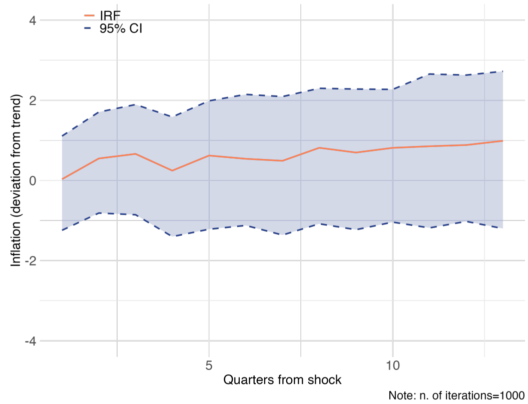
(A) 1 sd temporary energy prices inflation shock



(B) 1 sd temporary food prices inflation shock



(C) 1 sd temporary shortages index shock



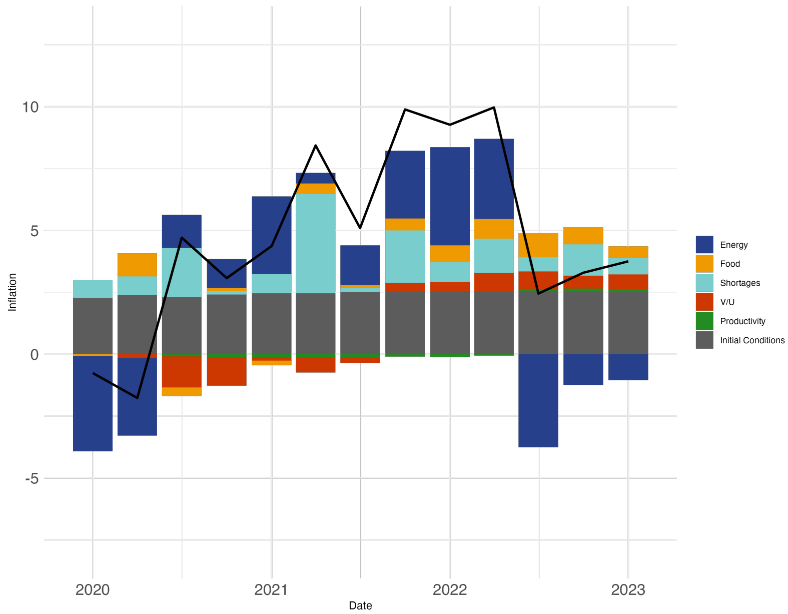
(D) 1 sd permanent V/U shock

Adding one equation to the model outlined in Section 4, I endogenize the variable *shortages* as follows

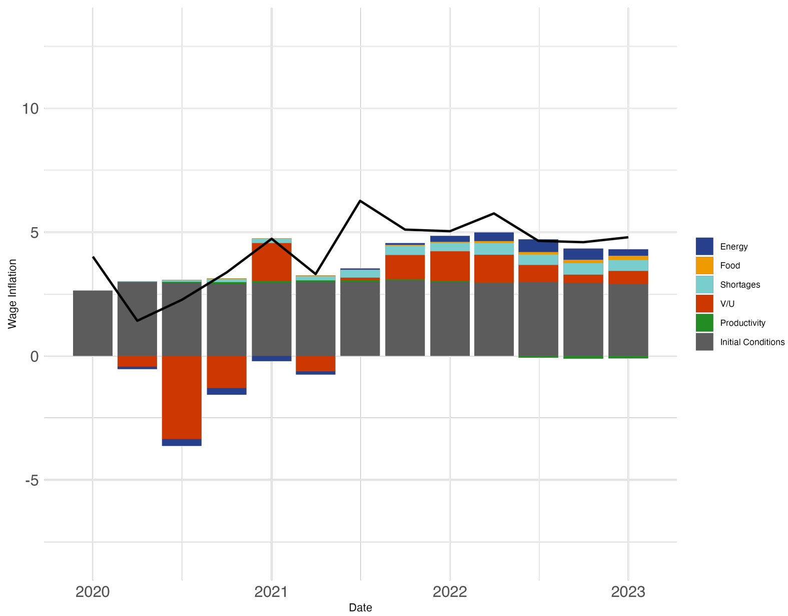
$$shortages_t = \sum_{i=1}^{4} \gamma_{1,i} shortages_{t-i} + \sum_{i=0}^{4} \gamma_{2,i} fiscal_{t-i} + u_{5,t} \tag{6}$$

where $fiscal$ is the de-seasoned fiscals-surplus-to-GDP ratio, constructed from data available on FRED. It is particularly appropriate to use the fiscal surplus rather than government purchases to measure fiscal stance because it captures movements in taxes and transfers, which were particularly relevant in the postpandemic response of fiscal policy, as highlighted by Mori (2025). Figure 11 shows the two series with two different seasonal adjustments. The series I estimate the model with is the rolling mean adjustment. Estimated coefficients for this equation are reported in Table 6. With this new system in hand, it is possible to replicate all the standard exercises provided above. A possible threat to identification here is simultaneity bias; that is, to say both the saliency of supply constraints and fiscal policy are determined by unobservables. I partially address it including lags for fiscal deficits, but this is clearly a limitation to the identification strategy. Figure 12 shows the shortage simulation, while the simulation of the other four equations does not differ much from the standard one provided in Section 6 and is included in Appendix 2, as well as implied

FIGURE 9: Historical decomposition of price inflation (panel A) and wage inflation (panel B), replication of Bernanke and Blanchard (2025). Each bar represents the fully dynamic–both direct and indirect–contribution of each component to price and wage inflation. The black line represents the actual path of inflation. The difference between the bars and the actual values line represents model residuals.

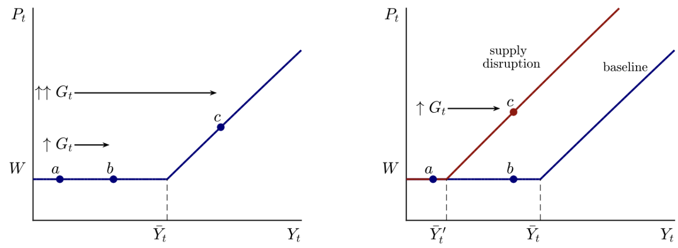


(A) Price inflation



(B) Wage inflation

FIGURE 10: From Fornaro (2024), a framework with fiscal policy and supply constraints.



IRFs (5). Figure 13 shows the historical decomposition of price and wage inflation within the new specification of the system. There is one notable differences with Figure 9 and with the conclusions drawn in Bernanke and Blanchard (2025): the major role played by shortages in the paper is largely displaced by the fiscal variable. While energy prices appear to remain the main driver of the (dis)inflation path in the United States, supply chains seem not to be a driver of inflation *per se*. Rather, it is demand that pushes the economy against its supply constraints. This is the story I tell in this paper. In an economy sensitive to supply constraints, as many globalized ones after the pandemic, it is plausible to think that an impulse to aggregate demand in the form of unprecedented public spending takes the economy in regions where these constraints are binding, ultimately resulting in inflation. The Google Trend variable measures precisely the saliency of such supply constraints. This mechanism is formalized in Fornaro (2024), and its intuition can be grasped from Figure 10. The coincidence of an unprecedented expansionary fiscal stance and major supply disruptions can be easily seen in this framework to pave the way for a surge in inflation. This happens because when firms try to increase production beyond their supply constraints, marginal costs increase and so do prices. Production under binding supply constraints is akin to an increase in markup, where inflation spikes with constant wages.

FIGURE 11: Fiscal surplus/GDP ratio, original and de-seasoned series. (Source: FRED)

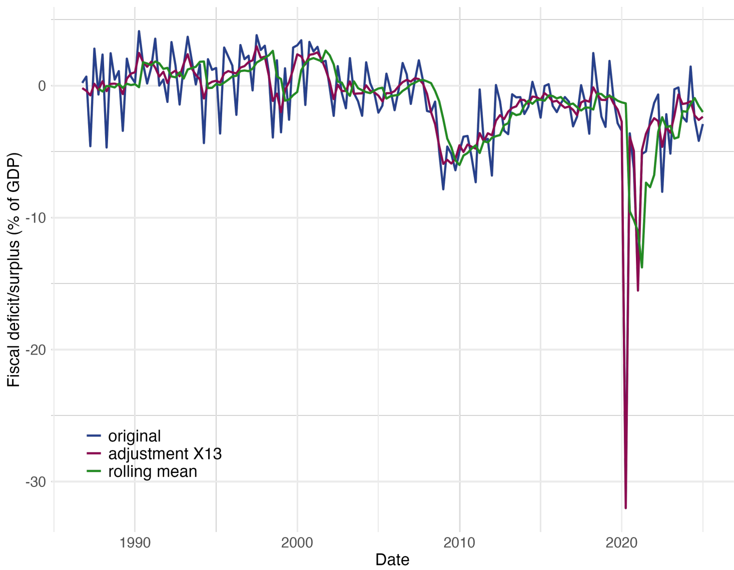
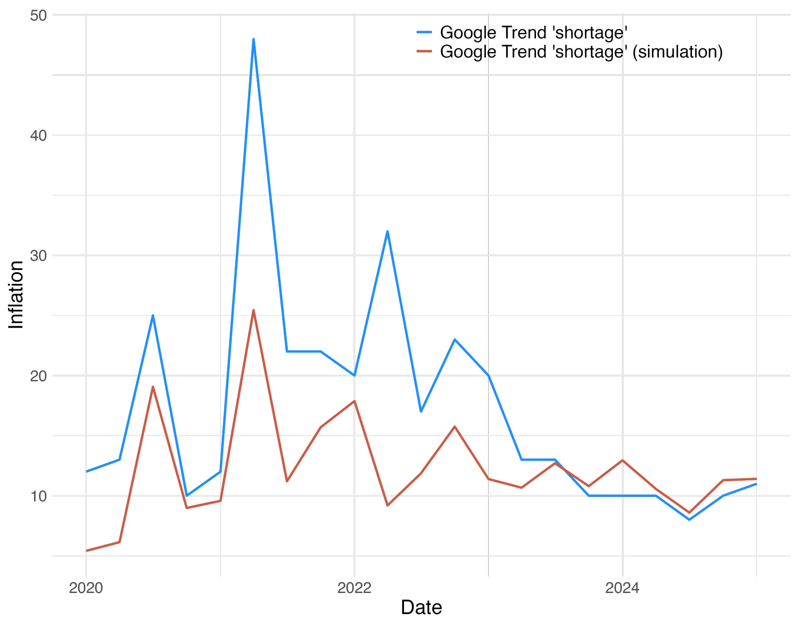


TABLE 6: Shortages equation (see (6)) estimated coefficients. Sample period is 1990Q1:2025Q1. Sample period includes the pandemic as the variation in shortages in the prepandemic period was low. Standard errors in parentheses.

	Shortages equation
Constant	0.65 (0.46)
shortage_l1	0.27 (0.08)***
shortage_l2	−0.03 (0.09)
shortage_l3	0.16 (0.09)*
shortage_l4	0.46 (0.08)***
fiscal	−1.61 (0.27)***
fiscal_l1	1.78 (0.41)***
fiscal_l2	−0.40 (0.43)
fiscal_l3	−1.22 (0.42)***
fiscal_l4	1.08 (0.29)***
eq1: R^2	0.74
eq1: Adj. R^2	0.73
Num. obs. (total)	141

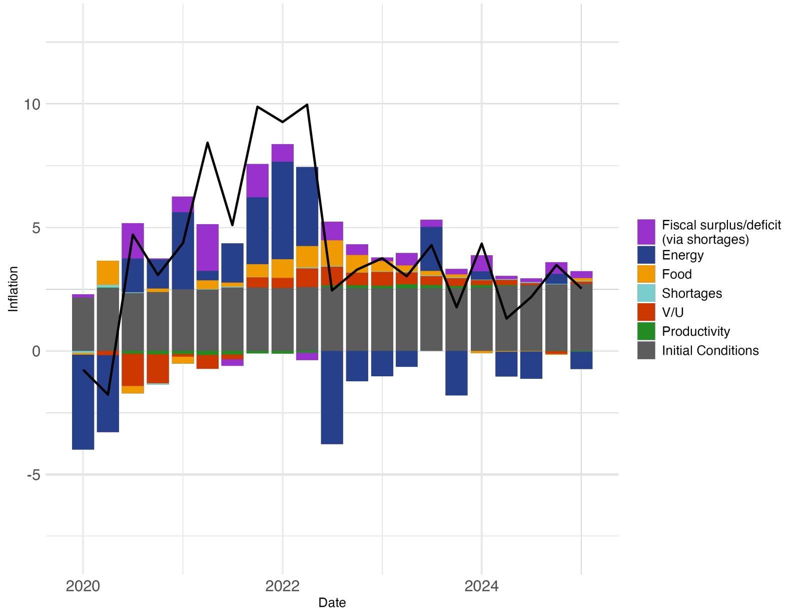
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

FIGURE 12: Shortages: real v. dynamic simulation (source: Google Trend)

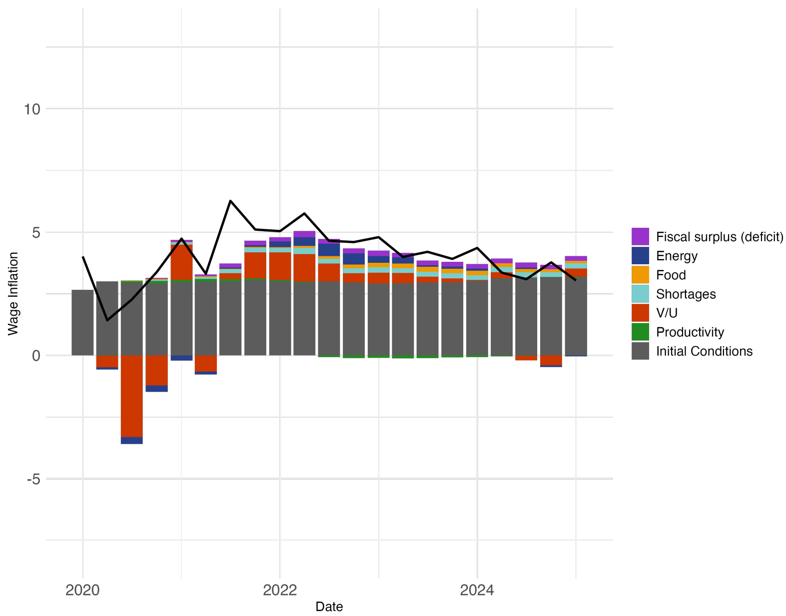


Even though a historical decomposition as that proposed by BB makes the case for supply, there seems to be an underlying *demand* mechanism that did not operate via the traditional labor market channel, whose relevance remains low. Indeed, as shortages become determined endogenously, one is able to estimate a "fiscal component" of shortages. What previously looked as being entirely determined by supply-side factors now can be disentangled. Fiscal shocks are seen to have a positive and significant effect on price inflation on impulse, which soon becomes non-statistically significant (see Figure 20b). Fiscal stance has driven a non-negligible share of inflation in the last five years and its influence is still present in the most recent data. Indeed, it sums in the period

FIGURE 13: Price inflation decomposition (Panel A) and wage inflation decomposition (Panel B). I extend Bernanke and Blanchard (2025) endogeneizing the shortages variable with the inclusion of a fiscal proxy. This permits to estimate a fiscal (via shortages) component of inflation. Each bar represents the fully dynamic–both direct and indirect–contribution of each component to price and wage inflation. The black line represents the actual path of inflation. The difference between the bars and the actual values line represents model residuals. Decomposition goes from 2020Q1 to 2025Q1.

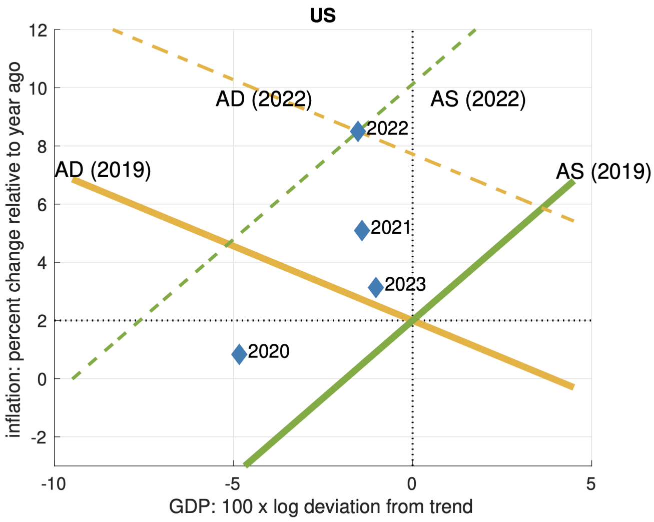


(A) Price inflation



(B) Wage inflation

FIGURE 14: From Giannone and Primiceri (2024): a simple AS-AD framewor.k



2020Q1:2025:Q1 to 2.4%. Notably, inflation expectations were unaltered, and the implied impulse response function of inflation expectations to fiscal shocks is flat (see Appendix 6).

The result I provide fits well in the mosaic of some recent works arguing how postpandemic inflation shall not be thought as mainly driven by supply factors. Above all, this is the thesis of Giannone and Primiceri (2024). In the paper, the authors estimate a bi-variate VAR identifying demand and supply shocks through sign restrictions. They show in a historical decomposition analogous to the one provided above that inflation in the US and the Euro-area was mainly driven by general demand factors. They rationalize this empirical finding in a simple aggregate supply-aggregate demand (AS-AD) framework, as in Figure 14. The pandemic induced an inward shift of both AS and AD. The credibility of the FED, modeled through a relatively flatter AD curve[3], led to initially observe large output effects with small effects on inflation. In this simple context, the only way to rationalize the surge in inflation with low output disturbances observed after 2020 is through either an outward shift of AD or a change in its slope–which the authors argue to be the same thing from a monetary policy perspective, namely a (too) ease monetary policy. More precisely, it was the case that AD bounced back *relatively faster* than AS. As said above, expansionary fiscal policy–what I look at in this work–enters this framework as a force shifting outward AD (but should not generally influence the slope of AD).

This interpretation of inflation is not necessarily in contradiction with the one provided by Bernanke and Blanchard. Rather, superficially, inflation was due to commodity prices and rising supply-chain pressures [4]. However, it is quite a strong assumption to treat these components as fully exogenous, as BB do. Thus, endogenizing these factors, one sees that the *pure* contribution of supply shocks is small, and the effect previously observed was mediating a demand component. This is an exercise done both in Giannone and Primiceri (2024) and in this work. BB indeed acknowledge the role of demand forces (including fiscal policy) in determining supply-side factors, but integrating them in their framework remains an open question that I have picked up in this work. Even if Giannone and Primiceri (2024) focus on energy prices and monetary policy and I

[3]In the extreme case of the most credible central banker possible, one that does not inflation deviate from its target, we would observe output fluctuations without any inflation at all; that is, to say the economy would be moving horizontally on a flat AD curve.

[4]"*in an accounting or reduced-form sense*"(Giannone and Primiceri, 2024)

look at supply chains and fiscal policy and the magnitudes are different, the two results seem to be similar in spirit and pointing to the same AS-AD mechanism.

A similar mechanism is formalized in Lorenzoni and Werning (2023). They propose a New-Keynesian model in which a demand shock (be it preference, monetary, or fiscal) hits an economy that produces employing a supply-constrained non-labor input whose price is endogenous. In this context, a "*supply-constrained demand shock*" would generate a dynamic response of inflation in three phases: firstly, a fast increase in non-core inflation, driven mainly by the supply-constrained input(s), followed by a period of sustained price and wage inflation (where prices outpace wages), with a third period of wage inflation and real wages "catching-up." Despite the fact that it is commonly intuitive to think of demand shocks as working mainly through output effects, labor market overheating, and wage inflation, in an economy sensitive to supply constraints, this may not be the case. Excess demand may make its own way to inflation without passing primarily through the "traditional" labor market channel but through endogenous commodity prices. Although not immediately intuitive, this is still demand-driven inflation. In this spirit, imagine again an AS-AD framework as the one in Figure 14. Lorenzoni and Werning (2023) nicely summarize this particular manifestation of excess demand as follows: "*inflation is in general caused by some form of scarcity on the supply side, relative to existing demand pressures. But there are multiple inputs on the supply side, labor inputs and non-labor inputs. Depending on the episode, scarcity can manifest itself more strongly in labor inputs or in non-labor inputs. When non-labor input scarcity dominates, price inflation will be faster than wage inflation*"(Lorenzoni and Werning, 2023). In this light, the classical intuitive distinction between "supply-driven" and "demand-driven" inflation looks more nuanced, and inflation may stem from the interaction between these two forces at play.

Some more technical notes: Firstly, a notable point is that when simply adding the fiscal variable on the Right-Hand side of Equation 2, the historical decomposition did not look very different from the one provided in Figure 13, with a direct displacement of the shortage component and other components mostly unaltered (see Appendix 7). This drove the choice to discipline demand shocks through shortages, but may be informative of some deeper dynamics. In particular it remains to figure out why the pure role of other commodity prices is robust to the inclusion of the fiscal variable. It would, in fact, be intuitive to think that as long as it is appropriate to endogenize the state of supply chains, so it is for commodity prices. The fiscal component here relates to the price level through a Keynesian mechanism, that is shifting outward aggregate demand. In this framework, credibly testing empirically an FTPL-like mechanism would be less immediate. Secondly, the relevance of the case for endogenizing variables treated as exogenous is a function of the size of the economy being treated. For what concerns commodity prices, treating them as exogenous is a sound approximation for a small open economy, as it hardly affects the determination of prices on global markets. The same does not hold for a large open economy, as the United States, which likely does not operate atomistically in the determination of the prices of, for example, energy. The same line of reasoning may be extended to the state of supply chains: a large economy reasonably influences the state of global supply chains, including through its aggregate demand.

It should be noted that Bernanke and Blanchard (2025) state in their paper: "*The collision of high demand and limited supply in some sectors can account for at least some of the increase in markups observed during the pandemic period. While other factors no doubt influenced markups, including for example the fiscal transfers that directly affected demand in product markets, at least in this simple specification we do not find that including these factors is needed to explain the behavior of pandemic-era inflation.*" While it is true that the inclusion of other regressors does not influence much the simulation of price inflation (see Appendix 3), coefficients vary a lot both in magnitude and significance (confront Appendies 1 and 7), which decomposition-wise is a key fact yielding the direct displacement we observe in Appendix 7.

The evidence I provide and the theory introduced in Fornaro (2024) imply various lessons for fiscal policy. Firstly, the "fiscal Phillips multiplier"–broadly intended as the increase in inflation induced by government spending–is large when supply constraints bind, or in presence of adverse supply shocks. Secondly, government spending concentrated on some sectors implies higher Phillips multipliers, as it is more likely to push the few firms to which it is directed on their supply constraints. Thirdly, government spending is not all the same: expenditures that foster innovation and productivity–e.g., R&D investments–are not as inflationary as transfers, that are funneled directly into consumption. Further empirical evidence is needed to evaluate more thoroughly these hypotheses.

8. CONCLUSIONS

The BB replication package I developed for this work and the extension I proposed will–hopefully–be useful for future research and deeper understanding of the topic. There are various open paths that could be taken to improve my approach: testing FTPL-like mechanisms; verify whether there are other demand components at play to disentangle working through, say, energy prices; taking the model to other countries, e.g., to Italy (extending Pisani and Tagliabracci, 2023), to see whether the recent significant increase in public expenditure translated into inflation; polish the specification so as to be able to estimate a global demand component and not capture only one single channel; integrate the criticisms to the BB approach put forward in Bergholt et al. (2023); and integrate monetary policy in this framework. I hope these will be the starting points of many interesting future works.

REFERENCES

Alvarez, S., A. Cavallo, A. Mackay, and P. Mengano (2024). Markups and cost pass-through along the supply chain. *Unpublished manuscript, Harvard Business School*.

Angeletos, G.-M., C. Lian, and C. K. Wolf (2024, Nov). Deficits and inflation: Hank meets ftpl. NBER Working Papers 33102, National Bureau of Economic Research, Inc.

Bai, X., J. Fernández-Villaverde, Y. Li, and F. Zanetti (2024). The causal effects of global supply chain disruptions on macroeconomic outcomes: evidence and theory.

Barnichon, R. (2010). Building a composite help-wanted index. *Economics Letters 109*(3), 175–178.

Barro, R. and F. Bianchi (2023, November). Fiscal influences on inflation in OECD countries, 2020-2022. Technical report, Cambridge, MA.

Benigno, P. and G. Eggertsson (2023). It's baaack: The surge in inflation in the 2020s and the return of the non-linear phillips curve. Technical report, CEPR Discussion Papers.

Bergholt, D., F. Canova, F. Furlanetto, N. Maffei-Faccioli, P. Ulvedal, et al. (2023). What drives the recent surge in inflation? the historical decomposition rollercoaster. Technical report, Working Paper.

Bernanke, B. and O. Blanchard (2024). An analysis of pandemic-era inflation in 11 economies. *Peterson Institute for International Economics Working Paper*, 24–11.

Bernanke, B. and O. Blanchard (2025). What caused the us pandemic-era inflation? *American Economic Journal: Macroeconomics 17*(3), 1–35.

Blanchard, O. (2021). In defense of concerns over the $1.9 trillion relief plan.

Cochrane, J. H. (2021). The fiscal theory of the price level: An introduction and overview. *Journal of Economic Perspectives*.

Colonna, F., R. Torrini, and E. Viviano (2023). The profit share and firm markup: how to interpret them? *Bank of Italy Occasional Paper* (770).

Comin, D. A., R. C. Johnson, and C. J. Jones (2023). Supply chain constraints and inflation. Technical report, National Bureau of Economic Research.

English, B., K. Forbes, and A. Ubide (2024). Monetary policy responses to the post-pandemic inflation.

Fornaro, L. (2024). *Fiscal stimulus with supply constraints.* BSE, Barcelona School of Economics.

Gagliardone, L. and M. Gertler (2023). Oil prices, monetary policy and inflation surges.

Galí, J. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications.* Princeton University Press.

Giannone, D. and G. Primiceri (2024). The drivers of post-pandemic inflation. *CEPR Discussion Paper No. 19377*.

Guerrieri, V., G. Lorenzoni, L. Straub, and I. Werning (2021). Monetary policy in times of structural reallocation. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2021-111).

Guerrieri, V., M. Marcussen, L. Reichlin, and S. Tenreyro (2023). Geneva report: Inflation and relative prices. *International Center for Monetary and Banking Studies (ICMB), Center for Economic and Policy Research (CEPR) Report, May*.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press.

Hazell, J. and S. Hobler (2024). Do deficits cause inflation? a high frequency narrative approach.

Jørgensen, P. L. and S. H. Ravn (2022). The inflation response to government spending shocks: A fiscal price puzzle? *European Economic Review 141*, 103982.

Lorenzoni, G. and I. Werning (2023). Wage price spirals. *Forthcoming, Brookings Papers in Economic Activity*.

Mori, L. (2025). Fiscal shocks and the surge of inflation. *Mimeo*. JMP.

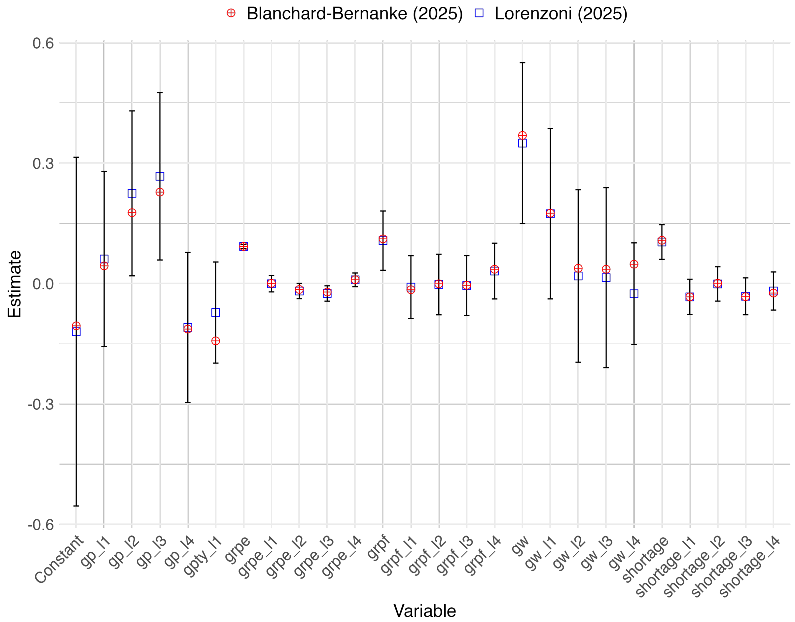
Pisani, M. and A. Tagliabracci (2023). What caused the post-pandemic inflation in italy? an application of bernanke and blanchard (2023). *Questioni di Economia e Finanza- Banca d'Italia Occasional Papers*.

Sargent, T. J., N. Wallace, et al. (1981). Some unpleasant monetarist arithmetic. *Federal reserve bank of minneapolis quarterly review 5*(3), 1–17.

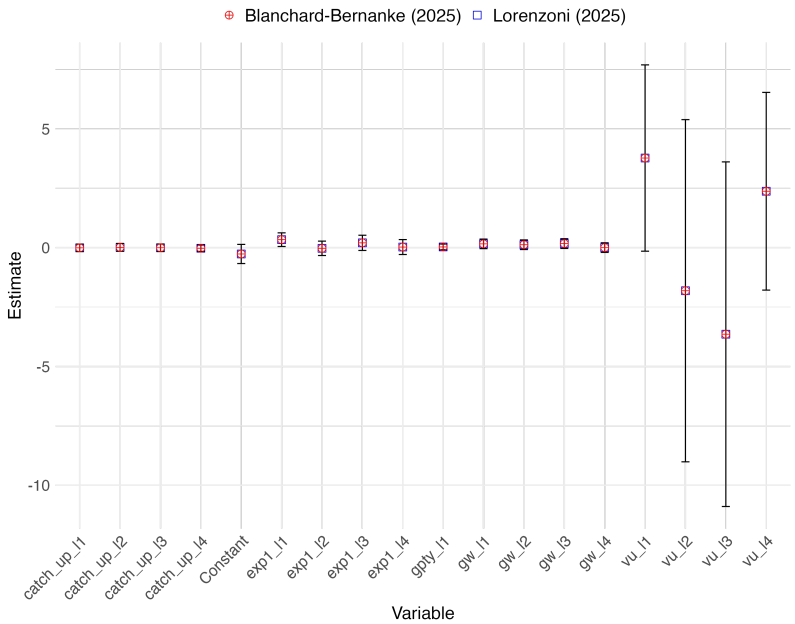
Sims, C. A. (2024). Origins of us inflation. In *AEA Papers and Proceedings*, Volume 114, pp. 90–94. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

APPENDIX

1.

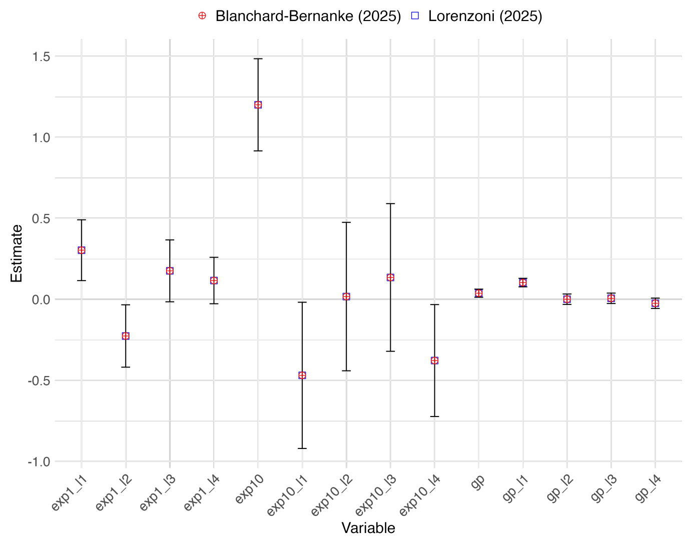


(A) Price equation

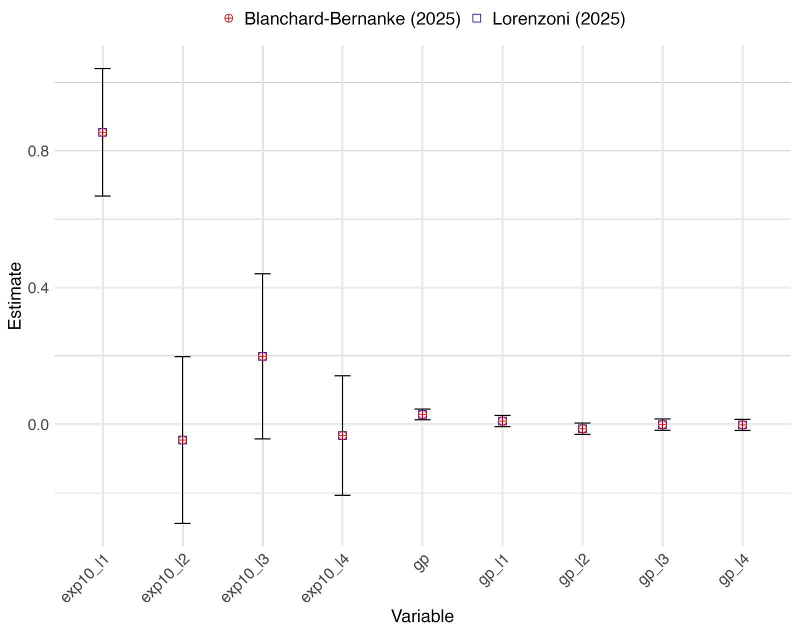


(B) Wage equation

FIGURE 15: Replication coefficients for U.S. (i). Confidence intervals refer to my estimates. Sample period is 1990Q1:2019Q4 for the wage equation and 1990Q1:2023Q1 for the price equation–as in Bernanke and Blanchard (2025).



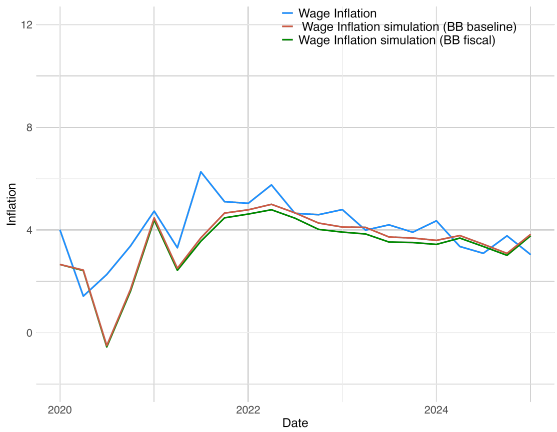
(A) 1y expectation equation



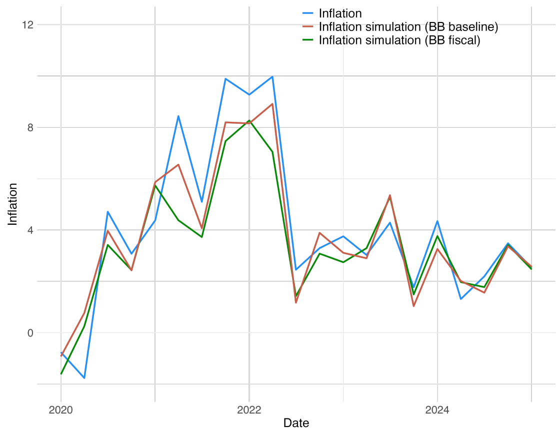
(B) 10y expectation equation

FIGURE 16: Replication coefficients for U.S. (ii). Confidence intervals refer to my estimates. Sample period is 1990Q1:2019Q4.

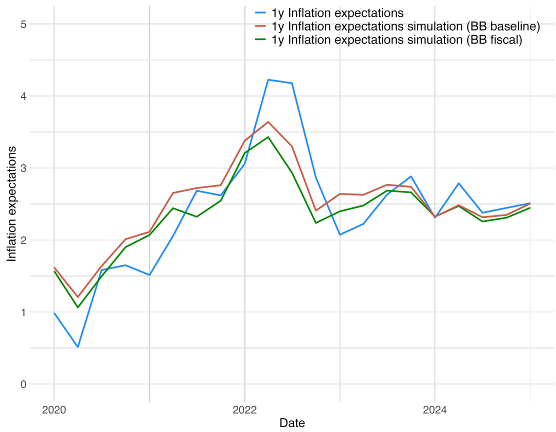
2.



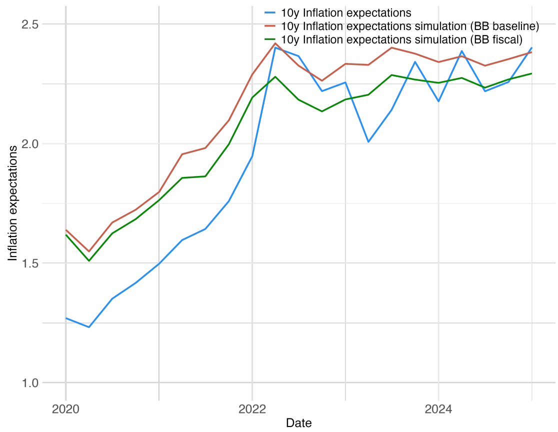
(A) Wage equation



(B) Price equation



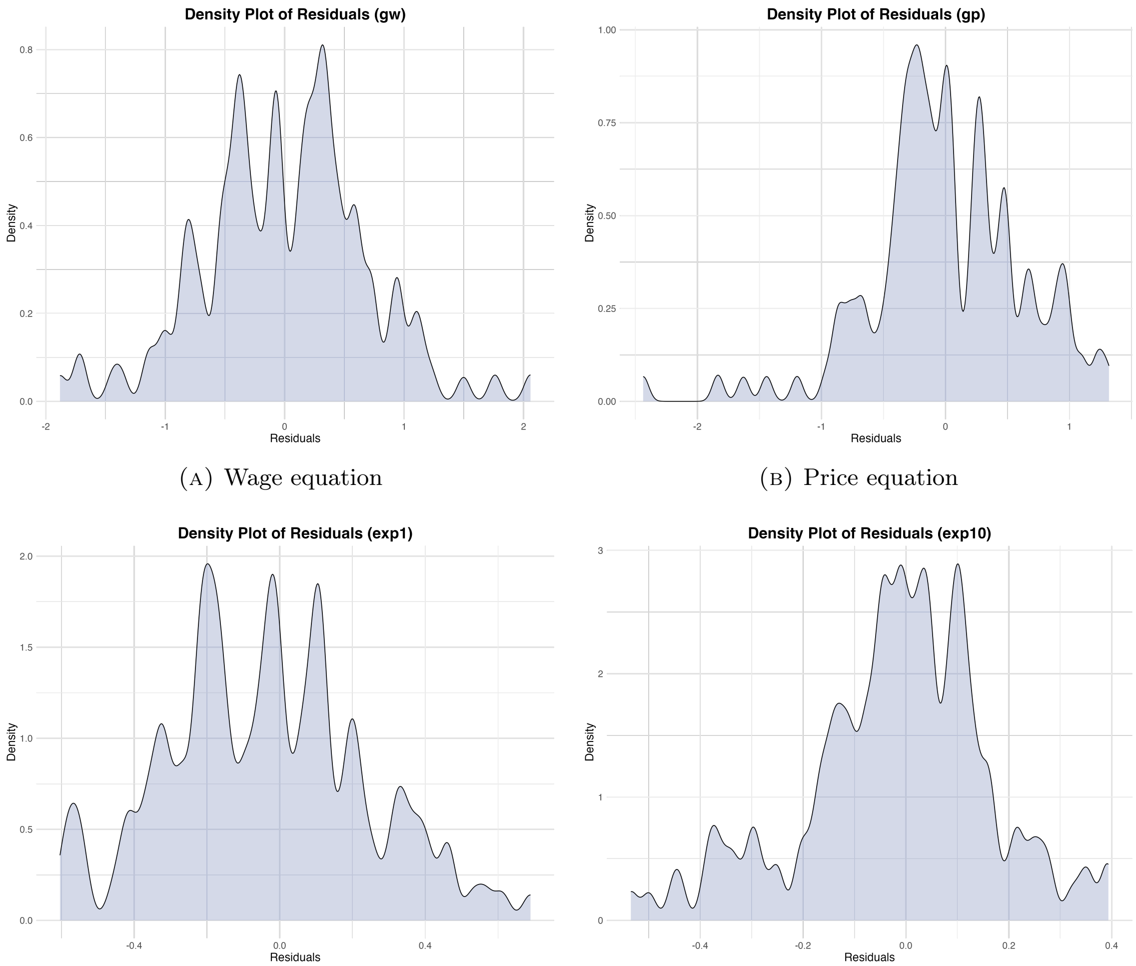
(C) 1y expectations equation



(D) 10y expectations equation

FIGURE 17: Dynamic simulation of the system. Simulations run from 2020Q1:2025Q1. Blue lines refer to actual values, orange lines refer to BB baseline simulation, and green lines refer to simulations of the system extended with a shortages equation.

3.



(A) Wage equation

(B) Price equation

(C) 1y expectations equation

(D) 10y expectations equation

FIGURE 18: Distribution of the residuals of each of the four equations.

4.

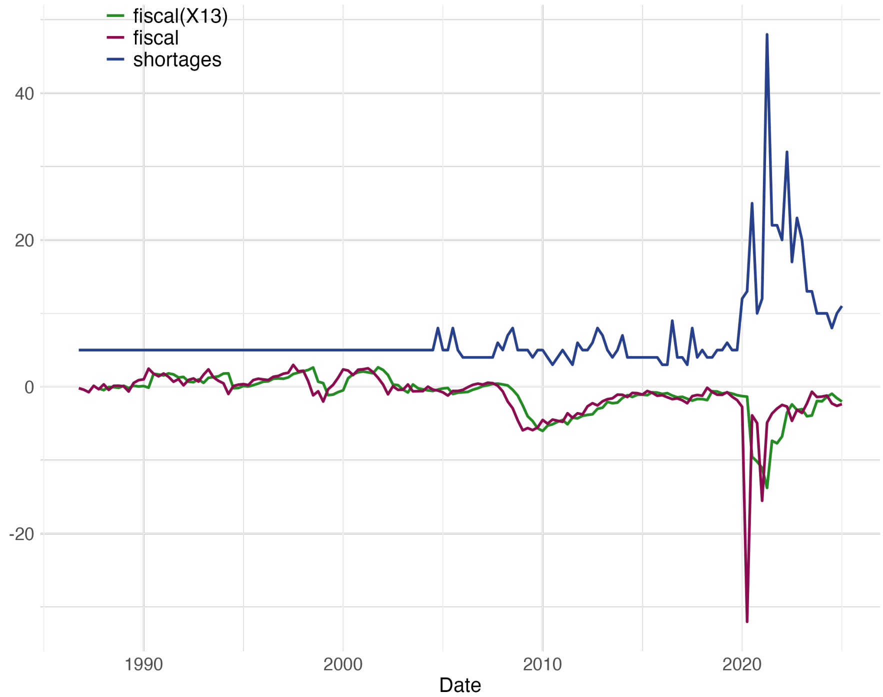
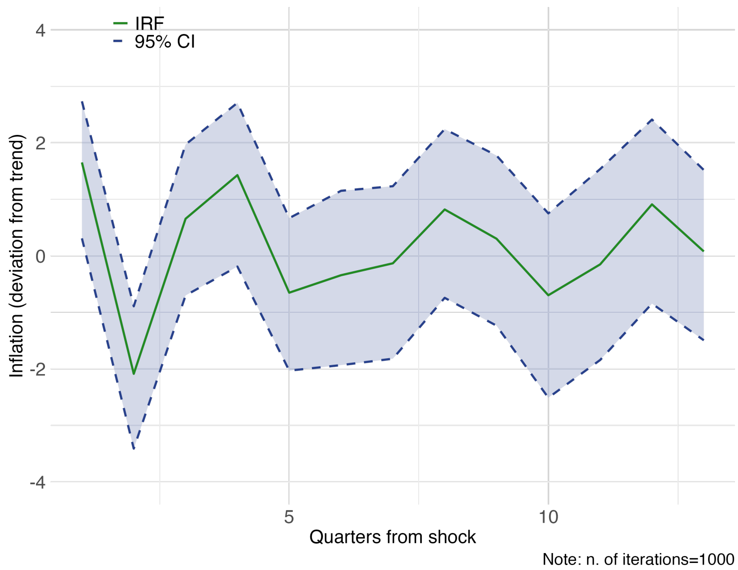
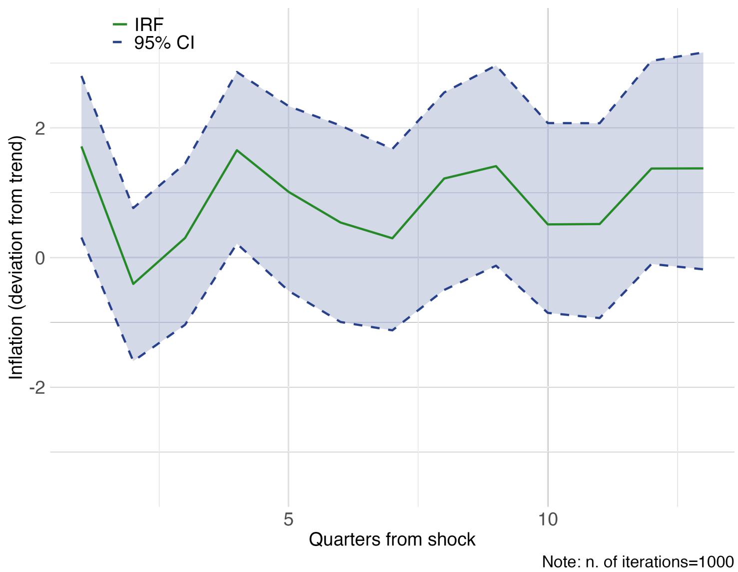


FIGURE 19: Fiscal defict/GDP ratio (Source: FRED) and shortages index (Source: Google Trend). There are two different seasonal adjustments for the fiscal variable. Rolling mean (green) and an adjustment implemented through X13(purple). The model was estimated employing the rolling-mean adjusted series.

5.



(A) 1sd temporary shock

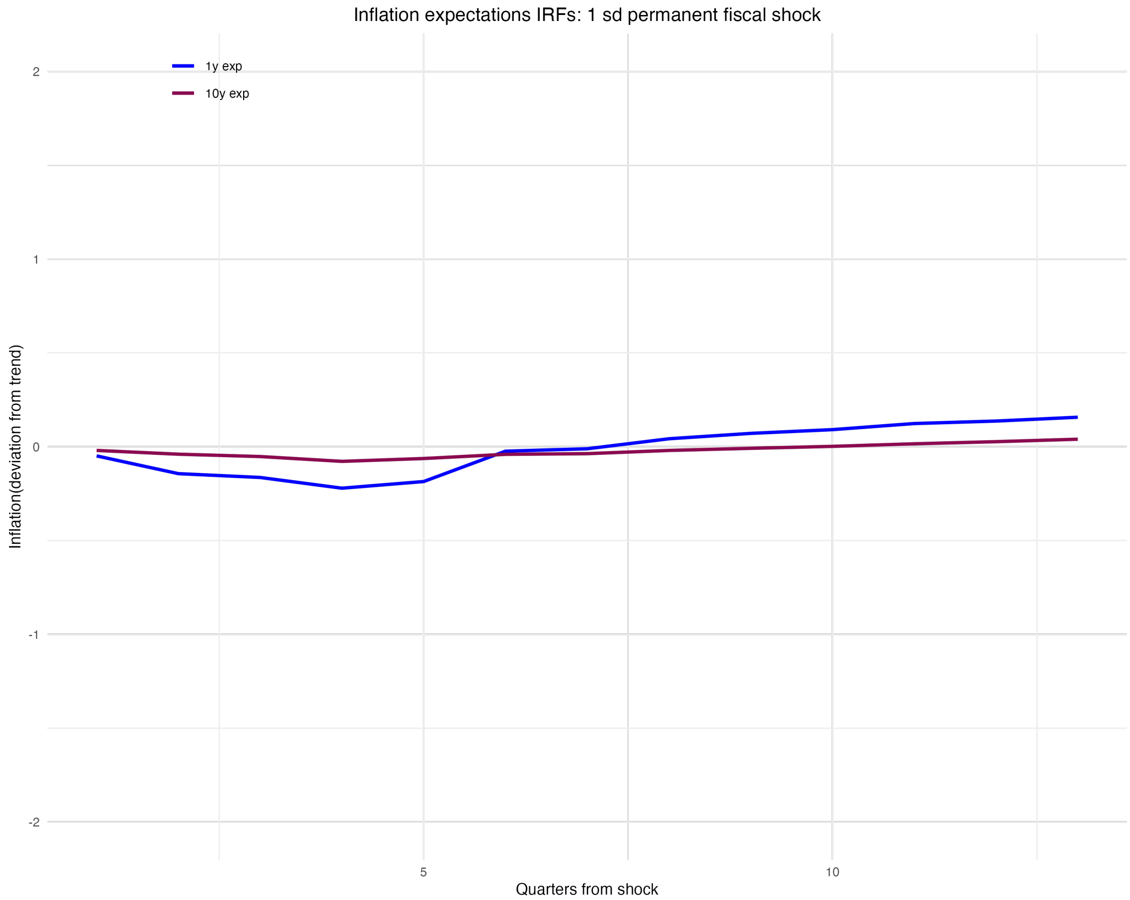


(B) 1sd permanent shock

FIGURE 20: Impulse responses of price inflation to a 1 standard deviation temporary (Panel A) and permanent (Panel B) fiscal deficit shock. Shaded areas represent 95% confidence intervals derived through bootstrapping.

6.

Figure 21: IRF of inflation expectations (long and short term) to a 1sd permanent fiscal deficit shock.



7.

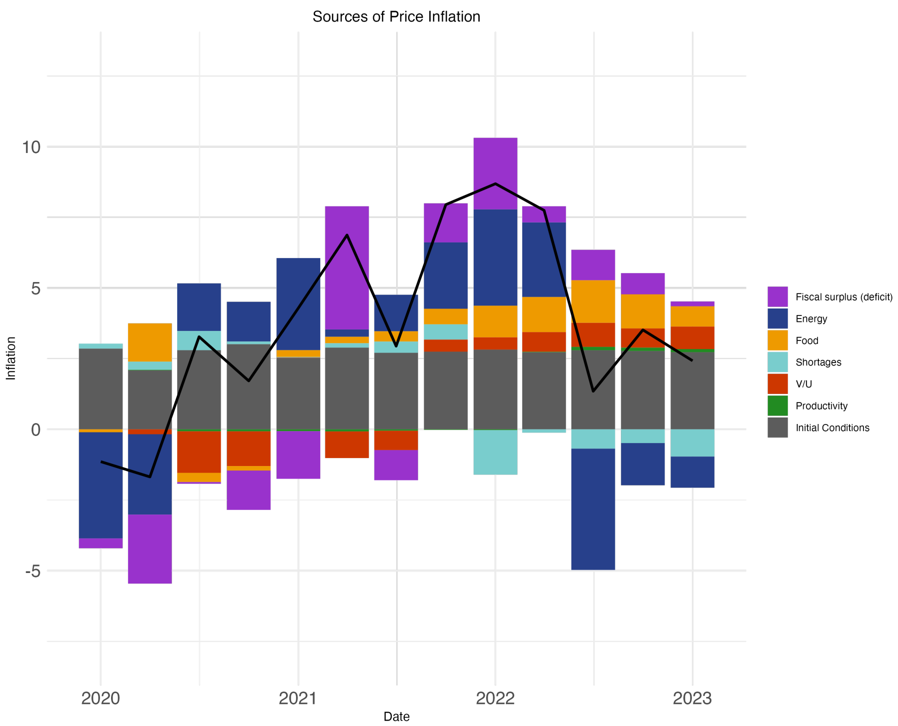


FIGURE 22: Price inflation historical decomposition. Model with "rough" inclusion of the fiscal variable on Right-Hand side of the price equation (as in (7)). Each bar represents the fully dynamic–both direct and indirect–contribution of each component to price and wage inflation. The black line represents the simulated path of inflation.

The price equation was modified as

$$\begin{aligned}\Delta p_t &= \sum_{i=1}^{4}\gamma_{1,i}\Delta p_{t-i} + \sum_{i=0}^{4}\gamma_{2,i}\Delta w_{t-i} + \sum_{i=0}^{4}\gamma_{3,i} grpe_{t-i} \\ &+ \sum_{i=0}^{4}\gamma_{4,i} grpf_t + \sum_{i=0}^{4}\gamma_{5,i}\text{ shortage }_t + \gamma_6\text{ gpty }_{t-1} + \sum_{i=0}^{4}\gamma_{7,i}\text{ fiscal }_t + u_t\end{aligned} \tag{7}$$

	Price equation
Constant	−0.15 (0.21)
gp_l1	0.05 (0.10)
gp_l2	0.20 (0.10)**
gp_l3	0.32 (0.10)***
gp_l4	−0.12 (0.09)
gw	0.37 (0.10)***
gw_l1	0.19 (0.10)*
gw_l2	0.01 (0.11)
gw_l3	−0.01 (0.11)
gw_l4	−0.02 (0.06)
grpe	0.09 (0.00)***
grpe_l1	0.00 (0.01)
grpe_l2	−0.02 (0.01)*
grpe_l3	−0.03 (0.01)***
grpe_l4	0.01 (0.01)
grpf	0.11 (0.04)***
grpf_l1	0.00 (0.04)
grpf_l2	0.00 (0.04)
grpf_l3	−0.02 (0.04)
grpf_l4	0.02 (0.03)
gpty_l1	−0.07 (0.06)
shortage	0.11 (0.02)***
shortage_l1	−0.04 (0.02)*
shortage_l2	−0.00 (0.02)
shortage_l3	−0.04 (0.02)**
shortage_l4	−0.00 (0.02)
eq1: R^2	0.94
eq1: Adj. R^2	0.93
Num. obs. (total)	141

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 7: Price equation with "rough" inclusion of the fiscal variable on the Right-Hand side. Sample period is 1990Q1:2023Q1. A homogeneity restriction was imposed in order to ensure that past price inflation and wage inflation sum to one. Standard errors in parentheses.

REGIONAL MINIMUM WAGES AND LABOR MISALLOCATION IN ITALY

Gabriele Nespoli[†]

DOI: 10.82029/2025013

Abstract. Using a large panel of Italian firms and building on previous work by Caselli et al. (2023), I estimate the *employment-maximizing* minimum wage in Italy in 2022 to be €13.85 per hour. This minimum wage level would feature highly heterogeneous impacts on local labor markets, with large disemployment effects in the South and little to no bite in the North. To address this spatial imbalance, I estimate region-specific *optimal* minimum wages for the same year, finding results ranging from €7.46 per hour in Sardinia to €18.73 per hour in the autonomous province of Bolzano. Introducing regional minimum wages in Italy reduces spatial heterogeneity in employment effects *vis à vis* the national minima, strengthening the link between idiosyncratic productivity and employment gains at the local (province) level. Additionally, I document a sizeable rise in firms' monopsony power between 2019 and 2022, underpinned by labor productivity growing faster than average wages over the same period.

JEL classification: C33, J31, J38, J42, R58

Keywords: Minimum Wage, Monopsony, Regional Wage Policy, Labor Misallocation.

1. INTRODUCTION

In the past months, the Italian political debate has devoted some attention to the proposed introduction of a national minimum wage. In particular, a minimum compensation level of 9 euros was discussed by the Italian parliament in autumn 2023 (Camera dei Deputati, 2023). As scrutiny has been stalling in the Senate, Italy remains one of the only five European Union member states without a statutory minimum wage in place, together with Austria, Denmark, Finland, and Sweden (Eurofound, 2024).

In this paper, I investigate what would be the optimal (employment maximizing) level of minimum wage to be introduced in Italy. That is the value at which disemployment effects are minimized and positive employment effects are maximized. The estimation features both a national minimum wage level and region-specific minimum wage levels, setting different minimum hourly compensations across the 21 Italian regions.

Using a large panel of manufacturing firms' balance-sheet data, I construct a measure of firm labor market power starting from the output elasticities of labor and materials and inputs' revenue shares. All these variables are retrieved from a translog production function, estimated following a large body of literature that uses inputs to control for unobservables (i.e., firm-level productivity) (Levinsohn and Petrin, 2003; Wooldridge, 2009; Ackerberg et al., 2015; Petrin and Levinsohn, 2012). Conceptually, this coefficient indicates if average within-firm wages are above or below the firm's marginal productivity, allowing to determine the market regime (efficient bargaining, perfect competition, or monopsony) in which it operates.

[†]Nespoli: gabriele.nespoli@studbocconi.it. I am grateful to Tito Boeri for his guidance and support. I thank Mauro Caselli for help in replicating his paper's methodology, and the editor, two anonymous referees, Leonardo Antonini, Francesco Doga, Matilde Dolfato, Luigi Lorenzoni, Giulio Savaré, Francesco Schivardi, Federico Pepe, Battista Severgnini, Katherine Jackson, Eric Gao and participants at the 24th Carroll Round conference at Georgetown University and the 15th fRDB Workshop at the University of Florence for useful comments and feedback.

From this I uncover a sizeable rise in monopsony power among Italian firms from 2019 onward and a correspondent decline in the share of firms operating under efficient bargaining. In 2022 roughly two-thirds of manufacturing firms operated in a monopsony setting, while only one third of them were classified in the efficient bargaining regime.

I then categorize Italian enterprises based on productivity, average wages, labor market power, and hypothetical minimum wage levels. This classification allows to retrieve the optimal minimum wage as the value minimizing the share of firms that would be characterized by negative employment effects and maximizing the share of firms that would increase their employment in response to its introduction. This optimal minimum wage was found to be relatively stable in the first period under analysis before significantly rising after 2020 and overcoming the national median wage in 2022. I attribute this result to post-pandemic increases in productivity–which affected firms of all sizes–and to a sizeable increase in the average firm size between 2020 and 2022.

Based on concerns of heterogeneous employment effects on local labor markets, I then apply the optimization procedure at the regional level, retrieving optimal region-specific minimum wage values for the year 2022. A spatial analysis of the potential effects of their introduction reveals how regional minimum wages would increase efficiency in the whole country, relegating heterogeneity (in employment effects) at the regional level and strengthening the negative relation between local productivity and province-level disemployment effects.

The remainder of the paper is structured as follows. Section 2 summarizes the main findings of the literature regarding optimal minimum wages. Section 3 details the data used and the dataset-construction procedure. Section 4 introduces the theoretical framework employed in the analysis, from the production function estimation to the optimization procedure that allows to uncover the optimal minimum wage value. Section 5 presents the main findings, articulated in national and regional minimum wage levels. Section 6 concludes.

2. LITERATURE REVIEW

The effectiveness and broader economic impacts of the minimum wage have been the subject of extensive empirical and theoretical research, with debates centering on its effects on employment, wage distribution, firm behavior, and overall economic efficiency (Dube and Lindner, 2024).

The economics literature broadly supports its use as a policy tool to raise the earnings of low-wage workers and reduce wage inequality, with generally modest effects on aggregate employment when set at moderate levels (Dube, 2019; Arribas Cámara et al., 2024; Dube and Zipperer, 2024; Dube and Lindner, 2024; Cengiz et al., 2019). However, the risk of job loss is higher for the least skilled and in sectors with limited ability to adjust (Neumark and Wascher, 2006).

Of the huge literature on minimum wages, only a rather small and recent part focuses on the determination of their *optimal* level. In fact, most studies on the matter chose to analyze employment effects of the minimum wage.

Central to this literature is what Manning (2021) calls the "New Minimum Wage Research," often regarded as taking its roots in the seminal paper by Card and Krueger (1994) and continuing in more recent times with works such as the one from Harasztosi and Lindner (2019), who estimate pass-through effects to consumers of the Hungarian minimum wage.

This paper goes in a different direction, studying aggregate employment effects of the minimum wage not in a real-world scenario but in the hypothetical case of its introduction in Italy, which currently lacks such a policy. Specifically, employment effects are not studied *per se*, but with the aim of determining an optimal minimum wage as the compensation level minimizing disemployment effects in the country, following the methodology set out by Caselli et al. (2023).

I thus follow a growing albeit small literature attempting to estimate *optimal* minimum wage levels. Ahlfeldt et al. (2018), for example, study the regional effects of the introduction of a statutory minimum wage in Germany in 2015. They show that the measure led to spatial wage convergence without reducing relative employment in low-wage regions. In a subsequent paper, they find that an *employment-maximizing* regional minimum wage would need to be set at 50% of the regional median wage in Germany. This policy would increase aggregate employment by 1.1% (Ahlfeldt et al., 2022).

Studying the interaction between federal and state policy with regard to minimum wages, Simon and Wilson (2021) conclude that decentralized and centralized policy setting exhibit strategic complementarity, and that joint policy setting leads to a small welfare gain over centralization.

The following paper aims to contribute to the minimum wage policy debate dialoguing with Caselli et al. (2023), who find that the optimal (national) minimum wage level ranged from 8.25 to 9.65 euro per hour in Italy in 2018.

Specifically, this paper makes three contributions to the existing literature: (i) it studies the spatially heterogeneous effects resulting from the potential introduction of a national minimum wage in Italy; (ii) it applies the procedure developed by Caselli et al. (2023) to retrieve region-level optimal minimum wages for Italy in 2022; and (iii) it expands the results by Caselli et al. (2023) regarding the national optimal minimum wage and the evolution of monopsony power, leveraging newly available firm-level data spanning from 2015 to 2022.

3. DATA

Firm-level yearly data come from the database Orbis by Bureau van Dijk and comprise balance sheet information for around 6 million Italian firms.[1] Data on operating revenues, number of employees, non-current assets (used as a proxy for physical capital), expenditure on materials, labor costs, and the industrial sector of activity (according to the EU statistical classification of economic activities NACE Rev. 2)[2] were retrieved for the years 2015–2022.

The data have been cleaned to avoid the influence of extreme values. Observations showing a growth rate greater than 400% or lower than −80% for both the variable *revenues* and all inputs (*materials, employees,* and *capital*) were thus deleted. Observations featuring missing values for all the variables of interest or a sum of labor costs and material costs greater than total revenues were deleted, as well. Around 3,375,000 firms passed this first cleaning procedure.

As a last step, I computed the shares of labor costs and material costs on total revenues and deleted the observations where the value of at least one of these shares was either lower than 5%, higher than 200%, or missing. Companies featuring non-consecutive year-level observations are also deleted, together with the ones for which fewer than five year-level observations are available. This last step left 250,190 distinct firms in the dataset, each featuring from 5 to 8 consecutive yearly observations, and the main analysis was restricted to this sub-group.

Sector-specific price indexes, revenues, and output data are retrieved from the ISTAT website.[3]

4. METHODOLOGY

4.1. Labor Market Power

The analytical framework employed in this paper closely follows the one developed by Caselli et al. (2023).

A firm-level time-varying measure of labor market power, ϕ_{it}, is defined as follows:

$$\phi_{it} = \frac{P_{it}^L}{MRP_{it}^L} \tag{1}$$

[1]Due to the large size of the dataset, data are obtained through an iterated automated download of many .csv files, which are later appended together by Stata.

[2]Note that the more recent (2023) classification NACE Rev. 2.1 was not used because both Orbis and Istat data were still classified according to NACE Rev. 2 at the time of the analysis.

[3]Please refer to appendix C for a detailed description of data on price indexes and revenues.

where P_{it}^{L} is the average labor cost paid by firm i at time t, and MRP_{it}^{L} is the marginal revenue product of labor, a measure of firm-specific productivity. If $\phi = 1$, the labor market is competitive, $\phi < 1$ indicates a labor market regime characterized by monopsony power, and $\phi > 1$ signals some degree of market power by firms' employees (efficient bargaining).

Following Mertens (2019, 2020) and Caselli et al. (2021), ϕ_{it} is rewritten in the following form:

$$\phi_{it} = \frac{\theta_{it}^{M}/\alpha_{it}^{M}}{\theta_{it}^{L}/\alpha_{it}^{L}} \tag{2}$$

where θ_{it}^{M} and θ_{it}^{L} are, respectively, the output elasticities of materials and labor, while α_{it}^{M} and α_{it}^{L} are the shares of materials and labor costs over total revenues.[4]

While the input shares α_M and α_L can be computed starting from the available data, the output elasticities θ_M and θ_L need to be estimated starting from a suitable production function.

4.2. Production Function Estimation

In order to compute unbiased estimates of the output elasticities, I follow a large literature using inputs to control for unobservables in production function estimations. (Levinsohn and Petrin, 2003; Wooldridge, 2009; Ackerberg et al., 2015; Petrin and Levinsohn, 2012)

As inputs and unobserved productivity shocks are potentially correlated, an endogeneity bias might arise in estimating the production function. I thus employ a Wooldridge–Levinsohn–Petrin (henceforth WLP) estimator to address the simultaneity issue. The WLP estimator does not assume constant returns to scale; is robust to the Ackerberg, Caves, and Frazer (ACF) critique (Ackerberg et al., 2015); and is programmed as an instrumental variable (IV) estimator. This approach consists of including lags of inputs to proxy for productivity and is detailed below.

A first step is run to get rid of the pure error term in the measure of revenues and markups De Loecker and Warzynski (2012):

$$q_{it} = g(l_{it}, k_{it}, m_{it}) + \varepsilon_{it} \tag{3}$$

where q it is the natural logarithm of deflated revenues of firm i at time t; and l_{it}, k_{it}, and m_{it} are the logarithms of labor, capital, and materials used by the firm, respectively. Specifically, equation (3) is estimated for every different sector, using a third-order polynomial on all inputs with year fixed-effects. This allows to separately identify expected output $\hat{q}_{it}$ and residuals $\hat{\varepsilon}_{it}$.

Then, a general production function is used:

$$\hat{q}_{it} = f_s(l_{it}, k_{it}, m_{it}, \mathbf{B}) + w_{it} + \eta_{it} \tag{4}$$

where $\mathbf{B}$ is the parameter vector containing the marginal effects and needs to be estimated, w_{it} is the firm-specific productivity (only observable by the firm), and η_{it} is the unobservable error term.

The function f_s is a revenue function, changes across different sectors and is assumed to be translog, which allows to obtain firm-level time-variant output elasticities:[5]

$$\begin{aligned} f_s = \alpha + \beta_L l + \beta_K k + \beta_M m + \beta_{L^2} l^2 + \beta_{M^2} m^2 \\ + \beta_{K^2} k^2 + \beta_{KL} kl + \beta_{KM} km + \beta_{LM} lm \end{aligned} \tag{5}$$

Labor is assumed to be a variable input and firms are assumed to be costs minimizers. Labor, materials, and their interactions are assumed to be endogenous and are thus instrumented with

[4]Please refer to Caselli et al. (2021) for the details behind this rewriting.

[5]With regard to the restrictions outlined by Christensen et al. (1973), the coefficients are symmetric by construction, while the estimated coefficients are not commodity-wise additive, confirming that aggregate input is not Cobb-Douglas.

the first and second lags of labor, and with the second lags of materials and capital.[6] This choice is justified by the assumption of dynamic input adjustment (in response to a productivity shock).

Estimation is then carried out via a two-step efficient generalized method of moments (GMM), which provides an estimate for the parameter vector $\mathbf{B}$.[7] Once marginal effects are retrieved from the vector, output elasticities of inputs can be computed for every firm-year observation:

$$\hat{\theta}_{it}^{M} = \hat{\beta}_M + 2\hat{\beta}_{M^2} m_{it} + \hat{\beta}_{KL} k_{it} + \hat{\beta}_{LM} m_{it} \tag{6}$$

$$\hat{\theta}_{it}^{L} = \hat{\beta}_L + 2\hat{\beta}_{L^2} l_{it} + \hat{\beta}_{KM} k_{it} + \hat{\beta}_{LM} l_{it} \tag{7}$$

$$\hat{\theta}_{it}^{K} = \hat{\beta}_K + 2\hat{\beta}_{K^2} k_{it} + \hat{\beta}_{KL} l_{it} + \hat{\beta}_{KM} m_{it} \tag{8}$$

where $\hat{\theta}_{it}^{M}$ is the estimated output elasticity of materials, $\hat{\theta}_{it}^{L}$ is the estimated output elasticity of labor, and $\hat{\theta}_{it}^{K}$ is the estimated output elasticity of capital.

4.3. **Labor Market Regimes**

Classification of firms in different regimes is based on previous work by Dobbelaere and Mairesse (2013). Once the relevant input elasticities are obtained from the production function, I use equation (2) to derive firm-level time-varying estimates of labor market power ϕ_{it}.

However, the input shares α_M and α_L are not used directly as observed in the data, but corrected *à la* De Loecker and Warzynski (2012). Specifically, a correction is needed because data on deflated revenues include an error unobservable to both the econometrician and the firm. I thus use equation (3) to obtain estimates of the error term and correct the revenue shares multiplying them by the exponential of the estimated error term.[8] As explained by Caselli et al. (2023), this adjustment cleans the revenue shares from any variations in output that are not related to variables affecting input demand.

Firm-year observations are classified into three different labor market regimes, according to the value of ϕ_{it}:

- Efficient Bargaining (EB) if $\phi_{it} > 1$. In this case the average available wage is greater than marginal productivity (as inframarginal gains are distributed across workers) and employees hold some degree of market power *vis à vis* the employer.

- Perfect Competition (PC) if $\phi_{it} = 1$. In this case the average available wage equals marginal productivity.

- Monopsony (MO) if $\phi_{it} < 1$. In this case the average available wage is lower than marginal productivity, and the firm holds some degree of labor market power towards the employees.

More precisely, a confidence interval is constructed to operate the firm classification. I start by defining ψ, a measure of labor market power that is conceptually equivalent to ϕ: [9]

$$\psi = \frac{\theta_M}{\alpha_M} - \frac{\theta_L}{\alpha_L} \tag{9}$$

where positive values of ψ correspond to values of ϕ which are > 1, and vice versa.

Then, I define a 95% confidence interval for ψ_{it}:

$$\text{C.I.} = \hat{\psi}_{it} \pm 1.96 \cdot \text{Var}(\psi_{it}) \tag{10}$$

[6] See Appendix A.1 for a more detailed description of the production function.
[7] Estimation was carried out by Stata/SE 18.0 with the user-contributed module *ivreg2* (Baum et al., 2002).
[8] Mathematically: $\alpha_X^{corr} = \alpha_X \cdot \exp(\hat{\varepsilon})$, where α_X is the output elasticity of input variable X.
[9] This choice follows Caselli et al. (2021) and permits a simpler characterisation of the confidence interval.

and classify each observation accordingly:

- Efficient Bargaining if $\hat{\psi}_{it} > 0$ and $0 \notin$ C.I.
- Perfect Competition if $0 \in$ C.I. Here, $\hat{\psi}_{it}$ is not significantly different from zero.
- Monopsony if $\hat{\psi}_{it} < 0$ and $0 \notin$ C.I.

See Appendix A.2 for details on the computation of $\text{Var}(\psi_{it})$.

4.4. **Optimal Minimum Wage**

Based on the degree of labor market imperfections ϕ_{it}, firm-specific average yearly available wages ($w_{it} = P^L_{it}$), the marginal revenue product of labor MRP^L_{it}, and a hypothetical minimum wage MW, firms can be classified into four different categories:

- Category 1: if $w_{it} < MW$, $MRP^L_{it} < MW$ and $\phi_{it} \geq 1$
- Category 2: if $w_{it} < MW$, $MRP^L_{it} < MW$ and $\phi_{it} < 1$
- Category 3: if $w_{it} < MW$, $MRP^L_{it} > MW$ and $\phi_{it} < 1$
- Category 4: if $w_{it} \geq MW$

Given a hypothetical yearly minimum wage MW, its introduction might cause heterogeneous effects on the employment levels of single firms, depending on which category that firm belongs to.

Firms in cat. 1 are characterized by wages and productivity which are low with respect to the hypothesized minimum wage and a labor market regime of perfect competition (PC) or efficient bargaining (EB). A minimum wage introduction is likely to reduce the labor demand of these firms, possibly increasing layoffs and thus reducing total welfare.

Firms in cat. 3 are characterized by marginal productivity levels that are higher than the hypothesized minimum wage, which in turn is higher than average wages, and operate in a regime of monopsony (MO). A minimum wage introduction is likely to increase labor supply without affecting labor demand, possibly increasing both employment and total welfare.[10] Evidently, the introduction of a minimum wage would be efficient for these firms.

Firms in cat. 2 are characterized by wages and productivity that are low with respect to the hypothesized minimum wage and a labor market regime of monopsony (MO). While a minimum wage introduction is likely to reduce the labor demand of these firms, with negative welfare effects, as it was the case for firms in cat. 1, it also has the potential of reducing monopsony power, generating positive welfare effects. The net welfare effect of a minimum wage introduction on these firms is then ambiguous, and for this reason they are excluded from the analysis carried out below.

Finally, firms in cat. 4 are characterized by average available wages that are already equal or higher than the hypothesized minimum wage. The net welfare effect of a minimum wage introduction on these firms is likely to be both small and unrelated to productivity. For this reason, also these firms are excluded from the main analysis.

4.4.1. *Yearly Minimum Wage* For every year t in the sample, the optimal (national) minimum wage MW^*_t is then defined as the value minimizing the share of firms in category 1 and maximizing the share of firms in category 3.

Analytically,

$$MW^*_t = \min_{MW} \ \%_{1,t} - \%_{3,t} \tag{11}$$

where $\%_{1,t}$ and $\%_{3,t}$ are, respectively, the shares of firms in cat. 1 and in cat. 3 at time t.

[10]Note that, for this to happen, a redistribution of welfare away from firms and toward workers is implied, with the latter gaining more than the former lose.

Both here and for the regional case (see below), I compute two values for the optimal minimum wage: one based on the above formula using firms shares, and another one using employment shares in the same formula.[11]

Due to heterogeneity concerns and based on previous results by Caselli et al. (2023), it can be well hypothesized that the introduction of a national minimum wage in Italy would lead to mixed results, with the areas of the country characterized by lower productivity experiencing a decline in labor demand, and more productive areas not feeling the bite of the new minimum. This is the main rationale for the pursuit of distinct optimal minimum wage levels for each Italian region, which largely guided the realization of this paper.

Thus, the optimal minimum wage $MW^*_{j,t}$ in region j at time t is defined as follows:

$$MW^*_{j,t} = \min_{MW} \%_{1,j,t} - \%_{3,j,t} \tag{12}$$

where $\%_{1,j,t}$ and $\%_{3,j,t}$ are, respectively, the shares of firms in cat. 1 and in cat. 3 in region j at time t.

4.4.2. *Hourly Minimum Wage* The equations above yield yearly values for the optimal minimum wage, also including social security contributions. As minimum wages are generally computed as gross of income tax but net of social security contributions, a conversion formula is applied to obtain the hourly minimum wage HMW:

$$HMW = \frac{MW \times (1 - SSC)}{168 \times 12} \tag{13}$$

where SSC is the share of social security contributions in the total wage bill, and the yearly minimum wage MW is divided by the average number of hours worked per month (168), times 12 months.[12]

5. RESULTS

5.1. Italian Optimal Minimum Wage

Based on the above methodology, the optimal yearly minimum wage in Italy was determined to be €39,000 in 2022. This amount corresponds to €13.85 per hour according to equation (13).[13] As shown in Figure 1, this is the value that minimizes the difference between the share of firms in cat. 1 and the share of firms in cat. 3 in 2022, with around 15% of firms that would be negatively affected by the minimum wage introduction, and more than 23% of firms on which the impact of the new minimum wage would be positive.[14]

In Table 1, estimates of output elasticities are reported, together with results for the labor market power parameter ϕ, for each of the 18 sectors considered. All of these results are to be interpreted as averages of the 2015–2022 period. The only sector characterized by monopsony is the one of electricity, gas, steam, and air conditioning supply (D), which has the lowest value of ϕ. All other sectors were either in a perfect competition or efficient bargaining regime during the 2015–2022 time span (on average), with the sector of education (P) showing the highest level of market power on the side of employees.

[11] In the second case, the shares of employees working in firms belonging to cat. 1 and to cat. 3 at time t are considered.

[12] As suggested by Caselli et al. (2023), SSC is computed for firms with labor costs around the average ±10%. As an approximation, I thus use a value of $SSC = 0.284$, which is the one found by Caselli et al. (2023) and relative to their 2011-2018 data.

[13] As the optimization problem is solved through an iterative process, the error margin is €250 per year, or €0.089 per hour.

[14] The result based on the weighted firm shares was found to be €63,500, or €22.55 per hour. Further analyses of this result are needed in order to explain the sharp wedge between this value and the median wage in Italy in 2022 (€13.84 per hour).

FIGURE 1: Optimal National Minimum Wage in 2022.

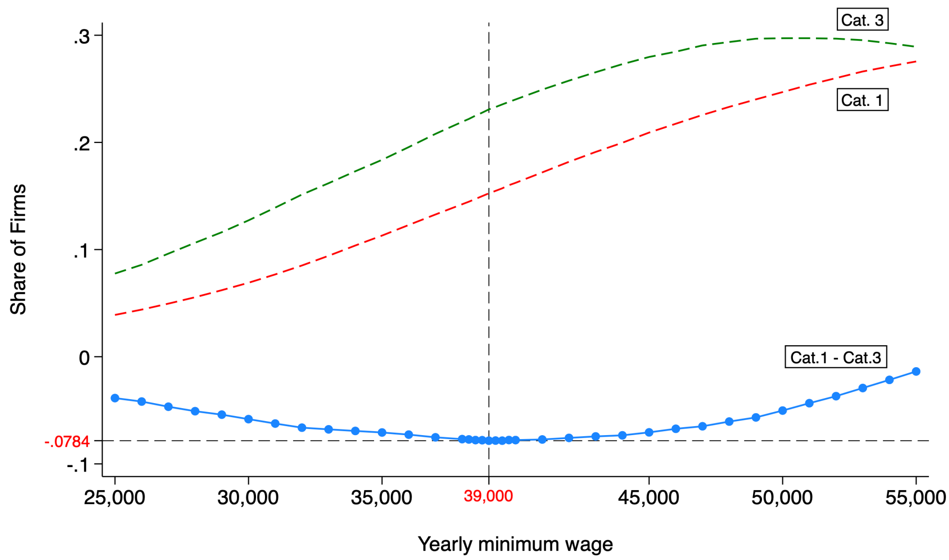


TABLE 1: Output Elasticities and Labor Market Power by NACE sectors.

Sector	N	θ_L	θ_M	θ_K	ϕ
A (Agriculture)	35,930 (2.0%)	0.265 (0.091)	0.584 (0.086)	0.123 (0.033)	2.024 (1.622)
B (Mining)	5,058 (0.3%)	0.425 (0.090)	0.483 (0.061)	0.113 (0.021)	1.224 (0.841)
C (Manufacturing)	564,090 (31.5%)	0.372 (0.108)	0.572 (0.115)	0.068 (0.026)	1.014 (0.596)
D (Electricity)	2,093 (0.1%)	0.355 (0.106)	0.444 (0.118)	0.193 (0.075)	0.775 (0.646)
E (Water and Waste)	15,938 (0.9%)	0.448 (0.024)	0.421 (0.031)	0.157 (0.044)	1.942 (2.034)
F (Construction)	287,130 (16.1%)	0.366 (0.102)	0.578 (0.076)	0.071 (0.022)	1.597 (1.056)
G (Trade)	424,089 (23.7%)	0.221 (0.096)	0.708 (0.122)	0.049 (0.028)	1.002 (0.601)
H (Transportation)	71,075 (4.0%)	0.375 (0.017)	0.513 (0.030)	0.114 (0.042)	2.576 (2.517)
I (Accommodation)	179,292 (10.0%)	0.257 (0.055)	0.623 (0.069)	0.141 (0.071)	2.427 (1.649)
J (Communications)	26,563 (1.5%)	0.459 (0.113)	0.484 (0.073)	0.086 (0.037)	1.516 (1.098)
K (Finance)	4,172 (0.2%)	0.271 (0.107)	0.530 (0.102)	0.218 (0.061)	1.064 (0.559)
L (Real Estate)	14,034 (0.8%)	0.313 (0.089)	0.554 (0.087)	0.114 (0.053)	1.565 (1.220)
M (Professional Act.)	28,179 (1.6%)	0.387 (0.092)	0.493 (0.083)	0.095 (0.033)	1.609 (1.440)
N (Administration)	38,102 (2.1%)	0.360 (0.014)	0.505 (0.019)	0.128 (0.037)	3.786 (3.823)
P (Education)	4,694 (0.3%)	0.286 (0.058)	0.613 (0.083)	0.096 (0.030)	10.227 (9.304)
Q (Human Health)	43,214 (2.4%)	0.313 (0.047)	0.582 (0.053)	0.096 (0.025)	6.270 (6.142)
R (Entertainment)	16,045 (0.9%)	0.211 (0.072)	0.631 (0.009)	0.112 (0.050)	5.759 (6.331)
S (Other Services)	28,862 (1.6%)	0.324 (0.089)	0.590 (0.040)	0.103 (0.034)	3.361 (2.605)
Total	1,788,560 (100.0%)	0.319 (0.116)	0.603 (0.119)	0.080 (0.047)	1.649 (2.076)

NOTES. Sectors O (Public administration and defense; Compulsory social security), T (Activities of households as employers), and U (Activities of extraterritorial organizations and bodies) were excluded from the analysis due to the small number of observations. Standard errors in parentheses.

Following the procedure detailed in section 4.4 and based on employment shares, the optimal yearly minimum wage is estimated for 2022, using both the simple firms shares and then weighting observations on employment. However, the results obtained when considering all sectors are considerably higher than the median wage in 2022 (€13.84 per hour), regardless of the weighting strategy (firms' shares or employment shares).

A solution can be found by restricting the analysis to a single industrial sector of interest, instead of considering all firms in the dataset. Looking at the values of ϕ from Table 1, I identify the sectors characterized, on average, by either monopsony or perfect competition. These are Manufacturing (C), Trade (G), Electricity (D), and Finance (K). Excluding the latter two because of the limited number of available observations, I choose to focus the analysis on the Manufacturing (C) sector. This is both the sector representing the relative majority of firms in Italy and a consolidated choice for labor economics studies in the past literature.

A new production function is thus estimated, considering the sub-sectors of manufacturing in equation (5). The resulting elasticities are displayed in Table 2, together with sub-sector-level estimates of labor market power ϕ.

TABLE 2: Output Elasticities and Labor Market Power by Manufacturing Sub-Sector.

Sub-sector	N	θ_L	θ_M	θ_K	ϕ
10 (Food Products)	47,290 (8.4%)	0.215 (0.049)	0.687 (0.059)	0.100 (0.037)	1.416 (1.062)
11 (Beverages)	6,282 (1.1%)	0.261 (0.030)	0.635 (0.051)	0.101 (0.039)	0.989 (1.043)
13 (Textiles)	16,883 (3.0%)	0.391 (0.129)	0.547 (0.156)	0.066 (0.038)	1.017 (0.491)
14 (Wearing Apparel)	16,858 (3.0%)	0.324 (0.070)	0.629 (0.081)	0.065 (0.027)	1.459 (1.440)
15 (Leather)	15,122 (2.7%)	0.345 (0.149)	0.590 (0.162)	0.077 (0.035)	1.108 (0.530)
16 (Wood and Cork)	19,442 (3.5%)	0.350 (0.120)	0.630 (0.103)	0.046 (0.014)	0.957 (0.524)
17 (Paper Products)	10,578 (1.9%)	0.322 (0.105)	0.632 (0.130)	0.056 (0.020)	0.741 (0.323)
18 (Printing)	17,912 (3.2%)	0.398 (0.100)	0.534 (0.070)	0.070 (0.020)	1.257 (0.786)
19 (Coke and Petroleum)	714 (0.1%)	0.356 (0.141)	0.635 (0.132)	0.064 (0.037)	0.481 (0.168)
20 (Chemical Products)	15,826 (2.8%)	0.312 (0.105)	0.641 (0.116)	0.073 (0.034)	0.793 (0.359)
21 (Pharmaceuticals)	2,022 (0.4%)	0.383 (0.157)	0.523 (0.140)	0.074 (0.032)	0.925 (0.375)
22 (Rubber and Plastic)	31,502 (5.6%)	0.325 (0.100)	0.602 (0.101)	0.063 (0.011)	0.943 (0.425)
23 (Other Mineral Products)	27,368 (4.9%)	0.385 (0.120)	0.588 (0.112)	0.075 (0.028)	0.988 (0.489)
24 (Basic Metals)	9,197 (1.6%)	0.362 (0.087)	0.566 (0.101)	0.067 (0.030)	0.796 (0.550)
25 (Fabricated Metal Products)	135,392 (24.2%)	0.430 (0.105)	0.493 (0.116)	0.083 (0.027)	0.993 (0.504)
26 (Electronic Products)	15,281 (2.7%)	0.383 (0.115)	0.558 (0.108)	0.053 (0.021)	1.052 (0.608)
27 (Electrical Equipment)	20,802 (3.7%)	0.309 (0.102)	0.637 (0.104)	0.047 (0.017)	1.178 (0.561)
28 (Machinery and Equipment)	69,359 (12.4%)	0.347 (0.082)	0.605 (0.107)	0.045 (0.021)	1.014 (0.489)
29 (Motor Vehicles)	7,238 (1.3%)	0.350 (0.127)	0.597 (0.138)	0.067 (0.029)	0.925 (0.404)
30 (Other Transport)	4,999 (0.9%)	0.412 (0.131)	0.557 (0.110)	0.065 (0.029)	1.115 (0.643)
31 (Furniture)	21,667 (3.9%)	0.325 (0.119)	0.641 (0.107)	0.048 (0.012)	1.257 (0.550)
32 (Other Manufacturing)	18,414 (3.3%)	0.378 (0.100)	0.547 (0.087)	0.070 (0.018)	1.144 (0.857)
33 (Repairs and Installations)	30,034 (5.4%)	0.429 (0.129)	0.517 (0.105)	0.064 (0.024)	1.141 (0.763)
Total	560,182 (100.0%)	0.361 (0.121)	0.576 (0.125)	0.069 (0.031)	1.071 (0.678)

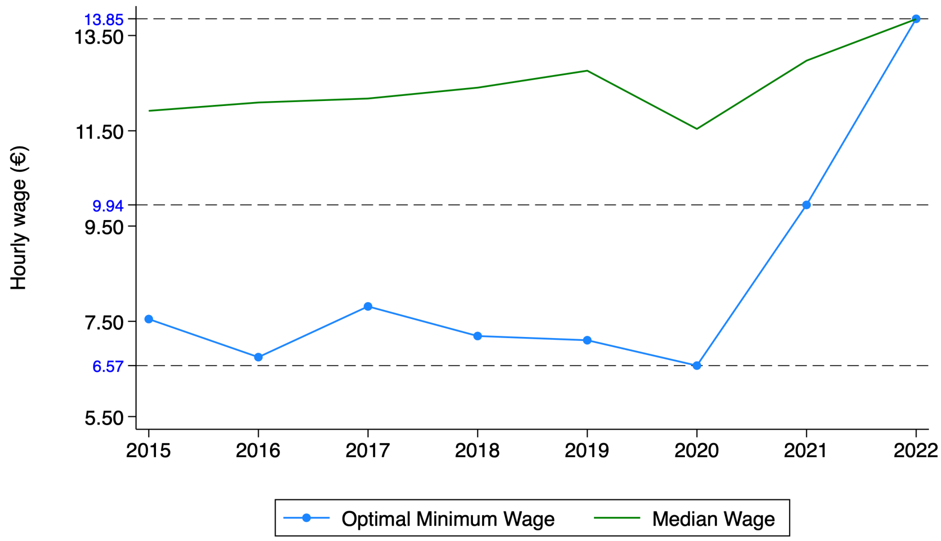
NOTES. Sector 12 (Manufacture of Tobacco Products) was excluded from the analysis due to the small number of observations. Standard errors in parentheses.

The optimal yearly minimum wage in Italy was thus found to be €39,000 in 2022.

Figure 2 illustrates the evolution of the optimal national minimum wage, as computed based on unweighted firm shares, from 2015 to 2022. Here, a rather surprising picture arises, as optimal minimum wages are relatively low and stable between 2015 and 2020, before showing a steep increase of 51% in 2021 and 39% in 2022.[15] In 2022, the estimated optimal minimum wage is €0.01 higher than the median wage. While this result might seem counterintuitive, the analysis of productivity and labor market regimes below will try to assess if it can be justified by monopsony power growth.

[15]This behavior might be partly explained by the "rebound" effect experienced by worldwide economies in the aftermath of the Sars-CoV2 pandemic. However, the reasons behind such a steep increase might be investigated by further research, also employing more recent data from 2023 to 2024.

FIGURE 2: Optimal National Minimum Wage from 2015 to 2022.



To explain such behavior, in Figure 3 I present evidence on the evolution of average labor productivity (as measured by the marginal revenue product of labor) by firm sizes.[16] Following the standard EU definition for small and medium enterprises (SME), microfirms are the ones employing fewer than 10 people and having a turnover below €2 million; small firms have fewer than 50 employees and a turnover below €10 million; medium-sized firms employ fewer than 250 people and do not exceed €250 million in yearly turnover. All remaining firms with values above these thresholds are classified as big firms (European Commission, 2003). Not surprisingly, there is a positive and stable correlation between firm size and labor productivity. From 2020 onwards, we observe a sizeable increase in productivity, with the marginal revenue product of labor increasing on average by 16.7% from 2020 to 2021 and by 17.6% between 2021 and 2022.[17] During both periods, this increase is more pronounced for medium-sized and big firms, while it is less important for microfirms and small firms.

As documented in Figure 4, the average firm size increased between 2020 and 2022.[18] This is due to a 5 percentage-points (pp) reduction in the share of microfirms, contemporaneous to a 3pp increase in the share of medium-sized firms and an increase of 1.5pp for small firms. Big firms also experienced a 1pp increase in their share over the period.

What is the overall relationship between all of these trends and labor market regimes? From Figure 5 we can see that the gap between the shares of firms operating under monopsony (MO) and the ones operating in an efficient bargaining (EB) regime significantly widens from 2019 onward. Between 2021 and 2022 the share of firms in the EB regime declined by 6.25 pp, at the expense of a 6.24pp increase in the share of firms operating under monopsony. In 2022, roughly two-thirds of manufacturing firms were classified in the MO regime, as opposed to less than one-third operating under EB.

[16]See appendix C for a similar graph showing the evolution of labor productivity by macroregion.
[17]Note that, over the 2020–2022 period, the average productivity increased by 37.3%, and the estimated optimal minimum wage increased by 110.8%.
[18]The average firm employed 38.06 people in 2020, 38.78 in 2021, and 40.83 in 2022. The median firm employed 11 people in 2020, 12 in 2021, and 12 in 2022.

FIGURE 3: Labor Productivity by Firm Size.

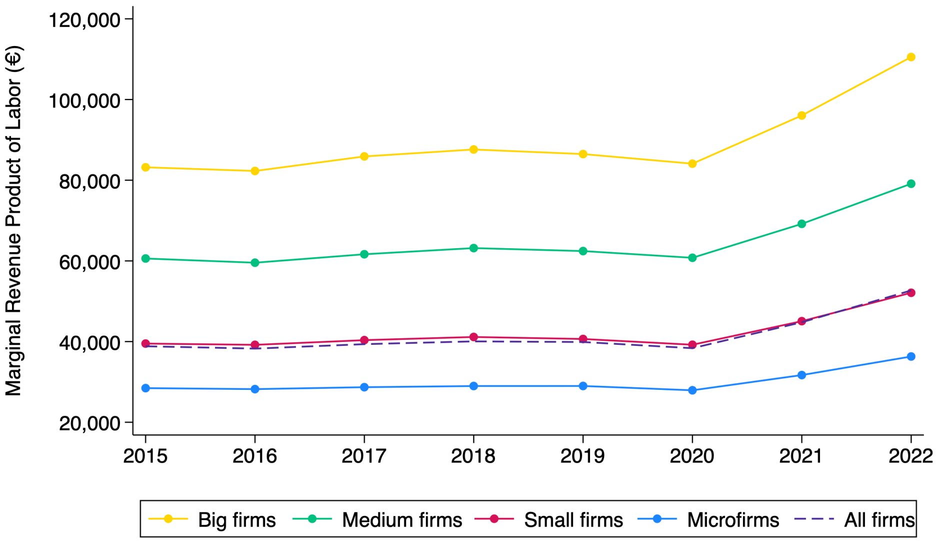
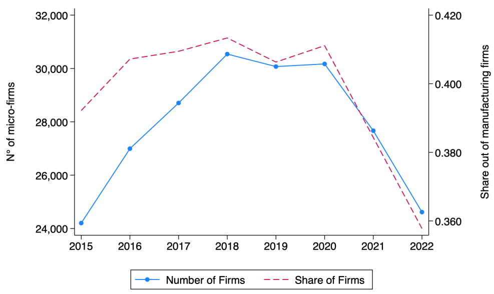
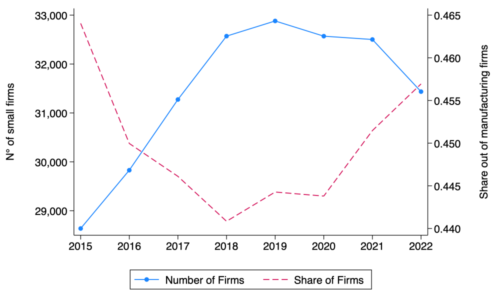


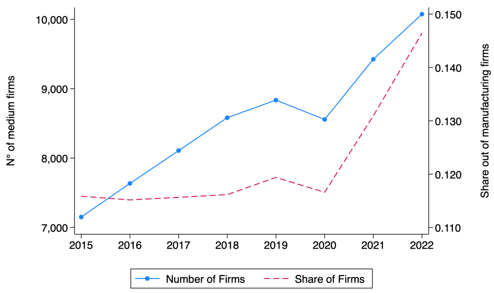
FIGURE 4: Share of Firms by Size Class.



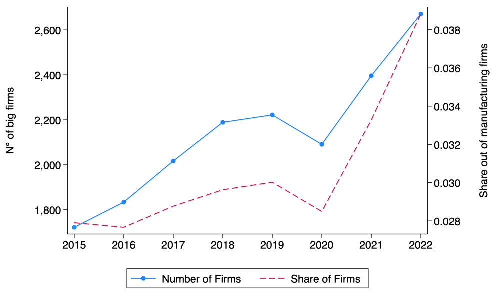
(A) Microfirms.



(B) Small Firms.

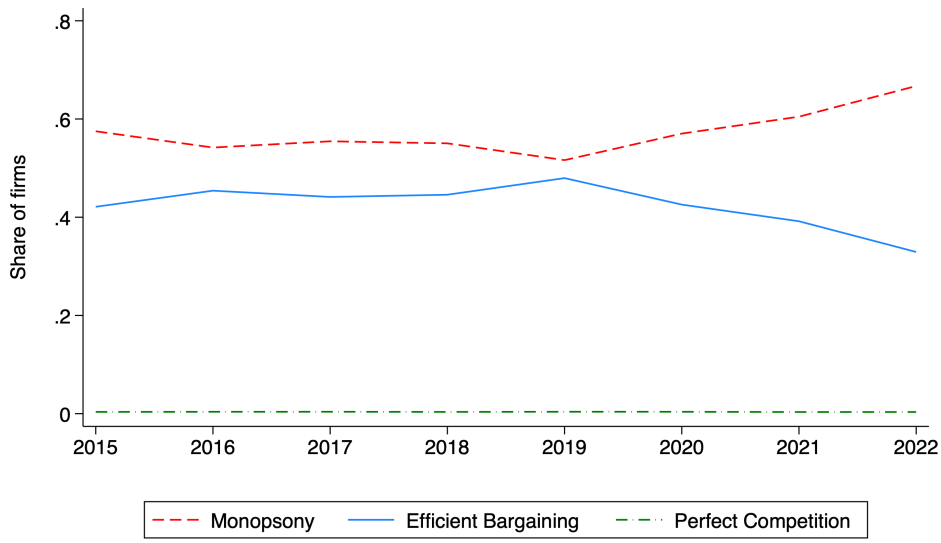


(C) Medium-Sized Firms.



(D) Big Firms.

FIGURE 5: Share of Firms by Labor Market Regimes.



5.2. **Local Effects of the National Minimum Wage**

What would be the impact of the proposed minimum wage (€13.85 per hour) on employment levels? I investigate this issue in Figure 6, where the share of firms in categories 1 and 3 (see section 4.4 for the classification) is reported for each province, with darker colors indicating higher values of the share of firms in a given category.[19]

Negative employment effects (corresponding to firms in category 1) are mainly concentrated in the Center and in the South of Italy, with a partial exception being the North-Western region of Liguria. Positive effects on employment (corresponding to firms in category 3) would also come from the Center–South of the country, which would experience a stark reduction both in monopsony power and (probably) in the number of active firms, if the proposed minimum wage were to be implemented. Moreover, its bite in Northern Italy would probably be very limited, with the majority of active firms in the area belonging to category 4.[20]

Considering Italy as a whole, this national minimum wage policy would result in around 15% of firms reducing their labor demand and more than 23% of firms increasing wages to their existing workforce and/or employing more workers.

The picture emerging from this analysis suggests that a more fine-tuned approach to the optimal minimum wage definition would both bring gains in terms of efficiency and improve the functioning of local labor markets.[21]

5.3. **Optimal Regional Minimum Wages**

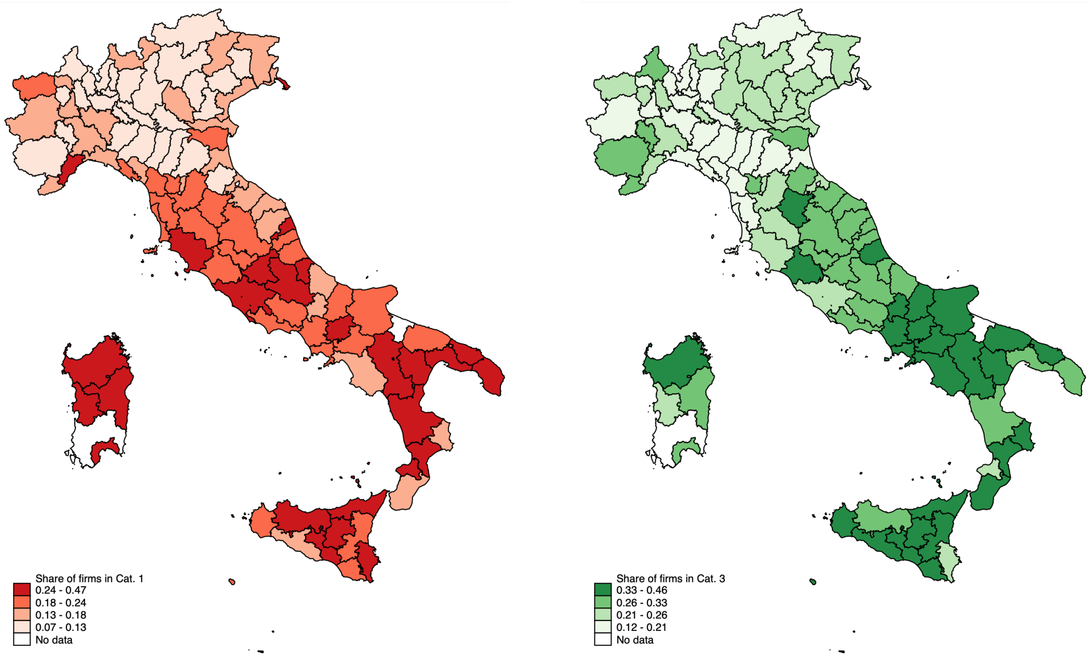
In this light, it seems reasonable to propose a differentiated approach where a different optimal minimum wage is defined at the level of every single Italian region. This endeavor, although very

[19] All choropleth maps in this paper are realized in Stata/SE 18.0 with the modules *shp2dta, mif2dta,* and *spmap* (Crow, 2006; Pisati, 2005, 2007).

[20] See Figure 11 in Appendix C for the shares of firms in categories 2 and 4.

[21] Note that the proposed national minimum wage would imply sizeable disemployment effects in certain southern provinces, with more than 25% of firms being negatively affected by its introduction. In these areas, a rather big number of firms could potentially be pushed out of the market.

Figure 6: Share of Firms by Minimum Wage Category (National MW).



(a) Category 1. (b) Category 3.

much theoretical in its nature, can possibly shade light on regional patterns in terms of wages and productivity.

As it was the case when studying the optimal national minimum wage, the regional analysis is again restricted to firms operating in the manufacturing sector.[22]

Table 3 displays the results of the optimization procedure,[23] with columns 2 and 3 showing the ranges found for yearly minimum wages, and columns 4 and 5 indicating the hourly minimum wage ranges for every region.[24] Considering the results originating from the procedure employing unweighted firm shares (i.e., the lower bounds of every range), there is wide between-regions heterogeneity in the optimal minimum wage values. The three highest minima (which are found in the autonomous province of Bolzano, Veneto, and Lombardy) are more than double the amount of the proposed Sardinian minimum wage, which, at €7.46 per hour, is the lowest value found in Italy.

These results are most likely driven by local productivity, whose geographical distribution in 2022 is plotted for reference in Figure 7. Average values of the marginal revenue product of labor (in manufacturing) range from a low of €36,627 in the province of Nuoro, in Sardinia, to a high of €64,420 for the province of Lecco, in Lombardy. Unsurprisingly, the North-South divide in average firm productivity is very pronounced, with a strong positive association between latitude and average marginal product of labor.[25]

[22]As before, further analyses are needed in order to explain the sharp wedge between the optimal regional MWs computed with data from all sectors and the median wage in Italy in 2022.

[23]As the optimization problem is solved through an iterative process, the error margin is €500 per year, or €0.178 per hour.

[24]As it was the case for the national result, a value of $SSC = 0.284$ was used in equation (13) to convert yearly values into hourly ones.

[25]The observation that productivity is higher in the North than in the South (with medium values in Central Italy) is a qualitative one, which can be drawn simply by looking at Figure 7.

Table 3: Optimal Regional Minimum Wages in 2022.

Regions	Yearly MW (€)		Hourly MW (€)	
	Unweighted	Weighted	Unweighted	Weighted
Abruzzo	38,500	48,000	13.67	17.05
Basilicata	34,500	55,500	12.25	19.71
Prov. Aut. Bolzano	52,750	65,500	18.73	23.26
Calabria	31,000	45,000	11.01	15.98
Campania	32,000	50,000	11.37	17.76
Emilia-Romagna	45,500	66,500	16.16	23.62
Friuli-Venezia Giulia	46,000	60,500	16.34	21.49
Lazio	27,000	57,000	9.59	20.24
Liguria	32,500	55,000	11.54	19.53
Lombardia	47,000	64,000	16.69	22.73
Marche	40,000	50,500	14.21	17.94
Molise	36,250	42,250	12.87	15.01
Piemonte	39,000	71,000	13.85	25.22
Puglia	36,500	47,500	12.96	16.87
Sardegna	21,000	44,000	7.46	15.63
Sicilia	32,000	40,000	11.37	14.21
Toscana	32,500	56,500	11.54	20.07
Prov. Aut. Trento	45,500	69,000	16.16	24.51
Umbria	38,500	53,000	13.67	18.82
Valle D'Aosta	45,500	60,750	16.16	21.58
Veneto	48,500	58,500	17.23	20.78

Notes. "Unweighted" indicates that simple shares of firms were used in the computation. "Weighted" indicates that the computation was carried out using employment shares.

While some minimum wage ranges are somewhat narrow, allowing for a precise definition of its optimal level (this is the case, e.g., in Molise, Sicilia, Abruzzo, and Veneto), others are rather wide, giving little information on the actual minimum wage value that would benefit the region in question (this happens, e.g., in Piemonte, Lazio, Toscana, and Trento). I attribute this to different distributions of firm sizes between regions: if all firms had the same number of employees, there would be no difference between the unweighted and the weighted result.[26] The share of employees working in big firms is positively correlated (0.44) with the gap between the weighted and unweighted optimal minimum wage at the region level.[27]

Firms with a number of employees greater than the average increase the weighted share of the category to which they belong and cause the two results (weighted and unweighted MW) to differ, with the direction of this effect depending on the minimum wage categories in which the bigger firms operate. Since bigger firms are more productive (see Figure 3) and pay higher wages,[28] I expect them to belong mainly to categories 3 and 4 for most given values of MW. Considering that (i) cat. 2 and cat. 4 are not directly part of the analysis; (ii) for any given value of MW, bigger firms are over-represented in cat. 3 and cat. 4 and under-represented in cat. 1 and cat. 2,[29] the

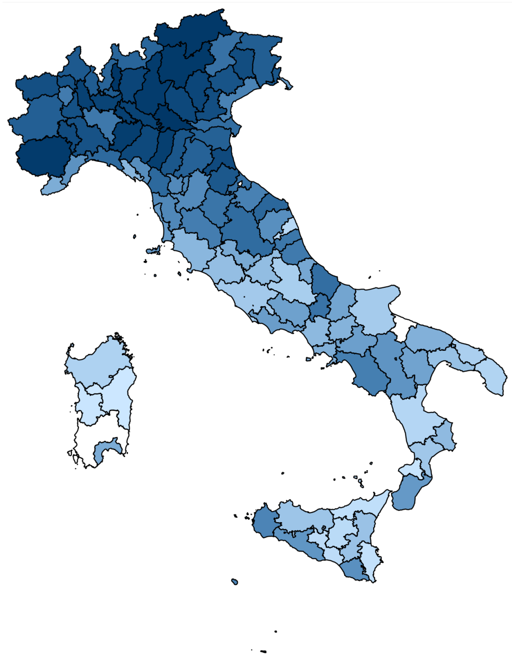
[26]See Tables 4 and 5 in Appendix C for the distribution of firms and employees in class-sizes across regions.

[27]See Figure 14 in Appendix C for the scatterplot visualizing this correlation.

[28]As shown in Table 6, median manufacturing wages in 2022 were €31,734 in microfirms, €39,856 in small firms, €48,968 in medium-sized firms, and €57,001 in big firms.

[29]This amounts to saying that there are always fewer "big" firms in cat. 1 than in cat. 3, where a "big" firm is defined as one having more than X employees, with X greater than the region-specific average employment level. (This also holds between cat. 2 and cat. 3, but note that cat. 2 is not directly included in the analysis.) To see why this is the case, please refer to Section 4.4 and consider the condition $MRP^L_{it} < MW$, required to be classified in

Figure 7: Labor Productivity by Province in 2022.



above causes regions having a higher share of employees working in big firms to display a wider gap between the share of firms in cat. 1 and cat. 3 *vis à vis* other regions. I claim that this differential wedge is concave in MW, the optimal minimum wage level on which we optimize.[30] Referring to Figure 1, this amounts to saying that the location of the maximum distance between the cat. 3 and cat. 1 curves will be further to the right on the *x-axis* for those regions.

In turn, this causes regions featuring a lot of "big" firms to display a higher level of optimal minimum wage when weighting observations using employment shares.[31]

For the reasons above, I always consider the lower bound of every interval (i.e., the unweighted MW) in the analysis that follows to draw a more realistic scenario when describing the potential consequences of the introduction of regional minimum wages.

5.4. **Local Effects of Regional Minimum Wages**

If the proposed regional minimum wages were in place, the deep North–South divide in employment effects shown in Figure 6 would give way to much more homogeneity in effects when comparing regions.

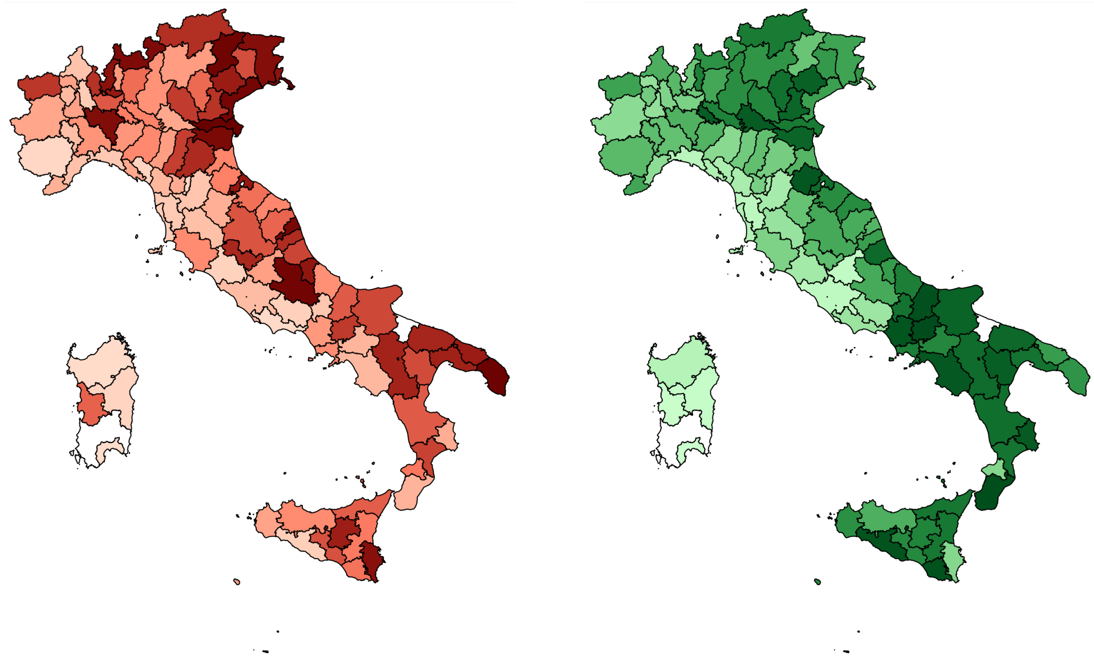
Indeed, Panel (a) of Figure 8 shows a rather homogeneous North–South distribution of provinces with the highest shares of firms in cat. 1. Panel (b) of the same figure shows how shares of firms

cat. 1. For bigger firms, having higher productivity MRP^L_{it}, it is more difficult to satisfy this condition and easier to satisfy $MRP^L_{it} \geq MW$, which is required to be classified in cat. 3, and this holds for any value of MW. Moreover, the condition $\phi_{it} = \frac{w_{it}}{MRP^L_{it}} \geq 1$ follows a similar path, with greater values of MRP^L_{it} making it more likely for a firm to be classified in cat. 3 (if $\phi_{it} < 1$, i.e., "big" firms enjoy monopsonistic power) or cat. 4 (if $w_{it} \geq MW$, i.e., "big" firms' wages are higher than the given MW).

[30] This claim should be verified by further research, ideally with a formal proof. It is nevertheless sensible to assume a relevant degree of monopsonistic power in the market, which would make it easier for "big" firms to fall into cat. 3 rather than into cat. 1 (please refer to Section 4.4 for the relevant conditions).

[31] Also here, a "big" firm is defined as one having more than X employees, with X greater than the region-specific average employment level.

Figure 8: Share of Firms by Minimum Wage Category (Regional MWs).



(a) Category 1. (b) Category 3.

benefiting from the introduction of regional minimum wages would be rather high across the whole country, with the highest value recorded in Benevento, Campania (46.73% of firms) and the lowest one in Cagliari, Sardinia (5.92% of firms).[32]

Within-region heterogeneity dominates the picture, with a weakly positive correlation (0.16) between productivity and the share of firms on which the introduction of a regional minimum wage would be positive (i.e., the ones classified in cat. 3) at the province level.[33]

In the whole of Italy, the regional minimum wage policy would induce positive employment effects on 27.16% of firms, while having negative labor demand effects on 17.75% of them. This is a net improvement with respect to the national MW case (where 23% of firms are in cat. 3 and 15% of them are in cat. 1), with the increase in the share of negatively impacted firms being more than compensated for by a larger increase in the share of positively affected ones.

Moreover, the marginal revenue product of labor is found to be weakly correlated (0.19) with net gains from the regional MWs (i.e., with the difference between shares of firms in cat. 3 and in cat. 1) at the province level, suggesting that optimal regional minimum wages likely depend on local productivity.[34]

6. Conclusion

Given the recent discussion about the proposed introduction of a statutory minimum wage in Italy, in this paper I attempted to assess its potential impact through an empirical study employing the latest available data on Italian firms. A measure of individual labor market power was computed from the estimates of input-output elasticities, which were retrieved from a translog production

[32]See Figure 12 in Appendix C for the shares of firms in categories 2 and 4.
[33]See Figure 13 in Appendix C for the scatterplot visualizing this correlation.
[34]See Figure 15 in Appendix C for the scatterplot visualizing this correlation.

function via a WLP estimator. This approach allowed to correct for the simultaneity between input choices and unobserved productivity shocks. Following the previous literature, I decided to restrict the analysis to the manufacturing sector, which is characterized by higher average monopsony power (with respect to the value found when considering all sectors) and is representative of the Italian labor market, comprising the relative majority of active firms in the country (more than 31%).

The optimal national minimum wage for 2022 was found to be €39,000 per year, or €13.85 per hour on average. This signals a steep increase with respect to both the optimal MW levels prevailing in Italy until 2020 and to the one found by Caselli et al. (2023) for 2018. I documented noteworthy gains in labor productivity from 2020 onward and a sizeable rise in monopsony power during the 2019-2022 period, and I attributed the increase in optimal minimum wage to both of these factors.

However, this national MW value was found to only be optimal for the country as a whole, featuring highly heterogeneous effects at the local level. Policymakers would thus be faced with both distributional and efficiency issues. In the short run, a national minimum wage in Italy would not only give way to significantly different employment effects between different geographical areas; it would also push a relevant number of firms out of the market.[35]

Aiming to obtain more homogeneous between-regions employment effects from the potential introduction of a minimum wage, I proposed a regional approach, where optimal MW levels were defined for each of the 21 Italian regions. Results from the spatial analysis based on these optimal values revealed how regional minimum wages would increase positive employment effects at the national level. The proposed regional minima would also have an efficiency-enhancing effect, as the share of firms potentially expanding the workforce would be highest in the most productive provinces. Therefore, regional minimum wages would have a higher bite than the national one, would increase aggregate positive labor supply effects and would reduce negative effects on the labor demand at the local level.

The main limitation of this study lies in the lack of detailed balance-sheet data for many of the smallest Italian firms, which prevents a fully comprehensive analysis of the demand-induced effects of a minimum wage. Moreover, using balance-sheets as a primary data source only allows to observe the mean cost of labor, not the median one. Data on reservation wages would also be useful to identify the pull-effect of the MW at the participation margin. A quantitative estimation of the firm-level employment effect (in terms of the number of employees involved) would allow to capture heterogeneity between firms of different sizes in different locations, leading to a more precise characterization of aggregate employment effects at the national level. Lastly, more recent data from 2023 to 2024 could provide a better understanding of the optimal MW trend in relation to the median wage.

Further research should investigate the general equilibrium effects associated with the introduction of national and regional minimum wages in Italy, with a focus on both workers' within-country migration decisions and firms' relocation choices.

REFERENCES

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Ahlfeldt, G. M., D. Roth, and T. Seidel (2018). The regional effects of Germany's national minimum wage. *Economics Letters 172*, 127–130.

Ahlfeldt, G. M., D. Roth, and T. Seidel (2022). Optimal minimum wages. CEP Discussion Papers dp1823, Centre for Economic Performance, LSE.

[35] Please note that firms' and workers' relocation incentives (or other spillover effects) are not taken into account in this paper.

Arribas Cámara, J., L. Cárdenas, and A. Rial (2024). The effects of the minimum wage on inequality : A literature review. *Work Organisation, Labour & Globalisation 18*(2), 196–218.

Baum, C. F., M. E. Schaffer, and S. Stillman (2002). IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation. Statistical Software Components, Boston College Department of Economics.

Camera dei Deputati (2023). Proposta di legge 1275 - disposizioni per l'istituzione del salario minimo.

Card, D. and A. B. Krueger (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review 84*(4), 772–793.

Caselli, M., J. Mondolo, and S. Schiavo (2023). Labour market power and the quest for an optimal minimum wage: evidence from Italy. *Applied Economics 55*(15), 1713–1727.

Caselli, M., L. Nesta, and S. Schiavo (2021). Imports and labour market imperfections: Firm-level evidence from France. *European Economic Review 131*(C), 103632.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics 134*(3), 1405–1454.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau (1973). Transcendental logarithmic production frontiers. *The Review of Economics and Statistics 55*(1), 28–45.

Crow, K. (2006). SHP2DTA: Stata module to converts shape boundary files to Stata datasets. Statistical Software Components, Boston College Department of Economics.

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review 102*(6), 2437–2471.

Dobbelaere, S. and J. Mairesse (2013). Panel data estimates of the production function and product and labor market imperfections. *Journal of Applied Econometrics 28*, 1–46.

Dube, A. (2019). Impacts of minimum wages: review of the international evidence. Technical report, UK Government.

Dube, A. and A. Lindner (2024). Minimum Wages in the 21st Century. Technical Report 2425, Rockwool Foundation Berlin (RF Berlin) - Centre for Research and Analysis of Migration (CReAM).

Dube, A. and B. Zipperer (2024). Own-Wage Elasticity: Quantifying the Impact of Minimum Wages on Employment. NBER Working Papers 32925, National Bureau of Economic Research, Inc.

Eurofound (2024). National minimum wages 2024. https://www.eurofound.europa.eu/en/data-catalogue/minimum-wages/national-minimum-wages-2024.

European Commission (2003). Commission recommendation of 6 may 2003 concerning the definition of micro, small and medium-sized enterprises.

Harasztosi, P. and A. Lindner (2019). Who pays for the minimum wage? *American Economic Review 109*(8), 2693–2727.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70*(2), 317–341.

Manning, A. (2021). The elusive employment effect of the minimum wage. *Journal of Economic Perspectives 35*(1), 3–26.

Mertens, M. (2019). Micro-mechanisms behind declining labour shares: Market power, production processes, and global competition. *IWH-CompNet Discussion Papers 3*, 104802.

Mertens, M. (2020). Labor market power and the distorting effects of international trade. *International Journal of Industrial Organization 68*, 102562.

Neumark, D. and W. Wascher (2006). Minimum Wages and Employment: A Review of Evidence from the New Minimum Wage Research. Working Papers 060708, University of California-Irvine, Department of Economics.

Petrin, A. and J. Levinsohn (2012). Measuring aggregate productivity growth using plant-level data. *RAND Journal of Economics 43*(4), 705–725.

Pisati, M. (2005). MIF2DTA: Stata module convert MapInfo Interchange Format boundary files to Stata boundary files. Statistical Software Components, Boston College Department of Economics.

Pisati, M. (2007). SPMAP: Stata module to visualize spatial data. Statistical Software Components, Boston College Department of Economics.

Simon, A. and M. Wilson (2021). Optimal minimum wage setting in a federal system. *Journal of Urban Economics 123*, 103336.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters 104*(3), 112–114.

A. MATHEMATICAL APPENDIX

A.1. **Production Function Estimation**

In what follows, I give a more detailed description of the production function presented in section 4.2.

Capital is denoted by k, materials are denoted by m and labor is denoted by l. The indexes $_{L1}$ and $_{L2}$ indicate, respectively, the first and second lags of a variable.

The variables assumed to be exogenous are capital k, its square k^2, its first lag k_{L1}, and the first lag of materials m_{L1}.

The endogenous variables are labor, material, and their interactions: l, m, l^2, m^2, lk, lm, mk. Accordingly, they are instrumented with the following lagged variables: l_{L1}, l_{L2}, k_{L2}, m_{L2}, l^2_{L1}, l^2_{L2}, m^2_{L2}, lk_{L1}, lk_{L2}, lm_{L1}, lm_{L2}, and mk_{L2}.

Year fixed-effects are added to control for sector-wide, time-specific shocks that might affect all firms in a given industry.

A.2. **Confidence Interval for ψ**

I hereby detail the procedure to derive the variance of ψ, used in the computation of the confidence interval presented in section 4.3.

Starting from equation (9) and following Caselli et al. (2021), note that α_M and α_L are observed, but θ_M and θ_L are estimated. As such, they must be considered as random variables. (Note that, in what follows, indexes i and t are left aside for simplicity.)

We can write:

$$\begin{aligned}\mathrm{Var}(\psi) &= \mathrm{Var}\left(\frac{\theta_M}{\alpha_M} - \frac{\theta_L}{\alpha_L}\right) = \\ &= \frac{\mathrm{Var}(\hat{\theta}_M)}{\alpha_M^2} + \frac{\mathrm{Var}(\hat{\theta}_L)}{\alpha_L^2} - \frac{2\ \mathrm{Cov}(\hat{\theta}_M, \hat{\theta}_L)}{\alpha_M \cdot \alpha_L}\end{aligned} \tag{14}$$

Given equations (6) and (7), we can derive the variances of $\hat{\theta}_M$ and $\hat{\theta}_L$ as follows:

$$\begin{aligned}\mathrm{Var}(\hat{\theta}_M) &= \mathrm{Var}\left(\hat{\beta}_M + 2\hat{\beta}_{M^2} \cdot m + \hat{\beta}_{KM} \cdot k + \hat{\beta}_{LM} \cdot l\right) = \\ &= \mathrm{Var}(\hat{\beta}_M) + 4m^2 \cdot \mathrm{Var}(\hat{\beta}_{M^2}) + k^2 \cdot \mathrm{Var}(\hat{\beta}_{KM}) + l^2 \cdot \mathrm{Var}(\hat{\beta}_{LM}) + \\ &\quad + 4m \cdot \mathrm{Cov}(\hat{\beta}_M, \hat{\beta}_{M^2}) + k \cdot \mathrm{Cov}(\hat{\beta}_M, \hat{\beta}_{KM}) + l \cdot \mathrm{Cov}(\hat{\beta}_M, \hat{\beta}_{LM}) + \\ &\quad + 4km \cdot \mathrm{Cov}(\hat{\beta}_{M^2}, \hat{\beta}_{KM}) + 4lm \cdot \mathrm{Cov}(\hat{\beta}_{M^2}, \hat{\beta}_{LM}) + 2lk \cdot \mathrm{Cov}(\hat{\beta}_{KM}, \hat{\beta}_{LM})\end{aligned} \tag{15}$$

$$\begin{aligned}\mathrm{Var}(\hat{\theta}_L) &= \mathrm{Var}\left(\hat{\beta}_L + 2\hat{\beta}_{L^2} \cdot l + \hat{\beta}_{KL} \cdot k + \hat{\beta}_{LM} \cdot m\right) = \\ &= \mathrm{Var}(\hat{\beta}_L) + 4l^2 \cdot \mathrm{Var}(\hat{\beta}_{L^2}) + k^2 \cdot \mathrm{Var}(\hat{\beta}_{KL}) + l^2 \cdot \mathrm{Var}(\hat{\beta}_{LM}) + \\ &\quad + 4l \cdot \mathrm{Cov}(\hat{\beta}_L, \hat{\beta}_{L^2}) + k \cdot \mathrm{Cov}(\hat{\beta}_L, \hat{\beta}_{KL}) + m \cdot \mathrm{Cov}(\hat{\beta}_L, \hat{\beta}_{LM}) + \\ &\quad + 4kl \cdot \mathrm{Cov}(\hat{\beta}_{L^2}, \hat{\beta}_{KL}) + 4lm \cdot \mathrm{Cov}(\hat{\beta}_{L^2}, \hat{\beta}_{LM}) + 2mk \cdot \mathrm{Cov}(\hat{\beta}_{KL}, \hat{\beta}_{LM})\end{aligned} \tag{16}$$

Then, the covariance between $\hat{\theta}_M$ and $\hat{\theta}_L$ is computed as follows:

$$\begin{aligned}\mathrm{Cov}(\hat{\theta}_M, \hat{\theta}_L) &= \mathrm{Cov}(\hat{\beta}_M, \hat{\beta}_L) + 2l \cdot \mathrm{Cov}(\hat{\beta}_M, \hat{\beta}_{L^2}) + k \cdot \mathrm{Cov}(\hat{\beta}_M, \hat{\beta}_{KL}) + \\ &\quad + m \cdot \mathrm{Cov}(\hat{\beta}_M, \hat{\beta}_{LM}) + 2m \cdot \mathrm{Cov}(\hat{\beta}_{M^2}, \hat{\beta}_L) + 4lm \cdot \mathrm{Cov}(\hat{\beta}_{M^2}, \hat{\beta}_{L^2}) + \\ &\quad + 2km \cdot \mathrm{Cov}(\hat{\beta}_{M^2}, \hat{\beta}_{KL}) + 2m^2 \cdot \mathrm{Cov}(\hat{\beta}_{M^2}, \hat{\beta}_{LM}) + k \cdot \mathrm{Cov}(\hat{\beta}_{KM}, \hat{\beta}_L) + \\ &\quad + 2kl \cdot \mathrm{Cov}(\hat{\beta}_{KM}, \hat{\beta}_{L^2}) + k^2 \cdot \mathrm{Cov}(\hat{\beta}_{KM}, \hat{\beta}_{KL}) + km \cdot \mathrm{Cov}(\hat{\beta}_{KM}, \hat{\beta}_{LM}) + \\ &\quad + l \cdot \mathrm{Cov}(\hat{\beta}_{LM}, \hat{\beta}_L) + 2l^2 \cdot \mathrm{Cov}(\hat{\beta}_{LM}, \hat{\beta}_{L^2}) + \\ &\quad + kl \cdot \mathrm{Cov}(\hat{\beta}_{LM}, \hat{\beta}_{KL}) + lm \cdot \mathrm{Cov}(\hat{\beta}_{LM}, \hat{\beta}_{LM})\end{aligned} \tag{17}$$

Solving equations (15), (16), and (17) first and then plugging the results into equation (14) allows to find an estimate for the variance of ψ. Finally, it is possible to use such result to compute the confidence interval of ψ presented in equation (10).

B. DATA APPENDIX

B.1. Price Indexes Data

To express all variables in real terms, adequate price indexes are used.[36] All publicly available sector-specific Producer Price Indexes (PPIs) are retrieved from the ISTAT website. These include PPIs for sectors A (Agriculture, Forestry and Fishing), B (Mining and Quarrying), C (Manufacturing), D (Electricity, Gas, Steam and Air Conditioning Supply), E (Water Supply; Sewerage, Waste

[36] The exception being labor, for which the number of employees (already available in the data) is used.

Management and Remediation Activities)[37], F (Construction)[38], H (Transportation and Storage), J (Information and Communication), M (Professional, Scientific and Technical Activities), and N (Administrative and Support Service Activities). Sub-sector-level PPIs are available in the service industry and are thus used individually. These include sub-sectors from 49 to 53 (H), from 58 to 63 (J), from 69 to 75 (M), and from 77 to 82 (N).[39]

All indexes have a standard value of 100 in the base year 2015. Although these sectors, for which price data are available, comprise the majority of Italian firms and employed workers, data on the remaining sectors are not publicly available. I thus constructed a national average PPI, based on available price data and using sectoral revenues as weights. Finally, I used this national weighted-average PPI to proxy for the PPI of sectors G (Wholesale and Retail Trade), I (Accomodation and Food Service Activities), K (Financial and Insurance Activities), L (Real Estate Activities), O (Public Administration and Defense; Compulsory Social Security), P (Education), Q (Human Health and Social Work Activities), R (Arts, Entertainment and Recreation), S (Other Service Activities), T (Activities of Households as Employers or for Own Use), and U (Activities of Extraterritorial Organizations and Bodies).

This amounts to assuming that the (weighted) average of PPI in sectors pertaining to agriculture, industry, and services gives a good approximation of producer price dynamics in the field of trade and in all other sectors.

B.2. **Revenues Data**

The sector-level revenues to be used as weights in the PPI average computation are proxied by the value of production in a given sector and year.[40] This amounts to assuming that the difference between inventories at the beginning and at the end of a given year is, on average, negligible. Production Value data are retrieved from the Istat website for sectors B, C, D, E, F, H, J, M, and N. Output data from Istat are used for sector A.

B.3. **Services Sectors – Approximation Methodology**

As PPI data on the services (H, J, M, and N) sectors are often incomplete, a weighted (based on production value) average PPI of the available sub-sectors is constructed and then used as a proxy for the PPI of sub-sectors for which data are missing.

Specifically, the weighted averages are constructed as follows: in sector H, the (weighted) average between the PPIs of subdivisions 52.1 (Warehousing and storage) and 52.2 (Support activities for transportation) is used as a proxy for the PPI of sub-sector 52 (Warehousing and support activities for transportation); in sector J, the PPIs of sub-sectors 61 (Telecommunications), 62 (Computer programming, consultancy and related activities), and 63 (Information service activities) are averaged to obtain a proxy PPI for sub-sectors 58 (Publishing activities), 59 (Motion picture, video and television program production, sound recording and music publishing activities), and 60 (Programming and broadcasting activities); in sector M, the average between PPIs of sub-sectors 69 (Legal and accounting activities), 70 (Activities of head offices; management consultancy activities), 71 (Architectural and engineering activities; technical testing and analysis),

[37]Due to data availability and based on production value considerations, the PPI of sector E is proxied by the one of sub-sector 36 (Water collection, treatment and supply).

[38]Due to data availability and based on production value considerations, the PPI of sector F is proxied by the one of sub-sector 41.2 (Construction of residential and non-residential buildings).

[39]Due to both data availability and production value considerations, the following PPI approximations are chosen: division 49 (Land transport and transport via pipelines) is proxied by subdivision 49.4 (Freight transport by road and removal services); subdivision 52.2 (Support activities for transportation) is proxied by sub-subdivision 52.24 (Cargo handling); division 63 (Information service activities) is proxied by subdivision 63.1 (Data processing, hosting and related activities; web portals); division 70 (Activities of head offices; management consultancy activities) is proxied by subdivision 70.2 (Management consultancy activities); division 81 (Services to buildings and landscape activities) is proxied by subdivision 81.2 (Cleaning Activities).

[40]For sector A (Agriculture, Forestry and Fishing) only, revenues are proxied by output.

and 73 (Advertising and market research) proxies for the PPI of sub-sectors 72 (Scientific research and development), 74 (Other professional, scientific and technical activities), and 75 (Veterinary activities); finally, in sector N, the weighted average PPI of sub-sectors 78 (Employment activities), 80 (Security and investigation activities), and 81 (Services to buildings and landscape activities) is used as a proxy PPI for sub-sectors 77 (Rental and leasing activities), 79 (Travel agency, tour operator), and 82 (Office administrative, office support).

C. Additional figures and tables

Figure 9: Labor Productivity by Macro-Regions.

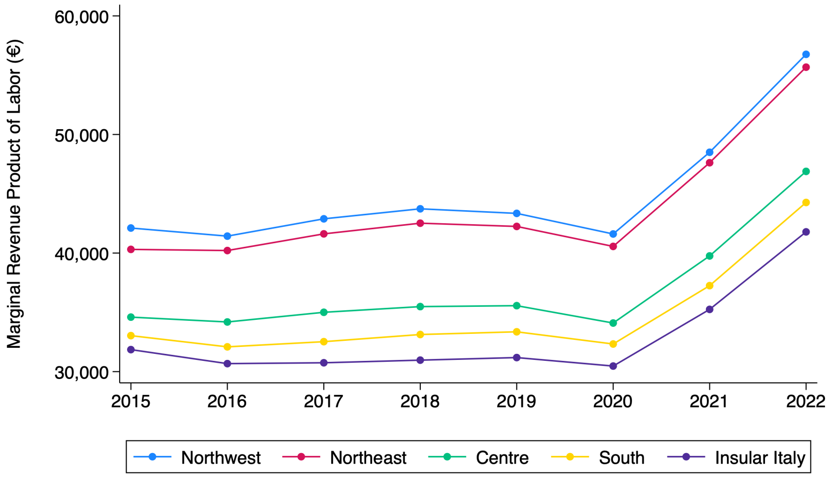


Figure 10: Single-Employee Firms.

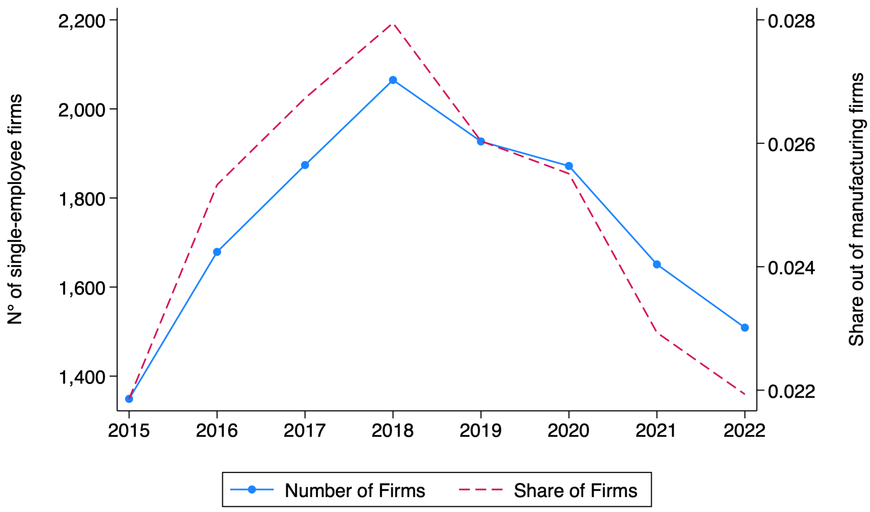
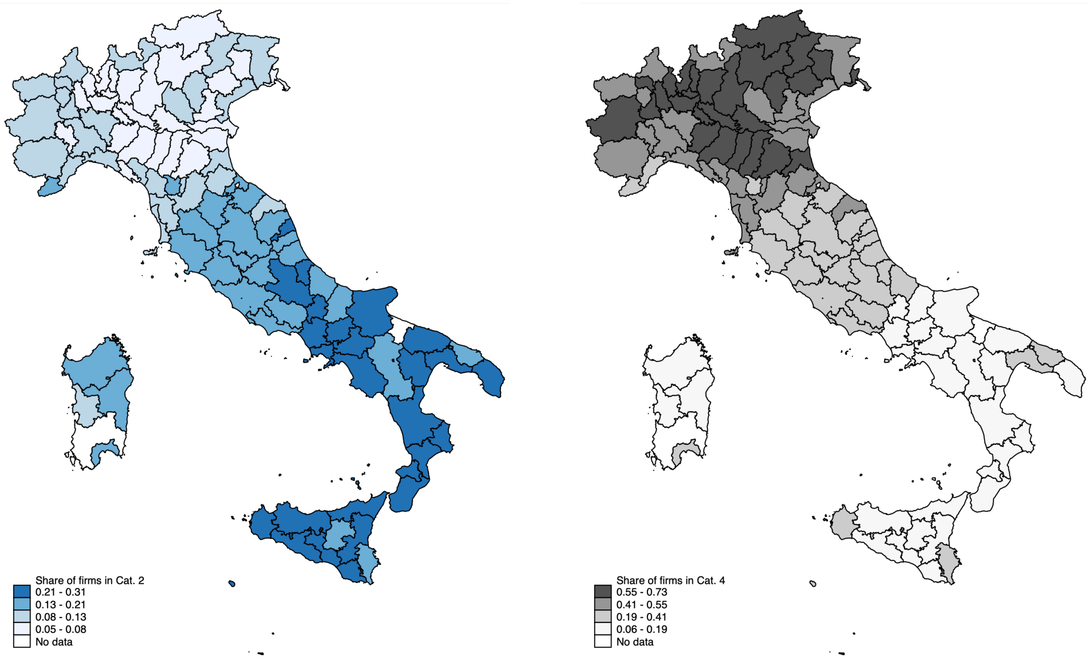


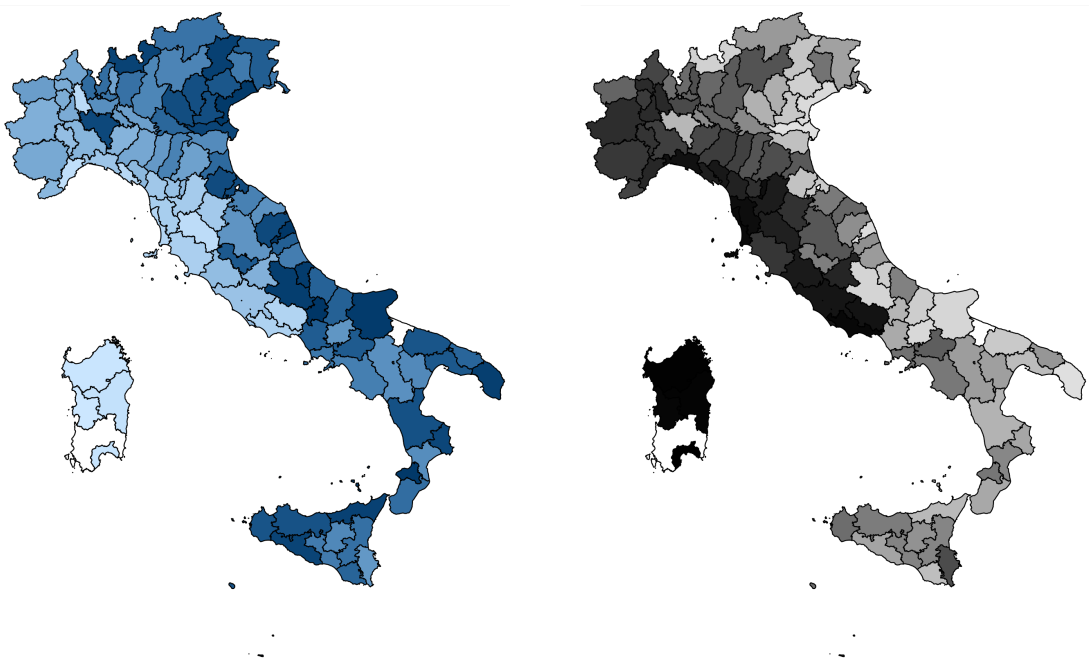
FIGURE 11: Share of Firms by National Minimum Wage Category (2 and 4).



(A) Category 2.

(B) Category 4.

FIGURE 12: Share of Firms by Regional Minimum Wage Category (2 and 4).



(A) Category 2.

(B) Category 4.

FIGURE 13: Correlation between Productivity and Share of Firms in Cat. 3.

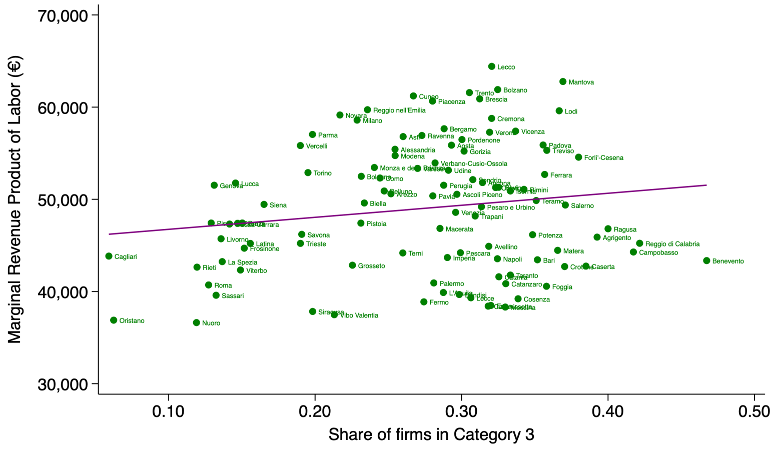


FIGURE 14: Correlation between Share of Employees in Big Firms and MW Gap.

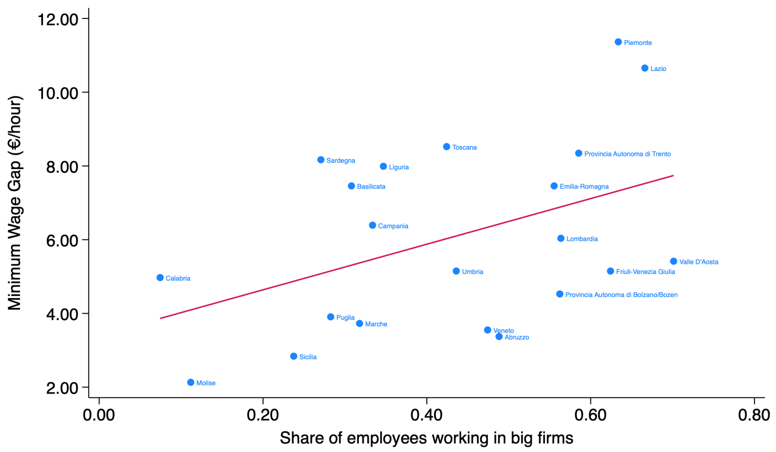


FIGURE 15: Correlation between Productivity and Net Gains from Regional MWs.

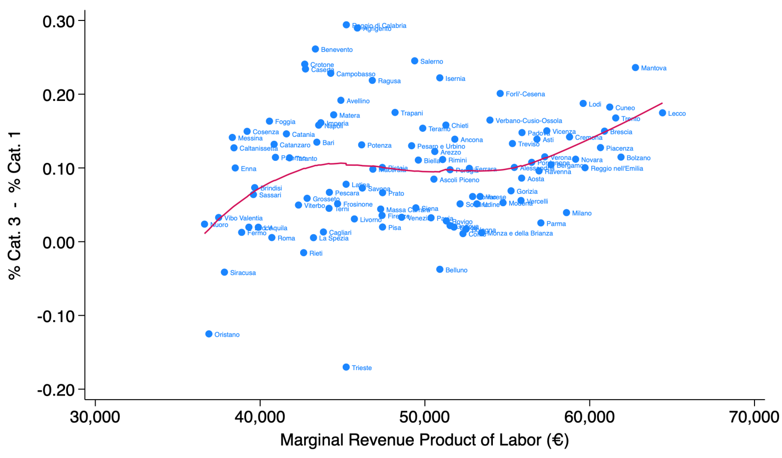


Table 4: Share of Firms by Size Class.

Regions	Microfirms	Small Firms	Medium Firms	Big Firms
Abruzzo	0.38	0.47	0.11	0.04
Basilicata	0.46	0.41	0.11	0.02
Prov. Aut. Bolzano	0.30	0.44	0.19	0.07
Calabria	0.56	0.38	0.05	0.00
Campania	0.46	0.41	0.10	0.02
Emilia-Romagna	0.32	0.47	0.16	0.05
Friuli-Venezia Giulia	0.31	0.48	0.16	0.05
Lazio	0.52	0.37	0.09	0.02
Liguria	0.39	0.45	0.13	0.03
Lombardia	0.33	0.46	0.16	0.05
Marche	0.36	0.48	0.14	0.02
Molise	0.46	0.44	0.08	0.01
Piemonte	0.31	0.47	0.17	0.05
Puglia	0.46	0.43	0.10	0.01
Sardegna	0.54	0.38	0.08	0.01
Sicilia	0.52	0.40	0.07	0.01
Toscana	0.38	0.48	0.12	0.03
Prov. Aut. Trento	0.33	0.45	0.16	0.06
Umbria	0.36	0.46	0.14	0.04
Valle D'Aosta	0.40	0.38	0.16	0.07
Veneto	0.30	0.48	0.18	0.04
Italy	0.36	0.46	0.15	0.04

Notes. The Table displays the distribution of manufacturing firms by size-class in each Italian region in 2022. Microfirms employ fewer than 10 people and have a turnover below €2 million; small firms have fewer than 50 employees and a turnover below €10 million; medium-sized firms employ fewer than 250 people and do not exceed €250 million in yearly turnover. All remaining firms with values above these thresholds are classified as big firms.

Table 5: Share of Employees by Size Class of Firm.

Regions	Microfirms	Small Firms	Medium Firms	Big Firms
Abruzzo	0.05	0.25	0.21	0.49
Basilicata	0.09	0.32	0.28	0.31
Prov. Aut. Bolzano	0.03	0.15	0.26	0.56
Calabria	0.19	0.48	0.26	0.07
Campania	0.09	0.30	0.28	0.33
Emilia-Romagna	0.04	0.19	0.22	0.56
Friuli-Venezia Giulia	0.03	0.15	0.20	0.62
Lazio	0.05	0.14	0.14	0.67
Liguria	0.07	0.28	0.30	0.35
Lombardia	0.04	0.18	0.23	0.56
Marche	0.06	0.30	0.32	0.32
Molise	0.12	0.46	0.31	0.11
Piemonte	0.03	0.15	0.19	0.63
Puglia	0.09	0.34	0.28	0.28
Sardegna	0.14	0.35	0.25	0.27
Sicilia	0.14	0.38	0.25	0.24
Toscana	0.06	0.27	0.24	0.42
Prov. Aut. Trento	0.03	0.17	0.22	0.59
Umbria	0.05	0.24	0.28	0.44
Valle D'Aosta	0.03	0.10	0.17	0.70
Veneto	0.04	0.21	0.28	0.47
Italy	0.04	0.20	0.23	0.52

Notes. The Table displays the distribution of manufacturing employees by firm-size class in each Italian region in 2022. Microfirms employ fewer than 10 people and have a turnover below €2 million; small firms have fewer than 50 employees and a turnover below €10 million; medium-sized firms employ fewer than 250 people and do not exceed €250 million in yearly turnover. All remaining firms with values above these thresholds are classified as big firms.

Table 6: Median Wage by Firm Size.

Regions	Microfirms	Small Firms	Medium Firms	Big Firms
Abruzzo	27,774	34,033	41,758	49,255
Basilicata	26,375	30,641	43,222	48,911
Prov. Aut. Bolzano	39,993	45,900	52,825	61,755
Calabria	22,325	29,085	39,981	40,097
Campania	23,728	29,344	36,538	46,928
Emilia-Romagna	35,199	43,188	52,366	58,967
Friuli-Venezia Giulia	34,027	41,440	48,283	55,768
Lazio	28,200	35,068	46,706	63,317
Liguria	31,050	40,952	50,907	58,810
Lombardia	36,285	43,484	52,385	59,855
Marche	29,324	36,557	43,747	48,910
Molise	25,663	31,339	36,379	71,556
Piemonte	33,808	40,831	49,929	57,433
Puglia	24,684	30,386	37,190	44,084
Sardegna	27,075	32,172	40,026	43,818
Sicilia	22,654	30,260	39,330	44,807
Toscana	31,426	38,117	47,650	57,319
Prov. Aut. Trento	36,384	43,319	50,455	58,418
Umbria	29,766	35,298	43,939	53,452
Valle D'Aosta	32,427	38,197	51,899	58,088
Veneto	33,341	40,862	48,045	53,651
Italy	31,734	39,856	48,968	57,001

Notes. The Table displays the manufacturing median wage (in Euros) by firm-size class in each Italian region in 2022. Microfirms employ fewer than 10 people and have a turnover below €2 million; small firms have fewer than 50 employees and a turnover below €10 million; medium-sized firms employ fewer than 250 people and do not exceed €250 million in yearly turnover. All remaining firms with values above these thresholds are classified as big firms.

CERTAINTY CRISES, AMBIGUITY AVERSION AND SELF-CONFIRMING EQUILIBRIUM

Matteo Pozzi †

DOI: 10.82029/2025014

> "Ignoranti quem portum petat nullus suus ventus est.[1]"
>
> Lucius Annaeus Seneca, *Epistula ad Lucilium LXXI*

Abstract. I explore how doubt affects economic outcomes in strategic settings. Doubt is defined as a revision of subjective beliefs over payoff-relevant statistical models consistent with objective evidence, leading to consideration of more extreme possibilities. This work builds upon research on Self-Confirming Equilibrium (SCE) with ambiguity-averse players developed by Battigalli, Cerreia-Vioglio, Maccheroni, and Marinacci (2015), extending it to games with payoff uncertainty and exploring the consequences of differences in the perception of ambiguity. The smooth ambiguity model of Klibanoff, Marinacci, and Mukerji (2005) is applied, for it allows clear separation of attitudes toward ambiguity from its perception. The key finding is that players with heightened awareness of uncertainty present in their decision environment are more likely to exhibit *status quo* bias, favoring past behavior and leading to objectively worse outcomes. This occurs even when ambiguity aversion is mild. This result is demonstrated through a comparative statics exercise on the equilibrium set, where players' beliefs are adjusted to reflect varying degrees of ambiguity perception. An application to a simple contracting game illustrates how ambiguity aversion and perception can lead to market shutdowns, such as chronic worker shortages or underinvestment, in spite of mechanisms designed to mitigate these issues.

JEL classification: C72, C73

Keywords: Ambiguity Aversion, Uncertainty Perception, Self-Confirming Equilibrium.

1. INTRODUCTION

Doubt is a common occurrence in decision-making. Worldly wisdom suggests doubt is desirable, for it helps to carefully evaluate all available information, and act upon considerate judgment thereof. Second-guessing oneself is thus arguably reasonable within environments characterized by persistent uncertainty to the extent that this may aid the deliberation process. I clarify the game-theoretic consequences of doubt, which can be informally defined as the revision of subjective beliefs over plausible statistical models relevant to a decision problem, to consider a larger set of possibilities consistent with objective evidence. Such a revision is not put into formal relation with the standard rules of Bayesian updating. It serves exercises of model equilibrium comparative statics in the event of believing that more state distributions are possible than was previously thought. The formal model developed in the thesis combines elements from Decision Theory and Game Theory, building on the existing literature of equilibrium analysis with ambiguity averse-players.

†Pozzi: m.pozzi@studbocconi.it; matteo.pozzi@ecb.europa.eu. I thank Prof. Massimo Marinacci for his excellent supervision throughout my thesis work, and Prof. Pierpaolo Battigalli for insightful comments and feedback in view of my defense. I also thank the editor and two anonymous referees for useful comments and feedback.

[1]Translation: "No wind is favorable to the sailor who knows not the harbor he seeks."

The relevance of this topic can be understood from a historical perspective. The enthusiastic intellectual attitudes and ingenuity that spread at the time of the Industrial Revolution fostered the drive for innovation that led to marvelous inventions. In turn, this brought to unforeseen increases in the marginal efficiency of capital. Almelhem et al. (2023) recently provided robust evidence in support of this explanation for modern economic growth, attributable to Mokyr (2017). Burn-Murdoch (2024) instead presented evidence that Western cultures are currently abandoning their culture of progress, turning "toward one of caution, worry and risk-aversion", which echoes the notion of doubt described above. In parallel with Mokyr's argument, he observes that this shift has been accompanied by increasingly sluggish growth, suggesting a connection between these phenomena. In what follows, I offer a solid framework to rationalize this hypothesis. Before, however, we must make a series of nontrivial modeling decisions on which tools to use to represent these behavioral traits. The proposed notion of doubt does not easily lend itself to a proper formalization.

A relatively straight-forward approach is to adapt the framework of Battigalli et al. (2015) to study equilibrium sets constrained by alternative sets of conjectures, carefully constructed to reflect varying degrees of ambiguity perception. Note that beliefs cannot truly abstract from the decision environment, as the probabilistic weight assigned to alternative hypotheses necessarily depends on what the latter are. It may then seem unappealing to consider preferences whose subjective probabilistic component is constrained *a priori*. However, the analysis reveals that the extent of ambiguity perception indeed has normative implications. Since this is the approach we take, we will need to combine contemporary findings from Game Theory, particularly Self-Confirming Equilibrium analysis, with a solid Decision-theoretic model for representing uncertainty averse preferences.

Following the abstract axiomatization of Expected Utility by von Neumann and Morgenstern (1944), a staple within the theory of Choice Under Risk, Decision Theory has long been concerned with modeling Choice Under Uncertainty, where the latter is interpreted in the sense of Knight (1921). The ground-breaking leap of Savage's (1954) Subjective Expected Utility (SEU) representation was challenged by field and thought experiments in the spirit of Ellsberg (1961). The first proposals within the Rational Choice Theory approach to represent preferences with alternative attitudes toward this type of uncertainty appeared starting with Schmeidler (1986) and Gilboa and Schmeidler (1989). Then, a vast literature blossomed, whose works all share the underlying common motivation to accommodate for ambiguity aversion. The ensuing representation Theorems account for this feature but are often limited operationally in that they do not cleanly separate attitudes toward ambiguity from its perception. The representation of Klibanoff et al. (2005) is a notable exception, which allows the modeler to ask questions about the effects of varying degrees of both uncertainty aversion and perceived uncertainty. While risk and uncertainty attitudes shall be regarded as stable personal traits, elicited through consistent revealed preferences, perception can be elicited but may vary upon the context of the decision problem.

At the same time and in the same opus, von Neumann and Morgenstern (1944) put forth the fundamental elements for the development of a new framework to deal with economic problems more rigorously and flexibly: Game Theory. This field of study proved extremely prolific, and allowed the formalization of a rich set of solution concepts. While the most widespread and appreciated notions in the theoretical and applied literature may be Rationalizability (Bernheim, 1984; Pearce, 1984) and Nash (1950) Equilibrium, growing attention has been given to Self-Confirming Equilibrium, learning dynamics, and their interplay (Battigalli et al., 1992; Fudenberg and Levine, 1993, 1998). Steady-states of a learning process are characterized by stationary strategy profile distributions which are optimal under the equilibrium private information and beliefs of each player. Self-Confirming Equilibrium is the solution concept we apply for it forms a natural domain to explore the relationship between decision-makers, their knowledge, and behavior under uncertainty (Battigalli et al., 2015).

The reader may question the appropriateness of such theoretical approach for the problem at hand. Within the rational choice theory, decision models shall capture tolerable economic behavior. Starting with a minimal set of assumptions that appeal to some "normative" intuition, a utilitarian representation of preferences is derived through logical inference. Such assumptions are always held to a standard of rationality, albeit the latter term may *prima facie* seem quite vague. For instance, the transitivity assumption present in almost all economic models has been predicated upon Money Pump arguments (Gustafsson, 2022). Hence, it would be a maintained normative assumption because of its desirability and not due to its supposed veracity. Indeed, it has been pointed out that such arguments are often not very compelling in motivating the normative desirability of transitivity.[2] Conducting ourselves a formal investigation of such issues is beyond the scope of the discussion. Nevertheless, Gilboa et al. (2010) provide us with an intuitive motivation for the axiomatic approach. Imagine a decision-maker is held accountable of her decisions; would she necessarily feel embarrassed if asked to explain her reasoning?

This approach differs from Behavioral Economics, where traditional economic assumptions are typically traded off in favor of reliable psychological and neurological foundations to decision-making, thus introducing notions of biases or entertaining that individuals' rationality is bounded.[3] Consider, for instance, a traditional consumer problem under constraint. Traditionally, enlarging the choice set of available alternatives shall weakly improve the consumer's welfare. Following a behavioral approach, however, we may take into account the neurological and psychological evidence that the so-called Set-Size Effect is nonmonotonic.[4] While behavioral models offer valuable insights, this is not the approach followed here; the aim is to reflect on the implications of normative models of ambiguity aversion and perception.

The representation of Klibanoff et al. (2005) does not impose restrictions on the elicited subjective second-order probability assigned to models compatible with the evidence. Since the precise shape of the representation depends on preferences themselves, the analysis confronts decision-makers who ultimately differ in their "taste" in the processing of information. Hence, it may not be concluded that more naïve judgments are irrational. They may be attributed to mood, intuition; perhaps *animal spirits* have an influence on beliefs (Keynes, 1936; Akerlof and Shiller, 2009). Still, it is sensible to confront decision-makers who differ in their naïveté when confronting uncertain environments. Results help highlight the role of human attitudes in economic problems even within standard economic frameworks. Simple formal conditions on beliefs needed to uphold some equilibria rather than others may even lead, *inter alia*, to a better understanding of the effects of potentially ambiguous communication between institutions and the public.

Therefore, I explore how markets can plausibly get stuck into "certainty traps" without departures from (subjective) rationality. The Smooth Self-Confirming Equilibrium solution concept introduced by Battigalli et al. (2015) is applied to perform a comparative statics exercise on the equilibrium set of a game with feedback. Formally, I compare alternative sets of beliefs that differ in their perceived uncertainty. I leverage an example showing how more awareness of uncertainty (a doubt, or more intuitively a *certainty crisis*) can bring about objectively worse equilibrium strategies. The analysis is generalized to encompass games where own payoff-relevant information is not fully observable to every player.

In the rest of this work, I adopt whenever possible the formal notation used by Battigalli et al. (2023) and specify when is done otherwise. In Section 2, I provide a brief discussion of decision models used to deal with ambiguity attitudes, focusing on the representation of Klibanoff et al. (2005), its positioning within the literature, and the motivation for applying this criterion. I present a general definition of ambiguity-neutral (or Bayesian) Self-Confirming Equilibrium for games

[2] For instance, they "require a non-sequitur that calls for the additional hypothesis that a [decision-maker] willingly participates in a process that leaves him worse off". For a nuanced discussion of the normative appeal of transitivity and the treatment of intransitive preferences, see Fishburn (1991).

[3] These models are thus said to be "descriptive": they translate empirical regularities into assumptions, whose reliability rests on the modeler's trust that they are natural, i.e. do not depend on contextual factors.

[4] See for instance Iyengar and Lepper (2000) and Seuanez-Salgado (2006).

with feedback and payoff uncertainty, and its generalization to games with ambiguity attitudes introduced by Battigalli et al. (2015). In Section 3, I define ancillary games characterized by the sets of conjectures players may be assumed to hold when entering the game; these are used to produce results concerning equilibrium inclusion that complement the Theorems on comparative ambiguity aversion of Battigalli et al. (2015). In Section 4, I present an application of this comparative statics exercise to a simple contracting game that resonates with the literature on adverse selection. In Section 5, I draw conclusions and discuss limitations and possible extensions to my work. All proofs are contained in the Appendix.

2. AMBIGUITY ATTITUDES AND SELF-CONFIRMING EQUILIBRIUM

Throughout the following, I analyze the strategic form of a finite multistage game with payoff uncertainty, feedback and ambiguity attitudes within the formal framework developed by Battigalli et al. (2015) (hereon, BCMM). Section 2.1 details the structure of the game and introduces notation. It is convenient to consider the strategic form of the dynamic game and assume that, upon entering the game, players choose a fully specified strategy (*covert commitment*) that is then automatically implemented as the play unfolds.

It must be clear that this analysis is not equivalent to the extensive-form analysis of the same game whenever players are not ambiguity-neutral decision-makers. Dynamic consistency underpins the notion that "we can analyze the essential aspects of Self-Confirming Equilibrium for multistage games with feedback by looking at the strategic (or normal) form [since] any fixed strategy profile is a Self-Confirming Equilibrium of [the extensive-form game] if and only if there is a Self-Confirming Equilibrium [of the strategic form] inducing the same terminal history".[5] (Battigalli et al., 2023). A game with non-Bayesian decision-makers may not satisfy this property. If the intrapersonal version of the One-Deviation Principle fails, as is the general case for players averse to ambiguity, *ex ante* optimal strategies do not correspond to unimprovable strategies in the sense of Battigalli et al., 2019a, meaning that sophisticated players in extensive-form games who cannot fully commit to carrying out the strategy established ex-ante may have a strict preference for deviating somewhere along the path (provided that at least one player moves more than once). At the same time, this may not rule out altogether the existence of equilibrium profiles, which are also sequentially optimal.

Nevertheless, the strategic form analysis is both convenient from an analytic standpoint, and meaningful in classes of games where covert commitment is credible and in the special case of static games (degenerate multistage games where all players are active only at the root). More broadly, it is relevant in light of the relationship between the Self-Confirming Equilibrium solution concept and learning dynamics: monotonicity results regarding the strategic form stem from the influence ambiguity aversion has on learning *regardless* of dynamic inconsistency, as more uncertainty averse agents tend to stifle their experimentation after a few repetitions of a game (Battigalli et al., 2019b).

2.1. Game description

I adopt the *Nash mass action* (or anonymous interaction) view detailed in BCMM, and generalize their definitions of Smooth and Maxmin Self-Confirming Equilibrium to a setting with payoff uncertainty. According to the anonymous interaction interpretation, each player *role* $i \in I$ represents a large population, from which an agent is "drawn at random and matched to play the game [with agents drawn at random from all other roles $j \neq i$], then separated and rematched with (almost certainly) different opponents. After each play in which he was involved, an agent obtains some evidence on how the game was played." This conveniently allows to assume that any one player always only focuses on the present stage and disregards future payoffs ("myopic"

[5] The two equilibrium strategy profiles are said to be "realization-equivalent" in this case.

agents)—because she assigns negligible probability to the event of rematching with the same person twice. In addition, we assume uncertainty stems both from coplayers' strategies $(S_i)_{j\neq i}$, and their private information types $(\Theta_i)_{j\neq i}$ (along with nature Θ_0). The uncertainty space is thus $S_{-i} \times \Theta_{-i} \times \Theta_0$, where $S_{-i} := \times_{j\neq i} S_j$ and $\Theta_{-i} := \times_{j\neq i}\Theta_j$, and we only consider finite sets for simplicity. As described above, I restrict the analysis to the strategic-form of the game. Thus analogously to BCMM, with the addition of information types, the rules of the game make up game form with feedback[6]

$$\langle I, \Theta_0, q_0, (\Theta_i, q_i, S_i, M_i, F_i)\rangle$$

- I is the (finite) set of player roles;
- $\Theta := \Theta_0 \times \bigtimes_{i\in I} \Theta_i$ is the set of information type profiles —which is relevant to the game form to the extent that information determines the (outcome associated with each) terminal history;
- $q := (q_i)_{i\in I\cup\{0\}}$ is the profile of (exogenous) information type distributions;
- $S_i := \bigtimes_{h\in H} \mathcal{A}_i(h)$ is the individual finite strategy set of a player $i \in I$;
- M_i is the set of messages player $i \in I$ can observe *ex post* (whenever a play of the game is terminated);
- $F_i := f_i \circ \zeta : \Theta \times S \to M_i$ is the private strategic form feedback function of $i \in I$.

The extensive form of the game is not very relevant to us; nonetheless, I offer a brief description to foster intuition and clarify notation. For each t in a finite number of periods T, there is a non-empty subset $J_t \subseteq I$ of players who are *active* in the sense that they have to choose an action during that turn. Every sequence of profiles of actions taken from the beginning of the game draws a map, or *history*, through the nodes crossed during the game. The root of the game is the empty sequence. As players choose their actions at t, they reach history $h = (a^1, a^2..., a^t) \in \overline{H}$. H is the set of non-terminal histories, i.e., all decision nodes where there are still actions to be chosen by at least some player, and the game has not terminated yet. By contrast, Z denotes the set of terminal histories. Clearly, $H \cup Z = \overline{H}$ (while their intersection is empty). The action set of each player i is A_i. We may however want to allow for the specific set of available moves at any point in the game to depend on how the game has unfolded so far. Thus we denote by $\mathcal{A}_i : H \rightrightarrows A_i$ the feasibility correspondence of i, so that by definition S_i contains all possible *plans* of i, i.e. prescriptions on which actions she should carry out among those available depending on the contingency. $\zeta : S \to Z$ denotes the *path function*, which associates any profile of strategies to the terminal node it induces.

The feedback structure of the game form is described by the profile of functions $f_i : \Theta \times Z \to M_i$. When these are combined with the path function, we obtain profile $F = (F_i)_{i\in I}$ which gives rise, for each $i \in I, \theta_i \in \Theta_i$ and pure strategy $s_i \in S_i$, to an ex-post information partition of the uncertainty space,

$$\mathcal{F}_{s_i,\theta_i} = \left\{F^{-1}_{s_i,\theta_i}(m_i) \subseteq \Theta_0 \times \Theta_{-i} \times S_{-i} : m_i \in M_i\right\} \tag{1}$$

Battigalli et al. (2015, 2016, 2023) offer thorough discussions of general feedback structures, together with definitions of games with feedback and ambiguity attitudes. Importantly, it is the case that a decision-maker's (partial) identification of the plausible statistical models does not hinge on (is independent of) preferences and ambiguity attitudes, i.e. it can be described relying solely on the game form. On the other hand, information types distributions q are needed to derive the objective distribution of messages for each player given the decision functions mapping types into strategies.

The payoff of a player i, $u_i : \Theta \times Z \to \mathbb{R}$, depends on the true information types of the players $(\theta_i)_{i\in I}$ and the terminal history $z \in Z$ that is reached at the end of game. The payoff function itself

[6]The notation is explained below. More detail can be found in Battigalli et al. (2023).

is the composition $u_i := v_i \circ g$ of an outcome function $g : \Theta \times Z \to Y$ and a Bernoulli (or vN-M) utility $v_i : \Theta_i \times Y \to \mathbb{R}$. The strategic form payoff function of each $i \in I$ is $U_i := u_i \circ \zeta : \Theta \times S \to \mathbb{R}$. We so obtain the strategic form $G = \mathcal{N}(\Gamma)$ of a finite[7] multistage game with payoff uncertainty and exogenous type distributions Γ:

$$(G, F) := \left\langle I, \Theta_0, q_0, (\Theta_i, q_i, S_i, M_i, F_i, U_i)_{i \in I} \right\rangle \tag{2}$$

The ensuing structure has an evident mathematical similarity to (simple) Bayesian games with exogenous type-independent beliefs. In this framework, however, it must be understood that the game with payoff uncertainty is simply endowed with exogenous information-type distributions.[8] Players are assumed to know I, S, M_i, F_i, U_i and, importantly, θ_i. The overall interactive situation is instead *not* assumed to be common knowledge among the players. Additionally, players will not always be assumed to be Bayesian. An important assumption that is maintained[9] in the following is that of Observed Payoffs:

Definition 1 (**Observed Payoffs**). *Feedback function F_i satisfies Observed Payoffs relative to payoff function U_i if payoff only depends on feedback, that is, if*

$$\forall (s_i, \theta_i) \in S_i \times \Theta_i, \forall (s'_{-i}, \theta'_{-i}), (s''_{-i}, \theta''_{-i}) \in S_{-i} \times \Theta_{-i},$$
$$F_{\theta_i, s_i}(\theta'_{-i}, s'_{-i}) = F_{\theta_i, s_i}(\theta''_{-i}, s''_{-i}) \Rightarrow U_{\theta_i, s_i}(\theta'_{-i}, s'_{-i}) = U_{\theta_i, s_i}(\theta''_{-i}, s''_{-i})$$

Finite game with feedback (G, F) satisfies Observed Payoffs if the feedback functions of every $i \in I$ do.

In simpler terms, this property of feedback requires that players are always able to discern their realized utility. As will be discussed below, this fundamental assumption underpins the monotonicity results in BCMM because it ensures measurability of the section of the payoff function at the strategy that has been played in the long-run. This fact extends seamlessly to payoff-uncertain settings. As per their discussion, while this assumption may be reasonable in many games of interest, it is not to be taken for granted *a priori.* For instance, if players had altruistic motivations with regard to others but could not directly observe their internal utility index, this assumption would be quite far-fetched. The description of the rules of the game is complete. Next I introduce some decision criteria used to deal with ambiguity attitudes. Through these, we will revisit the traditional, Bayesian definition of Self-Confirming Equilibrium under payoff uncertainty, allowing the analysis of broader patterns of behavior.

2.2. **Preferences and decision criteria**

The representation of behavior in uncertain environments requires a decision-maker (DM) to be able to rank acts that yield (possibly risky) consequences dependent upon each state of the world, whose probabilities are unknown. In our setting, an individual player of type $\theta_i \in \Theta_i$ from population $i \in I$ is faced with the choice of a strategy $s_i \in S_i$. The outcome of any such strategy is uncertain and depends on the choices and types of coplayers. The uncertainty space of this choice is thus described by the Cartesian Product of the sets of other agents' strategies, private information types, and residual uncertainty: $S_{-i} \times \Theta_{-i} \times \Theta_0$. The DM attaches to each strategy s_i a subjective

[7]All sets that are part of the structure are finite.

[8]The problem of terminology in Game Theory is a tricky one. Harsanyi (1967, 1968a,b) originally introduced the notion of Bayesian Games and Bayesian Equilibrium. His definition is appropriate to analyze game structures that contain a unique profile of prior beliefs over a relevant uncertainty space. In our context, the word "(non-)Bayesian" is used to denote the decision attitudes of these kinds of players. However, we shall not speak of a Bayesian game, because we allow for profiles of sets of priors, and therefore, if players are not ambiguity neutral, i.e. they do not evaluate events according to their predictive probability (cfr. Section 2.2), there is not an evident way to update their beliefs according to Bayes' rule.

[9]It will be stated explicitly whenever we do *not* make this assumption.

value, denoted $V_{\theta_i}(s_i)$, parametrized by her tastes, characteristics and beliefs. To directly relate the player's decision problem to a framework of decision under uncertainty, this outcome uncertainty space may be interpreted as the *grand state space* discussed e.g. in Gajdos et al. (2008) and Gilboa and Marinacci (2013). Beliefs regarding the state space are constrained solely by the objective information an agent possesses.

Savage (1954) axiomatized the well-known SEU criterion, which extended the classical Expected Utility framework used to analyze lotteries with known probabilities to the more general case of Knightian uncertainty, eliciting the DM's subjective probability (a belief over states of the world) by requiring that he be able to (consistently) rank bets. Applications of this model have proved incredibly prolific in Game Theory and Economics. An alternative axiomatization of the same decision criterion in a more flexible setting is due to Anscombe and Aumann (1963). However, the thought experiments proposed by Ellsberg (1961) challenged (the normative desirability of) the SEU criterion, and led to the formulation of decision models whereby, even assuming the DM had a unique probabilistic model (the equivalent of a lottery) in mind for the problem at hand, faint conviction in such model could drive her to avoid uncertainty. This critique conflicted with one of Savage's postulates, the "Sure Thing Principle", and correspondingly with the Independence axiom in the Anscombe-Aumann formulation.

Ellsberg's first experiment involves two urns, I and II, containing 100 balls each. All balls are either white or black; in Urn I there are 50 black and 50 white balls, whereas nothing is known about the composition of Urn II. A DM must choose an urn; then, a ball is drawn from each of the two urns, and the DM is given a prize if the ball extracted from the urn of her choice is black. After this, the balls are put back in their respective urn, and the DM must choose an urn again; the draw is repeated, but the new prize is given to her only if the ball is white. Oftentimes, the known bet of urn I is preferred in both cases, a phenomenon that is incompatible with the evaluation of acts through a unique, additive probability measure. Allowing for the decision patterns that emerged in Ellsberg's experiments necessarily requires abandoning at least some of the properties of the SEU representation. A variety of approaches have been proposed to deal with this type of decision-making. These include considering non-additive probabilities or *capacities*, allowing for multiple subjective probability measures, and incorporating context-dependent beliefs. I begin by introducing some decision models that explore these possibilities. The overview will provide support in favor of the specific decision criterion adopted for the game-theoretic analysis.

The first axiomatization of a non-Bayesian decision model, which does not impose additivity on the DM's elicited subjective probability was the Choquet Expected Utility (CEU) representation of Schmeidler (1986, 1989). By restricting the Independence axiom of the Anscombe-Aumann formulation to comonotonic acts, i.e. acts whose ranking is independent of the state of the world, all acts are evaluated by means of a subjective capacity, through Choquet integration (Choquet, 1953). Nonadditivity of capacities allows a DM to express mild confidence in the relative likelihood of events. Schmeidler also showed that, if a DM has preferences for hedging against uncertainty, the capacity itself is endowed with a peculiar convexity property, which makes Choquet integration over the capacity akin to minimization of the integral over the probability measures in the core. The core is the set of additive probabilities whose lower bound on any measurable event contained in the event algebra is given by the value of the capacity itself. There are, however, situations in which the objective information in possession of the DM lets her entertain a set of plausible prior beliefs that does not constitute the core of any capacity.[10]

Gilboa and Schmeidler (1989) further relaxed the Independence axiom, restricting it to mixing with lottery acts, i.e., acts whose consequences occur with known probabilities. Together with uncertainty aversion, this axiom leads to their Maxmin Expected Utility (MEU) representation. This model ranks acts according to their value in the worst-case scenario envisioned by the DM. In any decision environment, the DM deems possible some set of probability models, and this set is constrained by objective information. Let $\hat{\Sigma}$ represent such objective information. Then, in our

[10]See the example at p. 200 in Gilboa and Marinacci (2013).

game-theoretic setting, the DM can deem possible any $\hat{\pi} \in \hat{\Sigma} \subseteq \Delta(S_{-i} \times \Theta_{-i} \times \Theta_0)$. A frequentist MEU DM, who infers the probability of events strictly from their long-run relative frequency, and considers all possibilities that to her knowledge are compatible with her objective information, would select strategy $s_i \in S_i$ if

$$\forall s_i' \in S_i, \ \min_{\hat{\pi} \in \hat{\Sigma}} U_i(s_i, \theta_i, \hat{\pi}) \geq \min_{\hat{\pi} \in \hat{\Sigma}} U_i(s_i', \theta_i, \hat{\pi}) \tag{3}$$

where

$$U_i(s_i, \theta_i, \hat{\pi}) = \int_{\Theta_0 \times \Theta_{-i} \times S_{-i}} U_i(s_i, \theta, s_{-i}) \, \hat{\pi}(\mathrm{d}\theta_0, \mathrm{d}\theta_{-i}, \mathrm{d}s_{-i}) \tag{4}$$

is the Bernoulli utility of s_i under a risky but certain probability model $\hat{\pi}$ (a lottery). The above can often be reduced to

$$U_i(s_i, \theta_i, \hat{\pi}) = \sum_{(\theta_0, \theta_{-i}, s_{-i}) \in \Theta_0 \times \Theta_{-i} \times S_{-i}} U_i(s_i, \theta, s_{-i}) \, \hat{\pi}(\theta_0, \theta_{-i}, s_{-i}) \tag{5}$$

since typically only models characterized by finitely many consequences are considered. This is our case because game G is finite. When $\hat{\pi}$ is a Dirac measure $\delta(\theta_0, \theta_{-i}, s_{-i})$ on some deterministic profile, U_i is the vN-M utility of the certain outcome induced by that profile, which we have introduced as part of the game description in (2).

Criterion 3 is flexible and intuitive, if extreme, and motivates the MSCE definition of Section 2.3. It has been pointed out in the literature that it is not obvious why such a "paranoid" attitude shall characterize a DM who dislikes uncertainty. As stated by Gilboa and Marinacci (2013), however,

> "An individual who satisfies the axioms [of Gilboa and Schmeidler] can be thought of as if he or she entertained a set C of priors and maximized the minimal expected utility with respect to this set. Yet, this set of priors need not necessarily reflect the individual's knowledge. Rather, information and personal taste jointly determine the set C. Smaller sets may reflect both better information [or, we may add, more naïvete] and a less averse uncertainty attitude. "

The MEU model thus allows the DM to restrict her attention to a subset of the models which are not contradicted by objective information in her possession. If the set of priors entertained by the DM is some $C \subset \hat{\Sigma}$, she chooses $s_i \in S_i$ if

$$\forall s_i' \in S_i, \ \min_{\hat{\pi} \in C} U_i(s_i, \theta_i, \hat{\pi}) \geq \min_{\hat{\pi} \in C} U_i(s_i', \theta_i, \hat{\pi})$$

In the revealed preferences tradition, the MEU model doesn't require frequentist decision-making. As Gilboa et al. (2010) put it, an uncertainty-averse DM can convince others that decisions made through the Unanimity criterion (Bewley, 2002) are rational, while she cannot be convinced that she is wrong in making decisions by completing her preferences through the MEU criterion. Such completion through revealed preferences need not be frequentist in nature. There are, however, ways to more starkly separate the DM's attitude toward uncertainty from her perception of, and information regarding, uncertainty. The approach followed here is to assume information-type θ_i has SEU preferences over "second-order acts," i.e. expected utility profiles induced by the array of plausible objective models. This requires him to hold a second-order subjective probabilistic belief

$$\mu^i_{\theta_i} \in \Delta(\hat{\Sigma}) \subseteq \Delta^2(\Theta_0 \times \Theta_{-i} \times S_{-i})$$

over such plausible models.[11] Then, negative attitudes toward uncertainty may be described by a functional which "penalizes" more uncertain beliefs. Some belief μ is regarded as more uncertain

[11] As a matter of notation *and* interpretation, a conjecture of a player i may be common or differentiated across types and strategies chosen. For now, consider a level of generality whereby at least types may hold different conjectures, because this will be an important general concept for the later equilibrium analysis. That the conjecture may differ for players playing different strategies is something we informally take for granted here, whereas it will be stated explicitly when necessary (e.g. for equilibrium definitions).

(or uncertainty aware) than another ν if μ is a mean-preserving spread of ν (see Section 3). According to the smooth representation introduced by Klibanoff et al. (2005) [hereon, KMM], an uncertainty-averse DM evaluates available act (strategy) $s_i \in S_i$ under the subjective second-order prior according to criterion

$$V_{\theta_i}^{\varphi_i}\left(s_i, \mu_{\theta_i}^i\right) = \varphi_i^{-1}\left(\int_{\text{supp}\mu_{\theta_i}^i} \varphi_i(U_i(s_i, \theta_i, \hat{\pi}))\ \mu_{\theta_i}^i(d\hat{\pi})\right) \tag{6}$$

where φ_i is a continuous, strictly increasing and concave[12] real-valued function. One of the significant advantages of criterion (6) is that it is able to separate a DM's ambiguity attitudes from her ambiguity perception. This tractability allows to analyze the two features separately. Attitudes toward risk and ambiguity are typically regarded as stable personal traits; while perception of uncertainty may depend on stable personal features as well, it does not seem far-fetched to suppose it is often affected by contextual factors.

The shape of function φ_i captures the player's ambiguity attitudes. The support of the second-order prior $\mu_{\theta_i}^i$ may be interpreted "cognitively" (Wald, 1949; Gilboa and Marinacci, 2013): a motivation for a smooth ambiguity representation may be the desire to represent the behavior of a DM who does *not* exclude any plausible model consistent with objective information from her decision problem, and nevertheless does not evaluate acts solely based on the worst-case scenario for each. This is, however, far from necessary for the representation; in fact, supports of beliefs upholding criterion (6) will be conveniently used to describe varying degrees of uncertainty perception in the game theoretic analysis of Section 3. That is, although objective information might *not* rule out some objective statistical model, the DM may unwillingly ignore it. Note that if $\mu_{\theta_i}^i$ is a Dirac measure on some objective statistical model over the uncertainty space, the DM behaves as a naïve Bayesian; that is, she *de facto* entertains a unique first-order prior.

The presence of operator φ_i^{-1} in (6) constitutes an innocuous normalization that will facilitate comparison of subjective values with risky but unambiguous alternatives. "Ambiguity neutrality" corresponds to the case where φ_i is affine: in this case, the player computes a SEU calculation with respect to the *predictive probability* of each profile of strategies and information types, given by

$$\hat{\mu}_{\theta_i}^i(\theta_0, \theta_{-i}, s_{-i}) = \int_{\text{supp}\mu_{\theta_i}^i} \hat{\pi}(\theta_0, \theta_{-i}, s_{-i})\mu_{\theta_i}^i(\mathrm{d}\hat{\pi}) \tag{7}$$

This is no more the case for players who are even moderately averse to ambiguity; this key aspect will play a pivotal role in the analysis of Section 3. According to the KMM criterion, an agent i is characterized as *more ambiguity averse* than j if there exists a function h such that $\varphi_i = h \circ \varphi_j$, where h is strictly increasing and concave or, equivalently, when both φ_i and φ_j are twice continuously differentiable, if $\varphi_i''/\varphi_i' \leq \varphi_j''/\varphi_j'$. When the ambiguity aversion coefficient $-\varphi_i''/\varphi_i'$ converges to infinity, we obtain the limit case of *extreme* ambiguity aversion, denoted ω, corresponding to the Maxmin criterion:

$$V_{\theta_i}^{\omega}\left(s_i, \mu_{\theta_i}^i\right) = \min_{\hat{\pi} \in \text{supp}\mu_{\theta_i}^i} U_i(s_i, \theta_i, \hat{\pi}) \tag{8}$$

Note that, in this case, the set $\text{supp}\mu_{\theta_i}^i \subseteq \hat{\Sigma}$ does *not* reflect ambiguity attitudes—it only embodies objective information *and its processing* on the part of the DM. With this, we are now equipped to embark on the game-theoretic analysis. The models presented above constitute fundamental contributions to Decision Theory, as they axiomatize revealed preferences inconsistent with SEU within Rational Choice Theory. The survey in this Section is far from exhaustive; its scope is limited to the purpose of understanding uncertainty averse decision making, how it favors the *status quo*, and the normative interplay that exists among ambiguity attitudes (personal

[12] If the DM is uncertainty-loving instead, φ_i is convex.

tastes), knowledge, and perception. In later sections, only the KMM model will be used; the above exposition is, however, crucial to interpret correctly the results of Section 3.

In economic problems, usage of uncertainty-averse (or loving) decision criteria may heavily affect predictions, as they characterize individual incentives differently and, therefore, also equilibrium sets according to any solution concept that does not assume perfect monitoring. In other words, departure from the *Bayesian paradigm* allows to encompass broader choice patterns, which address in a descriptively consistent while normatively desirable way Knightian uncertainty. This is especially palatable given that (i) "unmeasurable uncertainty" is pervasive in the evolution of recurrent anonymous interaction phenomena, e.g., in the stock market, as pointed out by Keynes (1936), and (ii) Ellsberg's Paradox, together with the ensuing academic discussion, has shown that neutrality toward uncertainty may not be necessarily regarded as the most normatively appropriate attitude.

2.3. Self-confirming equilibrium with ambiguity attitudes

The first definitions for Conjectural or Self-Confirming Equilibrium were put forth by Battigalli and Guaitoli (1988) and Fudenberg and Levine (1993). Both assumed the SEU decision criterion, and were suited for the analysis of games in extensive-form. Battigalli and Guaitoli had subsumed the results of their independent undergraduate theses, and their framework allowed for general feedback functions. On the other hand, while Fudenberg and Levine's work more specifically assumed that players could perfectly observe ex-post the path of play, their analysis would apply to large population games, justifying the assumption that players exclusively focus on their current-stage payoff even if they are not impatient.

A general definition of anonymous-interaction Self-Confirming Equilibrium for games with feedback and payoff uncertainty, which incorporates the contributions above, and in which all players are SEU maximizers, can be adapted from Battigalli et al. (2023). In a Self-Confirming Equilibrium, players respond rationally to confirmed conjectures about the behavior of coplayers. Information feedback for each player i is described by the private feedback function $F_i : \Theta \times S \to M_i$.[13] A player of type θ_i playing strategy s_i who holds conjecture $\mu^i_{s_i,\theta_i}$ expects to receive each message m_i with probability

$$\mathbb{P}^{F_i}_{s_i,\mu^i}(m_i|\theta_i) := \sum_{\theta_0,\theta_{-i},s_{-i}:F_i(\theta,s)=m_i} \mu^i_{s_i,\theta_i}(\theta_0,\theta_{-i},s_{-i}) \tag{9}$$

whereas, if coplayers are collectively enacting mixed strategy profile σ, random matching and a standard law of large numbers argument imply that in a steady state the long-run observed frequency of each message is given by

$$\mathbb{P}^{F_i}_{s_i,\sigma_{-i},q_{-i}}(m_i|\theta_i) := \sum_{\theta_0,\theta_{-i},s_{-i}:F_i(\theta,s)=m_i} q_0(\theta_0) \prod_{j\neq i} \sigma_j(s_j|\theta_j) q_j(\theta_j) \tag{10}$$

To understand the relevance of these formulae, consider a game that has already been repeated several times, such that a stationary state has been reached. That is, we are in "equilibrium," which I shall define explicitly shortly below. A player of type θ_i from population i has been playing s_i for a long time, and wants to determine whether it is sensible to keep doing so. I implicitly assume that players exhibit inertia, that is, unless they have a *strict* incentive to deviate, they do not do so.[14]

[13] To ease intuition and exposition, the notation is kept consistent throughout with the strategic-form framework introduced at the beginning of Section 2. It should be kept in mind, however, that the scope of the definition of Bayesian Self-Confirming Equilibrium is more general and adapts seamlessly to games in extensive-form.

[14] This assumption is standard in equilibrium analysis, and the reason should be intuitive: if we assumed the opposite, players may keep bouncing from one strategy to another. Even though this does not affect their own payoffs *directly*, it may dynamically alter coplayers' incentives, leading to chaotic interaction. Although it is in principle totally legitimate to conduct such an analysis, this would be formally much more complex, and equilibria (which would then need to be defined differently) may fail to exist even in relatively simple games.

Through the long-run objective distribution of messages described by Equation 10, θ_i is able to partition the uncertainty space (cfr. Equation 1) and infer the relative frequency of (unobservable) subsets of the uncertainty space, which we also call *events*. At the same time, θ_i enters a play of the game holding a conjecture $\mu^i_{s_i,\theta_i}$ which describes what he believes of the uncertainty space and, in turn, of the relative frequency of the messages he can observe. Is the player free to hold *any* conjecture? In principle, yes, but a key assumption we make in Self-Confirming Equilibrium analysis is that players have sufficient cognitive ability not to hold beliefs that are inconsistent with the stationary distribution of observed messages.

Suppose you are playing Texas Hold 'em against a friend whom you suspect is a *cheater* (this would be his information type). In the beginning, you do not have evidence for or against this hypothesis, so you are free to trust him as you deem fit. This remains the case as long as what you observe does not definitively imply that he is either cheating or not. However, after a long time, you notice that he folds 10% of the time, and *every single time* he does not, he was "lucky" enough to have drawn Double Aces from the deck. Next time, would it not be foolish *not* to assume he has Double Aces in his hand? Assuming the deck is fair, the number of plays does not have to be very large for you to rule out that the observed message frequency (i.e., the outcome of the games he keeps winning) is inconsistent with him being of type *non-cheater*. This naturally leads us to the definition of (Bayesian) Self-Confirming Equilibrium.

A conjecture is confirmed in the long run if it is not disproven, that is, if the subjective probability assigned to the event of observing any given message corresponds to its observed frequency. When a player's conjecture is confirmed, she has no incentive to deviate to another strategy, if she is acting optimally already. These facts provide the two conditions for stationarity that uphold equilibrium.

Definition 2 (SCE)**.** *Consider a finite game G and a profile of feedback functions $F = (F_i)_{i\in I}$. A profile of mixed strategies and conjectures*

$$\left(\sigma_i^*(\cdot|\theta_i), \left(\mu^i_{s_i,\theta_i}\right)_{s_i\in supp\sigma_i(\cdot|\theta_i)}\right)_{i\in I,\theta_i\in\Theta_i} \in \bigtimes_{i\in I} \left(\Delta(S_i)\times\Delta^{S_i}(S_{-i})\right)^{\Theta_i}$$

is an ***anonymous self-confirming equilibrium*** *(SCE) of the game with feedback (G,F) if, for every role $i\in I$, information type $\theta_i\in\Theta_i$ and strategy $s_i\in supp\sigma_i^*$, the following conditions hold:*

1. *(rationality)* $\sigma_i(s_i|\theta_i) > 0 \Rightarrow \forall \tilde{s}_i \in S_i, U_i\left(s_i,\theta_i,\mu^i_{s_i,\theta_i}\right) \geq U_i\left(\tilde{s}_i,\theta_i,\mu^i_{s_i,\theta_i}\right)$
2. *(confirmation)* $\mathbb{P}^{F_i}_{s_i,\mu^i}(\cdot|\theta_i) = \mathbb{P}^{F_i}_{s_i,\sigma^*_{-i},q_{-i}}(\cdot|\theta_i)$

This definition allows for fractions of agents in the same population to hold different assessments depending on their information type and the strategies they are playing. This is important as the partition of the uncertainty space can differ depending on these parameters. Note that an equilibrium is identified as a profile of strategies *and beliefs*—the former alone only make up an *equilibrium (strategy) profile*. BCMM discuss the generalization of the SCE solution concept to non-neutral attitudes toward uncertainty, comparing sets of equilibria underpinned by increasing degrees of ambiguity aversion, assuming that players are KMM decision-makers. Their definition is easily generalized to a payoff-uncertain setting. Let $\nu^i := \mathbb{P}^{F_i}_{s_i,\sigma^*_{-i},q_{-i}}(\cdot|\theta_i) \in \Delta(M_i)$ denote the long-run distribution of messages observed by θ_i playing s_i within equilibrium strategy profile σ^*; then

$$\hat{\Sigma}_{-i,\theta_i} := \hat{\Sigma}_{-i,\theta_i}(s_i,q_{-i},\sigma^*_{-i}) = \left\{\hat{\pi}\in\Delta(\Theta_0\times\Theta_{-i}\times S_{-i}) : \hat{F}_{s_i,\theta_i}(\hat{\pi}) := \hat{\pi}\circ F^{-1}_{s_i,\theta_i} = \nu^i\right\} \tag{11}$$

denotes the *partial identification set*, that is, the set of all distributions of profiles of strategies and types deemed possible from θ_i's long-run empiricist perspective. This is the hard information constraint faced by each DM, which in Section 2.2 we had more informally denoted by $\hat{\Sigma}$. Since the distributions contained in this set are undistinguishable by construction from θ_i's perspective, they form the domain of uncertainty of his or her decision problem. This fact motivates a more general definition of Self-Confirming Equilibrium, which encompasses the possibility of non-Bayesian DMs. Denote a game with feedback and ambiguity attitudes by (G, φ): this structure differs from (G, F) in that we add a profile $\varphi := (\varphi)_{i \in I}$ of ambiguity smoothing functions, in the sense of Equation 6.

Definition 3 (SSCE). *A profile of strategy distributions $\sigma^* = (\sigma^*_{\theta_i})_{\theta_i \in \Theta_i, i \in I}$ is a **smooth** Self-Confirming Equilibrium (SSCE) profile of game with feedback and ambiguity attitudes (G, φ) if, for every $i \in I$, $\theta_i \in \Theta_i$, and $s_i \in \text{supp}\sigma^*_{\theta_i}$, there exists μ_{s_i,θ_i} such that*

1. *(rationality) $\forall \tilde{s}_i \in S_i, V^{\varphi_i}_{\theta_i}\left(s_i, \mu_{s_i,\theta_i}\right) \geq V^{\varphi_i}_{\theta_i}\left(\tilde{s}_i, \mu_{s_i,\theta_i}\right)$*

2. *(confirmation) $\text{supp}\mu_{s_i,\theta_i} \subseteq \hat{\Sigma}_{-i,\theta_i}$*

For a strategy profile to constitute a SSCE, there must exist a profile of assessments *consistent with long-run feedback* (that is the role of $\hat{\Sigma}_{-i,\theta_i}$) such that the subjective value of the game as evaluated through φ_i by each player is greatest when choosing that strategy. It is important to note that the existence of the profile of assessments here constitutes a necessary condition for a strategy profile to constitute an equilibrium, and it is not included as a component itself. In other words, the equilibrium profile *is* the equilibrium strategy profile. This feature is formally convenient to describe the monotonicity results of Section 3. For this reason, in the remainder, when I write "SCE" to denote an equilibrium set, I will also be referring to equilibrium strategy profiles only, leaving beliefs in the background.

This definition differs from BCMM in that the partially identified set of distributions $\hat{\Sigma}_{-i,\theta_i}$ is a subset of the larger space $\Delta\left(\Theta_0 \times \Theta_{-i} \times S_{-i}\right)$. The original analysis does not explicitly discuss payoff uncertainty.[15] BCMM refer to games with chance moves; the latter can be alternatively modeled as payoff parameters (see Section 3.5). At the same time, separately modeling payoff uncertainty allows to describe more flexibly the *ex ante* distribution of information in a game. One can more naturally interpret information types as private, stable features of individuals.[16] BCMM's results rely on an important Lemma upheld by the assumption of Observed Payoffs (Definition 1), extended here to enclose payoff uncertainty. First, note that we can define the objective expected payoff of playing strategy s_i for type θ_i when coplayers are behaving according to σ^*_{-i} as[17]

$$U_i(s_i, \theta_i, q_{-i}, \sigma^*_{-i}) := \sum_{(\theta_0, \theta_{-i}, s_{-i}) \in \Theta_0 \times \Theta_{-i} \times S_{-i}} U_i\left(s_i, \theta, s_{-i}\right) q_0(\theta_0) \prod_{j \neq i} \sigma^*_j\left(s_j | \theta_j\right) q_j(\theta_j) \tag{12}$$

Lemma 1. *If payoffs are observable in game G, then for every $i \in I, \theta_i \in \Theta_i, s_i \in S_i$ and $\sigma^*_{-i} \in \Delta(S_{-i})$,*

$$\forall \hat{\pi} \in \hat{\Sigma}_{-i,\theta_i}\left(s_i, q_{-i}, \sigma^*_{-i}\right), \qquad U_i(s_i, \theta_i, \hat{\pi}) = U_i(s_i, \theta_i, q_{-i}, \sigma^*_{-i})$$

[15] If the game has no payoff uncertainty, it does not necessarily feature complete information, which would instead correspond to the informal assumption of the occurrence of event

$$\bigcap_{n \in \mathbb{N}_0} \text{(everybody knows that)}^n \text{ everybody knows all the rules of the game.}$$

[16] Modeling the distribution of information through chance moves would also introduce significant technical complications for SCE analysis. A game representation equivalent to the one above, if carried out through alternative paths of play induced by realizations of chance moves, would hinder clarity and require a fastidious description of what players can observe regarding those chance moves.

[17] Computation of expected payoff according to Equations 4 and 5 requires one-step integration as subjective models $\hat{\pi}$ are joint distributions of strategies and types. These models can be decomposed to obtain marginal beliefs on types, and conditional beliefs on strategies depending on types. Hence the notation abuse in Equation 12, where U_i takes in an extra argument, is moot.

In words, the objective expected payoff of playing strategy s_i is constant across all distributions in the partially identified set. This is quite natural for stationary distributions as, under payoff observability, a DM shall be able to rule out distributions inducing expected payoffs different from the long-run time average observed when playing s_i. BCMM's main result relies on this fact: as long as beliefs are consistent with evidence, their support and shape do not influence the value of the strategy being played—whereas they indeed (potentially) affect the value of untested strategies. The analysis in Section 3 also hinges on this important observation.

As a limit case of KMM, aversion to uncertainty takes the maxmin form of Gilboa and Schmeidler (1989)—although recall from the discussion in Section 2.2 that the set of priors is not to be interpreted in the same way. The following definition is provided separately, because the Maxmin criterion may admit a strictly frequentist interpretation, intended here, where revealed preferences are not the result of a set of priors:

Definition 4 (MSCE)**.** *A profile of strategy distributions $\sigma^* = (\sigma^*_{\theta_i})_{\theta_i \in \Theta_i, i \in I}$ is a* ***maxmin*** *self-confirming equilibrium (MSCE) profile of a game with feedback (G, F) if, for every $i \in I$, $\theta_i \in \Theta_i$ and $s^*_i \in supp\sigma^*_i$,*

$$\min_{\hat{\pi} \in \hat{\Sigma}_{-i,\theta_i}(s^*_i, q_{-i}, \sigma^*_{-i})} U_i(s^*_i, \theta_i, \hat{\pi}) \geq \min_{\hat{\pi} \in \hat{\Sigma}_{-i,\theta_i}(s^*_i, q_{-i}, \sigma^*_{-i})} U_i(s_i, \theta_i, \hat{\pi}) \ \forall s_i \in S_i \tag{13}$$

BCMM establish the following important monotonicity result: if a game with ambiguity attitudes is more ambiguity averse than another, in the sense that they share the same underlying structure but every player is weakly more ambiguity averse than the former, the set of smooth equilibria weakly expands. This surely implies that, in general, the set of SSCE is a weak superset of the set of SCE determined by SEU-maximizing players (Definition 2).

Recall that, in the long-run, the value of tested strategies is known objectively, whereas that of untested ones remains ambiguous. BCMM state that "keeping beliefs fixed, [the latter] is higher when ambiguity aversion is lower": the more a player is ambiguity averse, the more he exhibits status quo bias. As will be shown, a related complementary result holds. Keeping ambiguity aversion fixed, the value of untested strategies is in general lower the more uncertain the belief (the more the DM perceives ambiguity, or is troubled by the imprecision in the information she is presented with).

The definition of SSCE assigns an instrumental role to beliefs. The underlying motivation is that, in principle, one may entertain any belief; thus, it is best to consider all of them. Section 2.2 clarifies how, absent additional assumptions, a prior whose support is a strict subset of $\hat{\Sigma}_{-i,\theta_i}$ characterizes personal taste and characteristics in the processing of objective information, given the context. It is thus a stretch to assume that players may not be characterized in terms of their perception, and that we cannot compare agents who differ in this respect. It is sensible to discuss how non-Bayesian DMs may interact differently depending on their naïveté, as embodied by the width of the support of their second-order prior, *for a fixed degree of ambiguity aversion.* Thus in Section 3 an ancillary framework is developed to apply the SSCE solution concept to "doubtful" thinking. That is, I will consider how alternative beliefs entailing varying degrees of ambiguity perception impact steady states. The implications will highlight the effect that doubts early in the game may have on equilibrium.

2.4. **A note on payoff observability**

In the definitions of SSCE and MSCE, the subjective value optimization under some profile of conjectures is sufficient to ascertain that the candidate strategy profile is indeed an equilibrium, due to Lemma 1. Since any statistical model in the support of a conjecture shall yield the same subjective value as the conjecture that corresponds to the true distribution of strategies and types, it does not matter which conjecture is actually held within the set $\hat{\Sigma}_{-i,\theta_i}$ for determining the subjective value of the strategy that is currently being played. As pointed out by BCMM, however, the set

$$\left\{U_i(\tilde{s}_i, \theta_i, \hat{\pi}) : \hat{\pi} \in \hat{\Sigma}_{-i,\theta_i}(s_i^*, q_{-i}, (\sigma^*_{\theta_k})_{\theta_k \in \Theta_k, k \neq i})\right\}$$

for a generic strategy $\tilde{s}_i$ distinct from s_i^* is in general not a singleton. This is because partitions of the uncertainty space induced by alternative pure strategies need not correspond for any given type—and this is what drives action in favor of the status quo.

Let $\hat{\sigma}$ be some Bayes-Nash strategy profile of the game. If payoffs are not observable, even in the long-run it is not granted that θ_i will learn the expected payoff associated with any $s_i \in \text{supp}\hat{\sigma}_{\theta_i}$. This implies that the role of uncertainty becomes all the more relevant in determining decision-makers' course of action, as the appeal of less ambiguous strategies inherently grows. In this case, the monotonicity result of BCMM does not hold.

For instance, consider the case of constant (no) feedback: $\hat{\Sigma}_{-i,\theta_i}$ necessarily corresponds to the entire uncertainty space. Even though players know their payoff function, they cannot observe their realizations so that, as more distributions become consistent with the evidence (the partition induced by feedback becomes coarser), more paranoid, uncertainty-averse DMs may envision a larger set of possibilities. More conservative courses of actions may then offer higher minimal expected payoff than all $s_i \in \text{supp}\hat{\sigma}_{\theta_i}$ also if coplayers were playing according to $\hat{\sigma}_{-i}$, because the expected payoff induced by $\hat{\sigma}_{-i}$ can never be observed. This is indeed the case of Example 1 below: without feedback, an extremely uncertainty averse DM will never enter the MP subgame. While it is desirable to keep this in mind, in what follows the assumption of Payoff Observability is always maintained.

3. COMPARATIVE AMBIGUITY PERCEPTION

In this section I extend the SCE analysis under uncertainty aversion to consider restrictions on exogenous beliefs, more precisely on profiles of second-order priors. Section 3.1 introduces an ancillary game structure to deal with such restrictions, and shows that priors predicting the same distribution on average may induce nested sets of equilibria if they entail varying degrees of (perceived) ambiguity. Section 3.2 illustrates this reasoning through an example, whereby the threshold of ambiguity attitude needed to justify more conservative courses of action depends on the spread of the prior. The example considers finite support measures for instructive purposes, but it is important to note that the result has a more general scope and, in some games, even very small probabilities assigned to more extreme events by priors with infinite support may affect the decision-making outcome. This important theoretical point is stated and discussed in Section 3.3. The implications of another relevant assumption on feedback, namely "own-strategy independence of feedback", are explored in Section 3.4. The last subsection instead considers how payoff uncertainty affects the formal description of the game.

3.1. Stochastic dominance of beliefs and equilibrium

Definition 5. *Consider a game with feedback and ambiguity attitudes* (G, φ). *Denote by* (G, φ, μ) *the* μ*-restriction of the game, that is, the same game with the additional restriction on exogenous beliefs whereby the belief profile must be described by* μ. *Denote by* (G, φ, C) *the* C*-restriction of the game, where* $C := \bigtimes_{i \in I} C_i$ *and* C_i *is the collection of beliefs that player* $i \in I$ *is allowed to hold in the restricted game.*

Definition 5 yields an ancillary class of games identified by prior profiles.[18] Clearly, a strategy profile σ^* can be a SSCE of $(G, \varphi, \tilde{\mu})$ if and only if it is a SSCE of (G, φ) such that for every i, θ_i

[18]The equilibrium set of a game restricted in this way may be empty even if mixed equilibria exist, because exogenously posed priors may be inconsistent with any Nash strategy profile.

and $s_i^* \in \text{supp}\sigma_i^*$, we have $\text{supp}\tilde{\mu}^i \subseteq \hat{\Sigma}_{-i,\theta_i}$ and

$$V_{\theta_i}^{\varphi_i}(s_i^*, \tilde{\mu}_{\theta_i,s_i^*}) \geq V_{\theta_i}^{\varphi_i}(s_i, \tilde{\mu}_{\theta_i,s_i^*})\ \forall s_i \in S_i$$

Thus more broadly, σ^* is a SSCE of (G, φ, C) if and only if it is a SSCE of (G, φ) where the confirmation condition on the profile of beliefs μ of Definition 3 ($\text{supp}\mu^i_{s_i,\theta_i} \subseteq \hat{\Sigma}_{-i,\theta_i}$) is replaced with the stronger condition

$$\mu^i_{s_i,\theta_i} \in C_i \cap \Delta(\hat{\Sigma}_{-i,\theta_i})$$

This definition allows us to analyze how the equilibrium set may change depending on the prior belief profile held by players. To this end, the following result equips us with a comparative notion on priors that relies on stochastic dominance. The proof relies on the foundational work of Rothschild and Stiglitz (1970), and allows the derivation of Proposition 1 below.

Lemma 2. *If payoffs are observable, second-order prior $\mu^i \in \Delta(\hat{\Sigma}_{-i,\theta_i})$ is a (weak) mean-preserving spread of $\nu^i \in \Delta(\hat{\Sigma}_{-i,\theta_i})$ if and only if μ^i (weakly) second-order stochastically dominates ν^i.*

Probability distribution μ^i is said to be a mean-preserving spread of ν^i if, formally,[19] $\mu^i \overset{D}{=} \nu^i + z$ where z is random noise with null conditional expectation. In simpler terms, a mean-preserving spread of some distribution is another distribution with the same expected value which, however, assigns more weight to the tails of the original distribution. Note that mean-preserving spreads only make up a partial (incomplete) order. It can be shown that if μ^i is a mean-preserving spread of ν^i, then it has higher variance and identical expected value; the converse is not true in general, which can be seen by simply observing that variance induces a complete ordering. A special case of MPS obtains when the predictive probabilities of two beliefs coincide and the *convex hulls* of their supports are nested in each other.

Lemma 3. *If $\mu^i, \nu^i \in \Delta(\hat{\Sigma}_{-i,\theta_i})$ induce the same predictive probability, and* $\text{conv supp}\nu^i \subset \text{conv supp}\mu^i$*, then μ^i is a mean-preserving spread (MPS) of ν^i.*

The notion of MPS has long been studied in the literature of choice under risk, motivated by the objective of defining what makes some lottery "riskier" than another. In this setting, we are dealing with beliefs rather than lotteries; thus an MPS of some belief shall be interpreted as the belief in the same average outcome which is, at the same time, more concerned with more uncertain outcomes.

Consistently with the interpretation in choice under risk, then, one may say that a belief is a MPS of another if it is "more uncertain", or it *perceives* more (Knightian) uncertainty. Indeed, KMM had already noted that ambiguity aversion is "an aversion to the subjective uncertainty about ex ante evaluations. Analogous to risk aversion, aversion to this uncertainty is taken to be the same as disliking a mean preserving spread" in the distribution of expected utility values induced by a belief and a strategy. They also state that a "useful comparative statics exercise is to hold ambiguity attitudes fixed and ask how the equilibrium is affected if the perceived ambiguity is varied", an exercise which we conduct through the following important result, which complements the analysis of BCMM:

Proposition 1. *Let (G, φ) be a finite game with feedback and ambiguity attitudes, which satisfies Observed Payoffs (Def. 1) and in which all players are weakly ambiguity averse. If μ and ν are profiles of beliefs, and for each i, θ_i and s_i, $\mu^i_{s_i,\theta_i}$ is a (weak) MPS of $\nu^i_{s_i,\theta_i}$, with* $\text{supp}\mu^i_{s_i,\theta_i} \subseteq \hat{\Sigma}_{-i,\theta_i}(s_i, q_{-i}, \sigma^*_{-i})$ *for every $\sigma^* \in SSCE(G, \varphi, \nu)$, then*

$$SSCE(G, \varphi, \nu) \subseteq SSCE(G, \varphi, \mu)$$

[19]Symbol $\overset{D}{=}$ means "is identical in distribution".

Proposition 1 establishes that greater ambiguity aversion need not be the unique determinant of "certainty traps" within the same game form. In fact, it is also possible that players who are equally averse to ambiguity are not all equally aware of the range of models compatible with the evidence, or they disregard some possibilities. To see this point more clearly, recall that the definition of SSCE admits beliefs whose support is strictly contained in $\hat{\Sigma}_{-i,\theta_i}$. Status quo biases, therefore, do not necessarily follow from a more ambiguity averse attitude; they may result from strong ambiguity *perception*.

How shall stochastic dominance be interpreted? It is wrong to conclude that an ambiguity averse decision-maker "prefers" holding beliefs that entail smaller uncertainty. It is behavior that elicits the representation, not the converse. In other words, the second-order prior held by a decision-maker in an interactive situation is symptomatic of the extent to which she would consider all available possibilities before making a choice. The stochastic dominance argument simply embodies the idea that the status quo bias is stronger for agents who are bewildered by the presence of uncertainty, irrespective of *how* averse to it they may be.

We are led to a startling paradox: provided that players slightly dislike uncertainty, outright ignoring it shall grant objectively better payoffs. Conversely, for any given prediction regarding the average outcome, frantically evaluating all possibilities, including the most extreme ones, increases the likelihood of certainty traps, even if relatively little weight is assigned to such extreme possibilities. The derivation of Proposition 1 allows a straight-forward corollary whereby wider ambiguity perception restricts the reachable equilibrium set for ambiguity-loving players.

The assumption of mean preservation means we require that players hold the same prediction regarding the behavior and types of others, while the spread implies that in one case they believe a wider range of outcomes to be possible. This affects their confidence in the prediction, so that the same average prediction may act as basis for action in different ways. Note that so far I have disregarded cases where the predicted outcome is different; this point will be addressed in Section 3.3. The KMM criterion is rather helpful in the derivation of Proposition 1, because it allows to deal with probabilities. Analogous reasoning would result, however, if we were to represent the DM's behavior e.g. through the CEU model (Schmeidler, 1986). Instead of comparing mean-preserving spreads, one may start from a capacity, and reduce the weight assigned to less uncertain objective distributions, i.e. those leading to expected utility values closer to the mean.

Proposition 1 complements the formulation and analysis conducted in BCMM of the status quo bias induced by aversion to uncertainty: when information is insufficient, for ambiguity averse players, the more contemplated possibilities the more equilibria are possible. Restricting the prior beliefs of players confines the comparison to beliefs that can be partially ordered. However, it enables a first evaluation of the consequences of players second-guessing themselves. When many priors are possible, an agent enters the game with a belief determined by her taste, objective information, and how the latter is processed, and acts accordingly. The evidence she accumulates thereafter depends on previous plays. Ex-ante more paranoid DMs, those who incur a certainty crisis, stop learning and get stuck more easily.[20] This is shown in the example of Section 3.2.

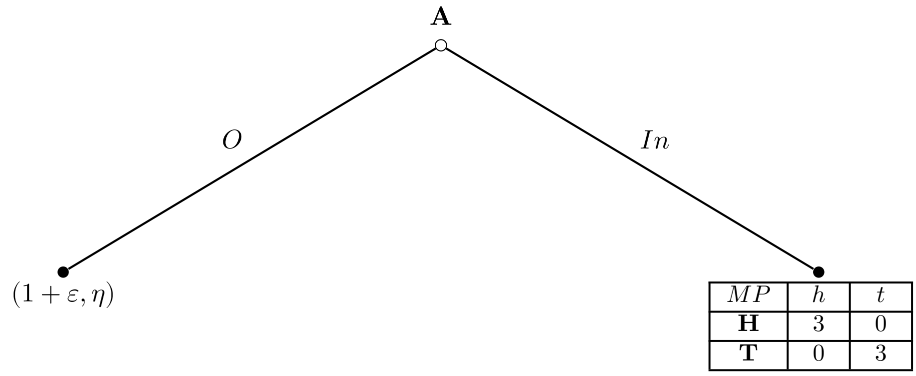
3.2. An illustration

Example 1. Consider the Matching Pennies game with an Outside option whose "tree" is depicted in Figure 1. Assume $\varepsilon \in \left(0, \frac{1}{2}\right)$. The game is a simplification of the running example in BCMM.

In this game without payoff uncertainty, player **A** repeatedly decides whether to leave the game (O) or enter and play matching pennies (MP) against Nature, which selects actions h, t according to a deterministic probability law $\sigma_0 \in \Delta(\{h, t\})$. Assume $F_{\mathbf{A}} = U_{\mathbf{A}}$, that is, **A**'s feedback coincides with her payoff function. If **A** plays **H** or **T**, in the long-run she learns her expected payoff and, given the game structure, also the exact distribution σ_0. However, if her initial belief is compatible with choosing strategy O, she may not learn anything and get stuck in the objectively

[20]For a rigorous argument on learning dynamics, see Battigalli et al., 2019b.

FIGURE 1: Matching Pennies with Outside Option



worse outcome. As in BCMM, this can never be the case for a SEU-maximizing player since for any conjecture $\overline{\mu}$ the subgame MP would necessarily have value

$$\max\{3\overline{\mu}, 3(1-\overline{\mu})\} \geq 1.5 > 1+\varepsilon$$

Now consider **A**'s decision-making for different priors. Let $\varphi_{\mathbf{A}} = U^{1/\alpha}, \alpha > 0$. For $\alpha \in (0,1)$, **A** would be ambiguity loving; for $\alpha = 1$, she'd be ambiguity neutral; otherwise, she is ambiguity averse. In general, the higher the value of α, the more $\varphi_{\mathbf{A}}$ is ambiguity averse.[21] Also let

$$\mu_{\mathbf{A},1} := \delta_{\frac{1}{2}\delta_h + \frac{1}{2}\delta_t}$$

$$\mu_{\mathbf{A},2} := \frac{1}{4}\delta_h + \frac{1}{2}\delta_{\frac{1}{2}\delta_h + \frac{1}{2}\delta_t} + \frac{1}{4}\delta_t$$

$$\mu_{\mathbf{A},3} := \frac{1}{2}\delta_h + \frac{1}{2}\delta_t$$

In words, if **A** believes $\mu_{\mathbf{A},1}$, she is certain that half the times Nature chooses h, and the other half it chooses t. If she believes $\mu_{\mathbf{A},2}$, she additionally deems it fairly plausible that either it always chooses h or always chooses t. If she believes $\mu_{\mathbf{A},3}$, she only believes these latter possibilities, and she assigns equal probability to them. Strategy O is a best-reply only if $\max\{V(H,\mu_{\mathbf{A}}), V(T,\mu_{\mathbf{A}})\} \leq 1$. Note that, by construction,

$$\hat{\mu}_{\mathbf{A},1} = \hat{\mu}_{\mathbf{A},2} = \hat{\mu}_{\mathbf{A},3} = \mu_{\mathbf{A},1}$$

where $\hat{\mu}$ is the predictive probability in the sense of Equation 7. Hence the expected utility profile for an ambiguity neutral agent would be the same for all three measures: on average, she expects the same behavior of Nature. For $\mu_{\mathbf{A},1}$, we have that

$$V^{\varphi}_{\mathbf{A}}\left(s_{\mathbf{A}}, \mu_{\mathbf{A},1}\right) = U_{\mathbf{A}}\left(s_{\mathbf{A}}, \mu_{\mathbf{A},1}\right) \ \forall s_{\mathbf{A}} \in S_{\mathbf{A}}$$

so that ambiguity attitudes have no bite and O can never be justified as a best-reply, because a prior with singleton-support believes no uncertainty is involved in the choice. On the other hand, observe that

$$V^{\varphi}_{\mathbf{A}}\left(\mathbf{H}, \mu_{\mathbf{A},2}\right) = V^{\varphi}_{\mathbf{A}}\left(\mathbf{T}, \mu_{\mathbf{A},2}\right) = \left(\frac{1}{4}3^{1/\alpha} + \frac{1}{2}1.5^{1/\alpha}\right)^{\alpha} \leq 1 \iff \alpha \geq 2.3 > \log_2(3)$$

[21]Note that this shape for $\varphi_{\mathbf{A}}$ entails a nonconstant degree of ambiguity aversion; see KMM for details.

and

$$V_{\mathbf{A}}^{\varphi}\left(\mathbf{H}, \mu_{\mathbf{A},3}\right) = V_{\mathbf{A}}^{\varphi}\left(\mathbf{T}, \mu_{\mathbf{A},3}\right) = \left(\frac{1}{2}3^{1/\alpha} + \frac{1}{2}0^{1/\alpha}\right)^{\alpha} \leq 1 \iff \alpha \geq \log_2(3)$$

By comparing $\mu_{\mathbf{A},1}$ with the other beliefs, we can observe that while the expected utility profile as expressed by the predictive probability is unchanged (and equal to $\mathbb{E}_{\mu_{\mathbf{A},1}}[U(In)] = 1.5$), for a sufficiently ambiguity averse player **A** these beliefs, which are mean-preserving spreads of the former, increase the likelihood of choosing the outside option. By comparing $\mu_{\mathbf{A},2}$ with $\mu_{\mathbf{A},3}$, we note that the latter *reduces* the support of the prior in terms of cardinality, but it does so while still increasing uncertainty in the payoff, because it assigns zero probability to the certain (in the sense of an objective lottery) outcome. Indeed, a moderately uncertainty averse **A** with $\alpha \in (\log_2(3), 2.3)$ may select O as a best reply if she held belief $\mu_{\mathbf{A},3}$, while she would not if she held $\mu_{\mathbf{A},2}$. $\triangleleft$

Note that all beliefs considered in the example are compatible with evidence observed by playing O. If **A** considers a wider range of possibilities, as long as she is sufficiently uncertainty averse she will avoid exploring even if she expects the same outcome on average. A greater perception of the threat presented by ambiguity makes the decision-maker more likely to get stuck into a suboptimal decision. In the example, it is especially interesting to confront beliefs whose supports are such that one is a superset of the other, as was the case for $\mu_{\mathbf{A},1}$ and $\mu_{\mathbf{A},2}$. Such a case can be more naturally interpreted as increased awareness of more remote possibilities on the part of the DM, rather than a belief envisioning exclusively different scenarios that just so happens to preserve the mean prediction.

It is also important to note that the objective value of the game given available information has not decreased, and that uncertainty aversion is present in the model in the same degree; it just is possible for it not to show and stay *hidden* as long as assessments are naïve enough, i.e. they are relatively unaware of the uncertainty involved in the choice. On the other hand, one can verify that the status of some strategy profiles as equilibria may not be affected by ambiguity perception. Without loss of generality, let $\sigma_0(h) > 1/2$ in the example and consider strategy $In.\mathbf{H}$; by playing the latter, **A** learns $\sigma_0(h)$ with certainty, and no ambiguity is present in her decision problem anymore, so that she may not hold conjectures that entail varying degrees of uncertainty.

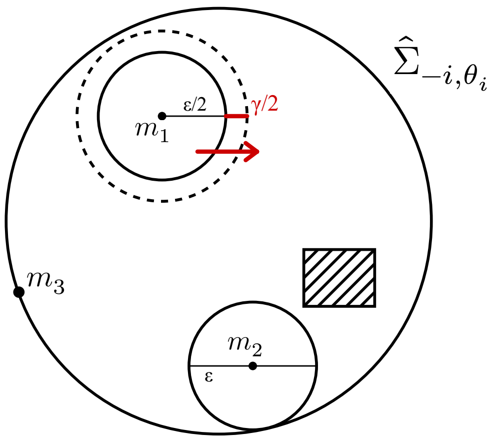
The example clarifies that, if players revise their prior beliefs in a way that does not alter predictions but is more aware of the underlying uncertainty, it may be easier to justify strategies and reach SSCE profiles which would otherwise be avoided. It highlights how the possibility of doubting own beliefs may bring about less ambiguous, though possibly suboptimal, equilibria.

3.3. **Generalization to sets of beliefs**

It is possible to derive a result on equilibrium inclusion which imposes weaker restrictions on exogenous beliefs, that is, where we compare nonsingleton sets of prior beliefs connected through second-order stochastic dominance. This is desirable as the assumptions of Proposition 1 might appear somewhat stringent. Recall that stochastic dominance only induces a partial order; therefore, we cannot compare beliefs which are not in such relation with each other. However, we may leverage pairwise relations among beliefs that belong to different sets.

Recall that for any subjective belief μ^i with nonsingleton support, we may find the associated predictive probability $\hat{\mu}^i$ through Equation 7. If we simply compare nested sets of beliefs, it is relatively trivial to conclude that the induced sets of justifiable strategies shall enlarge irrespectively of ambiguity attitudes. Thus we want to compare beliefs that can be put into the MPS relation with each other, i.e. vary in their perceived uncertainty, without imposing or forbidding other features of beliefs. In this respect, the convexity of $\hat{\Sigma}_{-i,\theta_i}$ is relevant to determine whether we can find in general a valid (in the sense of compatible with evidence) belief with support convex hull different from μ^i that yields the same predictive probability $\hat{\mu}^i$, so that moving from a set to another we are only adding beliefs that are second-order stochastically dominated by the rest. The potential non-convexity of $\hat{\Sigma}_{-i,\theta_i}$ is represented by the dashed square in Figure 2. Again following BCMM, we may grant convexity by assuming (informally) that players are totally ignorant about the matching process; that is, no one is aware that other players are randomly drawn *independently of each other*. This assumption is unimportant for the case of two-person games, as is the case for the game in Section 4.

FIGURE 2: An intuitive graphical representation of the partial identification set and supports of beliefs which belong to R_ε and R_ε^γ. m_k with $k = 1, 2, 3$ denote predictive probabilities.



The maximal degree of generality can be obtained by considering sets of beliefs where perceived uncertainty is varied *wherever possible*. In other words, we compare sets where we replace any belief that admits an MPS compatible with objective information with that MPS, and leave unchanged beliefs that do not allow so. This procedure avoids outright excluding some plausible expectations on the game play: note that the strategy profiles in $SCE(G, F)$ coincide with the set of $SSCE$ given by all singleton-support beliefs. Consider a belief with singleton support lying on the boundary of $\hat{\Sigma}_{-i,\theta_i}(s_i, q_{-i}, \sigma^*_{-i})$ for some $\sigma^*, s_i \in \mathrm{supp}\sigma^*_i$: there is no possibility to construct a MPS whose support lies within $\hat{\Sigma}_{-i,\theta_i}$, as the spread would necessarily involve including models from outside the set.

While dealing with infinite-support beliefs can be demanding, we may rely on Lemma 3, considering the support on the space of objective expected utility values induced by statistical models in $\Delta(\Theta_0 \times \Theta_{-i} \times S_{-i})$. Endow this space with the following topological structure: let objective models be subjectively related by the maximal L^1-distance on the space of expected utility values, which is a subset of the real line:

$$\forall p, r \in \Delta(\Theta_0 \times \Theta_{-i} \times S_{-i}), d_{\theta_i}(p, r) = \sup_{s_i \in S_i} \left| \int U_i(s_i, \theta_i, \cdot)\, \mathrm{d}p - \int U_i(s_i, \theta_i, \cdot)\, \mathrm{d}r \right|$$

Then, given $\varepsilon \in \mathbb{R}_+$, we can define a set R_ε as the smallest closed set of beliefs whose supports are composed of objective models with maximum pair-wise distance (diameter) ε.

$$R_\varepsilon := \bigtimes_{i \in I} \left\{ \mu^i \in \Delta^2(\Theta_0 \times \Theta_{-i} \times S_{-i}) : \forall \theta_i, \forall p, r \in \mathrm{supp}\mu^i_{\theta_i}, d_{\theta_i}(p, r) \le \varepsilon \right\}$$

The reader can verify that set R_ε is closed even when the partial identification sets are non-convex, because they are nonetheless compact-valued (BCMM). By increasing ε, we add profiles of beliefs with larger support convex hulls and which are necessarily MPSs of beliefs that are already present, provided that $\hat{\Sigma}_{-i,\theta_i}$ is always convex. As an auxiliary tool, let $R^*_\varepsilon \subseteq R_\varepsilon$ denote the subset whose beliefs already span the maximal convex hull compatible with objective information for every equilibrium. That is, R^*_ε includes each belief profile ν in R_ε such that no equilibrium $\hat{\sigma}$ exists under ν which would allow, for at least one i, θ_i and $s_i \in \text{supp}\hat{\sigma}_i$, constructing an MPS by strictly enlarging to a width larger than ε the convex hull of $\nu^i_{s_i,\theta_i}$ within the information constraint $\hat{\Sigma}_{-i,\theta_i}(s_i, q_{-i}, \sigma^*_{-i})$ consistent with equilibrium:

Note that this set is independent of specific values of γ when $\hat{\Sigma}_{-i,\theta_i}$ is convex. Finally, for any $\gamma \geq 0$ let $R^\gamma_\varepsilon := \left(R_{\varepsilon+\gamma} \backslash R_\varepsilon\right) \cup R^*_\varepsilon$. This last set serves the objective described above of varying perceived uncertainty wherever possible, without restricting the range of predictions an agent may hold. We can now state the main result of the paper: the stochastic dominance relation leads to weakly larger equilibrium sets when perceived uncertainty grows.[22]

Theorem 1. *Let (G, φ) be a finite game with feedback and ambiguity attitudes, which satisfies Observed Payoffs and in which all players are ignorant about the matching process and weakly averse to ambiguity. Then,*

$$\forall \varepsilon, \gamma \in \mathbb{R}_+, \qquad SSCE\,(G, \varphi, R_\varepsilon) \subseteq SSCE\,(G, \varphi, R^\gamma_\varepsilon) \subseteq SSCE\,(G, \varphi)$$

3.4. **Own-strategy independence of feedback**

An important, nontrivial assumption often considered when studying self-confirming equilibria is that of Own-Strategy Independence (cfr. Section 2.1 for notation):

Definition 6. *Feedback function F_i satisfies Own-Strategy Independence of Feedback about others (OSI) if, for every θ_i,*[23]

$$\forall s'_i, s''_i \in S_i, \mathcal{F}_{s'_i,\theta_i} = \mathcal{F}_{s''_i,\theta_i} \tag{14}$$

Game with feedback (G, F) satisfies Own-Strategy Independence if condition (14) holds for each player $i \in I$.

In words, i's feedback function satisfies this property if the partition of the uncertainty space it induces is independent of which strategy i actually chooses, for every type of player in that population. The learning patterns do not differ depending on one's own strategy, for one's given own type. In other words, changing from one strategy to another in between plays of the game reveals nothing more and nothing less about the behavior and types of coplayers. If payoffs are not observable, this property raises the incentives for ambiguity averse players to choose ex-ante less ambiguous options. It is natural to ask, then, what happens when it is combined with Observed Payoffs.

Lemma 4. *If (G, φ) satisfies Definitions 1 and 6, then for every $\varepsilon \in \overline{\mathbb{R}}_+$,*

$$BNE = SCE = SSCE(G, \varphi, R_\varepsilon) = MSCE$$

where BNE denotes the set of Bayes-Nash Equilibria.

As in BCMM, if OSI holds together with Observed Payoffs, in the long-run every player learns the expected payoff profile associated with every strategy, not only the one they have been playing. That is, set

$$\left\{U_i(\tilde{s}_i, \theta_i, \hat{\pi}) : \hat{\pi} \in \hat{\Sigma}_{-i,\theta_i}(s^*_i, q_{-i}, (\sigma^*_{\theta_k})_{\theta_k \in \Theta_k, k \neq i})\right\}$$

[22] One can easily show that Theorem 1 holds with equality if players are not averse to ambiguity.

[23] A less stringent formulation requires condition 14 to hold only for non-dominated strategies of each type θ_i in each role i, since rational players never pick dominated strategies irrespective of their beliefs and information.

is a singleton for every $\tilde{s}_i$, because all strategies induce the same partition of the uncertainty space. As a result, the role of prior beliefs becomes immaterial. The set of (S)SCE collapses to that of BNE. The following example shows this point, i.e. that there are no implications to alternative prior specifications in this class of games.

Example 2. Figure 3 represents the strategic form of a game with feedback and ambiguity attitudes. Consider only pure strategy profiles for simplicity. The unique pure BNE is $(\mathbf{M}, \ell)$. Let again $F_i = U_i$; the game thus satisfies both Observed Payoffs and OSI, as can be checked by the payoff structure. Given that Colin's feedback is perfect and independent of his strategy, $\mu_c = \delta_{s_{\mathbf{R}}}$ is necessary: no uncertainty is involved for him. Whenever Colin (c) is playing ℓ or m, Rowena ($\mathbf{R}$) may be uncertain of which; any belief encompassing such strategies yields the same (feedback and) payoff for $\mathbf{M}$, that is 3. Irrespectively of her beliefs, however, she knows that as long as c plays ℓ or m, her expected payoff profile is constant for every action she may play. In other words, whatever strategy she decides to play in the long-run, the expected payoff of alternative strategies involves no more uncertainty. Thus the support of her belief cannot induce any status quo in favor of any strategy that has been played for a long time. ◁

FIGURE 3: A game where stochastic dominance of priors does not matter.

$\mathbf{R}$ \ c	ℓ	m	r
U	$\mathbf{2}, \mathit{0}$	$\mathbf{2}, \mathit{1}$	$\mathbf{1}, \mathit{0}$
M	$\mathbf{3}, \mathit{1}$	$\mathbf{3}, \mathit{0}$	$\mathbf{0}, \mathit{1}$
D	$\mathbf{0}, -\mathit{1}$	$\mathbf{0}, -\mathit{1}$	$-\mathbf{1}, \mathit{2}$

3.5. **The role of payoff uncertainty**

In the examples made so far, payoff uncertainty has not been given a pivotal role, mostly for ease of exposition. In general, however, adding uncertain parameters plays an important role in SCE analysis. For any given feedback structure, the fewer elements known by the players, the more distributions are consistent with the evidence. Thus, uncertainty cannot but grow (though possibly weakly). As the discussion shall have made clear by now, this inherently reinforces the status quo bias implied by ambiguity aversion when players are sufficiently aware of it. This is shown in the following example.

Example 3. In the game of Figure 4, let $F_i = U_i$ for $i = 1, 2$, and $\theta = 0.1$. Suppose first that θ is private knowledge in possession of player 2. The game then satisfies Observed Payoffs and Own-Strategy Independence, thus has a unique smooth SCE, corresponding to the unique pure Bayesian Equilibrium, (T, ℓ). It is not possible to sustain (B, r) as a SCE since Player 2 will always want to deviate, knowing he can gain for certain.

FIGURE 4: A simple game with payoff uncertainty.

1 \ 2	ℓ	r
T	$1, -1$	$-1, -2$
B	$-1, \theta$	$0, 0$

Suppose now instead that θ is private knowledge possessed by player 1, and player 2 is only told (reliably) that $\theta \in [-0.5, 1]$. The game does not satisfy Own-Strategy Independence anymore. With a uniform conjecture over all available hypotheses, expected utility is 0.25, but if $\varphi_2(u) = -\frac{1}{\alpha}e^{-\alpha u}$ with $\alpha > 3.1$, it can be verified that the private value of the deviation to Player 2 is negative. Hence (B, r) is a SSCE of this game, given this ambiguity attitude. Interestingly, note that the status quo bias here ends up favoring Player 2, since this equilibrium is subjectively better than (T, ℓ), and profile (B, ℓ) cannot be sustained. Sometimes, certainty traps may turn out to be beneficial to some parties involved. $\triangleleft$

Many games of interest with no payoff uncertainty can be put into a formal equivalence with games with payoff uncertainty, so the latter does not always add significant insights into the analysis. The literature of choice under uncertainty has often drawn connections with games played against a malevolent nature. Indeed, games against nature can easily be reduced to one-player, payoff uncertain games.

Example 4. Go back to the game of Example 1. The same game against Nature can be thought of as the one-player game where the utility profiles of the subgame MP are determined by the residual uncertainty parameter $\theta_0 \in \{\theta_0', \theta_0''\}$, and $U_{\mathbf{A}}(In.\mathbf{H}, \theta_0') = U_{\mathbf{A}}(In.\mathbf{T}, \theta_0'') = 2, U_{\mathbf{A}}(In.\mathbf{H}, \theta_0'') = U_{\mathbf{A}}(In.\mathbf{T}, \theta_0') = 1$. Adapting the reasoning on priors to this payoff uncertain setting does not alter the results, as long as feedback is such that realized utility is observed, but an off-path realization of θ_0 is not. $\triangleleft$

It must be kept in mind, however, that while in complex games substituting a payoff uncertain representation with one that only features moves of chance and no payoff uncertainty may simplify the initial mathematical structure, it would also require more intricate descriptions of players' knowledge of the rules of the game and feedback. In Example 3, if θ were determined through a fictitious move of chance $q_0(\theta = 0.1) = 1$, in order to achieve a formal equivalence of the analysis with the initial interpretation as private knowledge of Player 2, it would be necessary to introduce a multistage structure, and it would not make sense to analyze the strategic form of the game. This is because Player 2 should always have knowledge of the actual value of the parameter before choosing her strategy.

4. An application

In this section, I present a simple application of the above framework to a contracting game, highlighting how certainty crises and non-Bayesian decision making may lead to inefficient outcomes. The game purposefully exhibits strong links with models of Adverse Selection and Moral Hazard (Akerlof, 1995; Mirrlees, 1999). Traditional versions of these theories typically rely on Bayes-Nash Equilibrium analysis. This application instead serves to appreciate the implications of Self-Confirming Equilibrium and ambiguity aversion for negotiation. Strategic interaction in the context of surplus sharing is also the fundamental feature of bargaining problems. A survey of models and equilibrium analysis (particularly, subgame-perfect equilibrium) for bargaining problems is provided by Osborne and Rubinstein (1991).

Section 4.1 describes the context of the game. Variables are introduced, together with specific numerical values, which are then used to clarify key insights. To this end, an example is conducted in Section 4.2, together with a more general comparative statics exercise, where Theorem 1 is applied to identify bounds on perception of ambiguity which may prevent certainty traps, as functions of the original variables.

4.1. Contract game description

The setting echoes the illustration of Section 3.2. A *thick* job market is composed of two large populations: Workers $\mathcal{W}$ and Firms $\mathfrak{F}$. At every repetition of the game, some $W \in \mathcal{W}$ and $F \in \mathfrak{F}$

are matched at random. W moves first, and must decide whether to apply for a job (J) at F at cost $c = 1$,[24] or not (NJ). If he applies, then he must choose whether to exert effort (E) at private cost $e = 3$, or shirk (S).

Simultaneously[25], the firm F decides whether to Collaborate (C) with the worker, which involves upfront cost $m = 4$, or Not (NC), which is costless. If F does Not collaborate, worker productivity is stale, and independent of effort. In this case, the worker receives a fixed payment schedule, making shirking optimal. Collaboration instead provides advanced technology, that crucially becomes productive *only if the worker has previously exerted effort.* Delivery of this technology occurs near the end of the contract, ensuring that W must commit to his effort level beforehand, without knowing F's choice. The payment structure in case of Collaboration is designed to incentivize effort: shirkers (S) receive no payment, while diligent workers (E) are compensated with a bonus above their effort cost.

Let R, P denote the Revenue and Payment (wage) functions, respectively. For notational convenience, also denote the share of revenues appropriated by the Firm by $\Pi_s := R_s - P_s$, where s denotes the (reduced) strategy profile. For the numerical example, assume the following specifications:

$$R(s) := \begin{cases} 12 & s = (J.E, C) \\ R_0 & s = (NJ, \cdot) \\ 3 & \text{otherwise} \end{cases} \qquad\qquad P(s) := \begin{cases} 6 & s = (J.E, C) \\ 3 & s_F = NC \\ 0 & \text{otherwise} \end{cases}$$

Let feedback be given by[26] $\overline{F}_i(s) = (R_s, P_s)$ for $i \in \{W, F\}$. One can quickly verify that this game satisfies Observed Payoffs in the sense of Definition 1, while it does not satisfy Definition 6. This is sensible as firms observe revenues, reward workers accordingly, and workers observe their salary and effort. As stressed above, $m > \Pi_{S,C}$, that is, if F recognizes a shirker then it does not reward him, but nonetheless incurs a loss. Aside from the necessary condition of Firm Cooperation to achieve higher profits, other assumptions on payoffs are consistent with typical models of Moral hazard. The firm is concerned with the level of effort on the part of the worker, and without monitoring technologies it cannot ex-ante ensure itself against shirking.

Let us introduce an additional strategy for the Firm: Cooperation with Insurance (I). If F wishes to cooperate, but is sufficiently worried that the worker might shirk, it can stipulate a parallel insurance contract[27] at premium $p = 3$, with potential payout $b = 6 > p$ in case of low revenue. By choosing I, worker compensation is the same as in the case of cooperation, while Firm profits change. If W exerts effort, the firm bears the insurance cost and incurs a loss equal to $\Pi_{E,C} - m - p = -1$. On the other hand, in case of shirking the insurance payout offsets losses and the Firm obtains $\Pi_{S,C} - p + b = 2$. In principle, insurance allows Firms to always achieve greater expected profits than by choosing NC. The tree of the game is shown in Figure 5, while Figure 6 depicts the strategic form, where numerical values for all relevant variables have been substituted in for clearer interpretation of the example.

The key strategic challenge is mutual commitment. Neither party can observe the other's choice in advance. The firm only wants to collaborate when confident of worker effort, while the worker would always exert effort if certain of collaboration. Post-contract, the firm can verify technology usage, creating a monitoring mechanism that aligns incentives. If the worker shirks, the advanced technology becomes worthless, and the firm incurs additional losses—even without

[24] The application cost c can be alternatively interpreted as reservation utility of the worker.

[25] Simultaneity is in fact non-essential and shall here be interpreted as a technical property of the rules of the game; it is key, however, that W's choice cannot be observed by F.

[26] I use $\overline{F}$ to denote feedback in this game in order to distinguish it from player index F used for Firms.

[27] This assumption allows flexible interpretation: in reality, firms may devise a number of financial and nonfinancial strategies to insure themselves against the possibility of hiring a bad employee. One simple strategy we can imagine is conducting parallel hiring processes at a cost to be able to replace the departing employee more quickly.

FIGURE 5: Contract game tree.

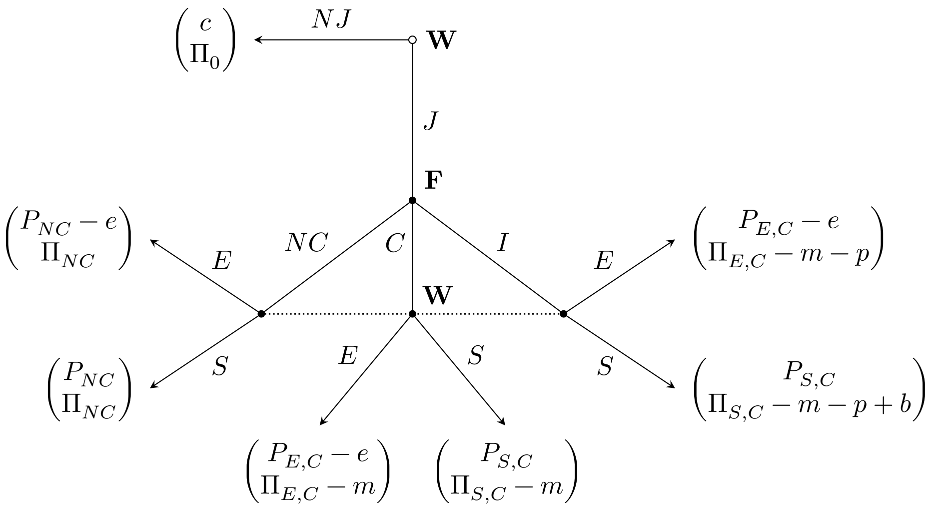


FIGURE 6: Strategic form of the Contract game.

W \ F	NC	C	I
NJ	$1, \Pi_0$	$1, \Pi_0$	$1, \Pi_0$
$J.E$	$0, 0$	$3, 2$	$3, -1$
$J.S$	$3, 0$	$0, -1$	$0, 2$

paying the worker. After the negotiation, the contract is carried out, W and F are separated, and the procedure is repeated.

The game is designed to capture *relational contracting*, a fundamental coordination challenge in labor markets.[28] Both firms and workers face incentives to make relation-specific investments only when they can credibly anticipate reciprocal commitment from their counterparty. The assumption that neither party can privately influence the market outcome adds a further layer to this bilateral hold-up problem, which characterizes many medium-term employment relationships.

4.2. **Comparative statics and market shutdown**

To fix ideas, consider the numerical example, whose strategic-form representation can be seen in Figure 6, and focus on SEU-maximizing players first. As in Section 3.2, a Bayesian worker W would never choose NJ, since for any $\sigma_F \in \Delta(\{NC, C, I\})$:

$$\max_{s_W \in \{J.E, J.S\}} V_W(s_W, \sigma_F) = \max\{3\sigma_F(NC), 3(1-\sigma_F(NC))\} \geq 1.5 > 1 = V_W(NJ, \sigma_F)$$

Showing that a Bayesian F has the possibility of insuring itself against shirkers and therefore it would never pick NC, regardless of its belief, is analogous. As a result workers, who do not pick NJ, perfectly observe that firms are always choosing either C or I. As the game repeats, by observing their payoffs they quickly learn that NC is not being chosen and thus it is optimal for them to exert effort. In turn, and also by observing their payoffs, firms can indirectly observe $J.E$, and the only rational reply is to choose C. Again, the same reasoning applies to firms.

[28]The issue of relational contracts and relation-specific investments has been widely discussed. For more articulate theoretical insights, see Crawford (1990) and Baker *et al.* (2002), and the references therein.

Whatever their conjecture, they always choose C or I. Although they initially may choose to stipulate the insurance, in the long run they would observe that all workers are exerting effort, and begin collaborating. Simply put, the game necessarily converges toward the efficient coordination outcome for both parties, i.e. $(J.E, C)$. This is the unique Bayes-Nash equilibrium of the game, and the unique SCE for Bayesian players. Note that, although insurance is never picked, its presence is key. Without it, equilibrium sets would be quite different: let (G, F) denote the game with feedback described in Section 4.1, and let $\overline{G}$ denote the modification of G where strategy I is removed from the choice set of players $F \in \mathfrak{F}$. Then

$$\begin{aligned}
BNE(G) = SCE(G) &= \{(J.E, C)\} \\
&\subseteq BNE(\overline{G}) = \left\{(J.S, NC), (J.E, C), \left(\frac{1}{3}\delta_{J.E} + \frac{2}{3}\delta_{J.S}, \frac{1}{2}\delta_{NC} + \frac{1}{2}\delta_C\right)\right\} \\
&\subseteq SCE(\overline{G}, F) = BNE(\overline{G}) \cup \{(J.E, \alpha\delta_{NC} + (1-\alpha)\delta_C) : \alpha < 0.5\} \\
&\qquad \cup \left\{\left(\beta\delta_{J.E} + (1-\beta)\delta_{J.S}, \frac{1}{2}\delta_{NC} + \frac{1}{2}\delta_C\right) : \beta > \frac{1}{3}\right\}
\end{aligned}$$

Now turn to ambiguity averse players. The structure of payoffs is such that uncertainty averse agents may pick NJ: e.g. for $\varphi_W = U_W^{1/\alpha}$ and $\alpha > \log_2 3$, NJ is justifiable as a best reply to $\mu_W = \frac{1}{2}\delta_{NC} + \frac{1}{2}\delta_C$. The same applies to firms: if they are sufficiently ambiguity averse and aware, they may choose NC, thereby cancelling the efficacy of the presence of insurance. The set of $SSCE$ is thus much larger, including mixed profiles which assign positive probabilities to subjectively optimal, yet objectively suboptimal strategies. For instance, sets $\{NJ\} \times \Delta(S_F)$ and $\Delta(S_W \backslash \{J.E\}) \times \{NC\}$ are both subsets of $SSCE(G, \omega)$. Simply introducing ambiguity aversion into to the model can easily make insurance useless and lead to market shutdowns, in the form of chronic worker shortages or pathological underinvestment. Insurance is unhelpful in this case because ambiguity aversion will either lead workers not to seek employment at all, or firms to underestimate the objective support offered by the decoy strategy to reach the coordination equilibrium.

It is also possible to make Theorem 1 operational, and obtain conditions for the occurrence of certainty traps, i.e., equilibria that are only possible when players are sufficiently ambiguity averse. We will conduct this exercise from the perspective of a worker; the reasoning is exactly the same for firms. Recall that the principle of reduction of compound lotteries applies to risky decisions. To avoid market shutdown it is necessary that

$$\begin{aligned}
c < V_{\mu_W}(J) &= \max\{V_{\mu_W}(J.E), V_{\mu_W}(J.S)\} \\
&= \max\left\{\varphi_W^{-1}\left(\sum_{\hat{\pi}\in\operatorname{supp}\mu_W} \varphi_W(P_{E,C} + \hat{\pi}(NC)(P_{NC} - P_{E,C}) - e)\mu_W(\hat{\pi})\right),\right. \\
&\qquad \left.\varphi_W^{-1}\left(\sum_{\hat{\pi}\in\operatorname{supp}\mu_W} \varphi_W\left(P_{S,C} + \hat{\pi}(NC)(P_{NC} - P_{S,C})\right)\mu_W(\hat{\pi})\right)\right\}
\end{aligned}$$

This condition simply requires that the entry cost c that would be saved by choosing NJ be smaller than the certainty equivalent of the subjectively optimal choice. If we consider the maximal degree of uncertainty perception, i.e. a second-order belief evenly split among the most extreme outcomes compatible with objective evidence, bounds on ambiguity aversion can again be found analogously to Section 3.2:

$$c < \varphi_W^{-1}\left(\frac{1}{2}\max\left\{\varphi_W(P_{NC} - e) + \varphi_W(P_{E,C} - e), \varphi_W(P_{NC}) + \varphi_W(P_{S,C})\right\}\right) \tag{15}$$

Explicit conditions on parameters can be derived through (15) by considering specific functional forms for φ_W. Conversely, we can consider a given degree of ambiguity aversion and calculate bounds on perception to avoid certainty traps. For instance, if W is extremely averse to ambiguity, we have the following sufficient condition:

$$c < \max\left\{\min_{\hat{\pi}\in\mathrm{supp}\mu_W} P_{S,C} + \hat{\pi}(NC)(P_{NC} - P_{S,C}), \min_{\hat{\pi}\in\mathrm{supp}\mu_W} P_{E,C} + \hat{\pi}(NC)(P_{NC} - P_{E,C}) - e\right\}$$

which reduces to:[29]

$$\min_{\hat{\pi}\in\mathrm{supp}\mu_W} \hat{\pi}(NC) > \frac{c - P_{S,C}}{P_{NC} - P_{S,C}} \qquad \vee \qquad \max_{\hat{\pi}\in\mathrm{supp}\mu_W} \hat{\pi}(NC) < \frac{P_{E,C} - c - e}{P_{E,C} - P_{NC}} \tag{16}$$

In the more general case of a given profile of smooth functionals φ, we can again consider the maximally uncertain belief within some given interval of utility values compatible with the information constraint. Then, we can find a sufficient condition on the maximal distance of models in the support to prevent certainty traps. Let

$$\overline{\pi} := \max_{\hat{\pi}\in\mathrm{supp}\mu_W} 1 - \hat{\pi}(NC) \qquad\qquad \underline{\pi} := \max_{\hat{\pi}\in\mathrm{supp}\mu_W} \hat{\pi}(NC)$$

denote the objective models which make subjectively most appealing choosing $J.E$ and $J.S$, respectively. Then $|\overline{\pi} - \underline{\pi}|$ must satisfy:

$$\begin{aligned} 2\varphi_W(c) < \max\{&\varphi_W\left(P_{E,C} - e + \overline{\pi}(P_{NC} - P_{E,C})\right) + \varphi_W\left(P_{E,C} - e + \underline{\pi}(P_{NC} - P_{E,C})\right) \\ &\varphi_W\left(P_{S,C} + \overline{\pi}(P_{NC} - P_{S,C})\right) + \varphi_W\left(P_{S,C} + \underline{\pi}(P_{NC} - P_{S,C})\right)\} \end{aligned}$$

Also in this case, an explicit solution for this sufficient condition can be derived only when a specific form for φ_W is considered.

5. CONCLUSIONS AND LIMITATIONS

An extensive general framework for Self-Confirming Equilibrium analysis of the strategic form of games with feedback, ambiguity attitudes, and payoff uncertainty has been provided. The results of BCMM are shown to generalize seamlessly to the case of payoff uncertainty. Most importantly, I have given a rigorous account of the game-theoretic interplay between uncertainty aversion and perception, clarifying that even in presence of mild aversion to ambiguity, agents who are more aware of the uncertainty present in their decision environment can get stuck in certainty traps.

The intuition for this fact is a direct consequence of the representation of KMM, who characterized ambiguity aversion as dread toward mean-preserving spreads of second-order beliefs in the space of expected utility values. A game-theoretic framework proves particularly illuminating in demonstrating how this feature can lead to objectively inferior decisions. Section 3 reveals the conditions required for this result are remarkably modest. It is sufficient to pose mild restrictions on the width of supports of beliefs to control flexibly the perception of ambiguity, while preserving the game's fundamental structure.

The application of Section 4 makes these theoretical results operational, demonstrating their relevance for microeconomic theory and particularly for models of adverse selection and relational contracting. Ambiguity aversion can fundamentally alter strategic interactions in employment relationships, potentially explaining observed patterns of inefficient contracting and underinvestment in human capital. This framework provides new insights into why seemingly beneficial long-term relationships might fail to materialize, even when they would objectively improve upon current arrangements. These results offer practical insights for organizational design and policy-making. They suggest that mechanisms for reducing strategic uncertainty, including reputation systems,

[29] Existence is taken for granted provided that $c + e \leq P_{E,C}$ is necessary for participation.

standardized practices, or industry norms, might be more valuable than previously recognized, as they could help overcome the coordination failures identified in the analysis.

The framework of Battigalli et al., 2019b would be most germane to the exploration of consequences on steady-state convergence of early conjectural revisions occurring in learning environments. Additionally, a more robust analysis would refer to extensive-form games where players can change continuation strategy at any non-terminal history (Battigalli et al., 2019a), in light of the implications of dynamic inconsistency pointed out in Section 2. In both cases, however, the complexity of mathematical objects involved grows noticeably. An exploration of comparative ambiguity perception in these settings is open to further research. In addition, a deeper understanding of the normative foundations of ambiguity perception would strengthen the economic significance of the analysis of games with restrictions on beliefs conducted in Section 3.

Nevertheless, core insights remains robust: heightened ambiguity aversion generates a status quo bias, whose likelihood increases with the perception of such ambiguity. Certainty crises, or deeply embedded uncertainty-averse cultures, can trap economic agents in inefficient equilibria. While multiple-prior models primarily address subjective rather than objective rationality, the existence of an objective probability distribution—if such exists—must be unique. Thus, although uncertainty aversion may be consistent with subjective rationality, and uncertainty averse decision-makers may never be convinced their choices are suboptimal, this trait can inadvertently become self-defeating.

REFERENCES

Akerlof, G. A. (1995). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Akerlof, G. A. and R. J. Shiller (2009). *Animal Spirits. How human psychology drives the economy, and why it matters for global capitalism.* Princeton University Press.

Almelhem, A., M. Iyigun, A. Kennedy, and J. Rubin (2023). Enlightenment ideals and belief in progress in the run-up to the industrial revolution: A textual analysis. IZA Discussion Paper No. 16674, IZA Institute of Labor Economics.

Anscombe, F. J. and R. J. Aumann (1963). A definition of subjective probability. *Annals of Mathematics and Statistics 34*, 199–205.

Battigalli, P., E. Catonini, G. Lanzani, and M. Marinacci (2019a). Ambiguity attitudes and self-confirming equilibrium in sequential games. *Games and Economic Behavior 115*, 1–29.

Battigalli, P., E. Catonini, and N. D. Vito (2023). Game theory. an analysis of strategic thinking. Typescript, Bocconi University.

Battigalli, P., S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci (2015). Self-confirming equilibrium and model uncertainty. *American Economic Review 105*(2), 646–677.

Battigalli, P., S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci (2016). Analysis of information feedback and selfconfirming equilibrium. *Journal of Mathematical Economics 66*, 40–51.

Battigalli, P., A. Francetich, G. Lanzani, and M. Marinacci (2019b). Learning and self-confirming long-run biases. *Journal of Economic Theory 183*, 740–785.

Battigalli, P., M. Gilli, and M. C. Molinari (1992). Learning and convergence to equilibrium in repeated strategic interactions: an introductory survey. *Research in Economics 46*(3), 335–378.

Battigalli, P. and D. Guaitoli (1988). Conjectural equilibria and rationalizability in a game with incomplete information. In P. Battigalli, A. Montesano and F. Panunzi (Eds.), Decisions, Games and Markets, Springer, Boston, MA, pp. 97–124.

Bernheim, B. D. (1984). Rationalizable strategic behavior. *Econometrica 52*, 1007–1028.

Bewley, T. (2002). Knightian decision theory: Part i. *Decisions in Economics and Finance 25*, 79–110.

Burn-Murdoch, J. (2024). Is the west talking itself into decline? *Financial Times.* January 5.

Choquet, G. (1953). Theory of capacities. *Annales de l'Institut Fourier 5*, 131–295.

Ellsberg, D. (1961). Risk, ambiguity and the savage axioms. *Quarterly Journal of Economics 75*, 643–69.

Fishburn, P. C. (1991). Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty 4*(2), 113–34.

Fudenberg, D. and D. K. Levine (1993). Self-confirming equilibrium. *Econometrica 61*(3), 523–545.

Fudenberg, D. and D. K. Levine (1998). *The Theory of Learning in Games.* Cambridge, MA: MIT Press.

Gajdos, T., T. Hayashi, J.-M. Tallon, and J.-C. Vergnaud (2008). Attitude toward imprecise information. *Journal of Economic Theory 140*, 27–65.

Gilboa, I., F. Maccheroni, M. Marinacci, and D. Schmeidler (2010). Objective and subjective rationality in a multiple prior model. *Econometrica 78*(2), 755–770.

Gilboa, I. and M. Marinacci (2013). Ambiguity and the bayesian paradigm. In D. Acemoglu, M. Arellano and E. Dekel (Eds.), Advances in Economics and Econometrics: Tenth World Congress, Cambridge University Press, pp. 179–242.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*(2), 141–153.

Gustafsson, J. E. (2022). *Money-Pump Arguments.* Cambridge: Cambridge University Press.

Harsanyi, J. C. (1967). Games with incomplete information played by "bayesian" players, i–iii part i. the basic model. *Management Science 14*(3), 159–182.

Harsanyi, J. C. (1968a). Games with incomplete information played by "bayesian" players, i–iii part ii. bayesian equilibrium points. *Management Science 14*(5), 320–334.

Harsanyi, J. C. (1968b). Games with incomplete information played by "bayesian" players, i–iii part iii. the basic probability distribution of the game. *Management Science 14*(7), 486–502.

Iyengar, S. S. and M. R. Lepper (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology 79*(6), 995–1006.

Keynes, J. M. (1936). *The General Theory of Employment, Interest, and Money.* Macmillan.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005). A smooth model of decision making under ambiguity. *Econometrica 73*(6), 1849–1892.

Knight, F. H. (1921). *Risk, Uncertainty and Profit.* Houghton Mifflin.

Mirrlees, J. A. (1999). The theory of moral hazard and unobservable behaviour: Part i. *The Review of Economic Studies 66*(1), 3–21.

Mokyr, J. (2017). *A Culture of Growth: The Origins of the Modern Economy.* Princeton University Press.

Nash, J. (1950). Non-cooperative games. *Annals of Mathematics 54*(2), 286–295.

Osborne, M. J. and A. Rubinstein (1991). *Bargaining and Markets.* Academic Press.

Pearce, D. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica 52*, 1029–1050.

Rothschild, M. and J. E. Stiglitz (1970). Increasing risk: I. a definition. *Journal of Economic Theory 2*(3), 225–243.

Savage, L. J. (1954). *The Foundations of Statistics.* John Wiley and Sons.

Schmeidler, D. (1986). Integral representation without additivity. *Proceedings of the American Mathematical Society 97*, 255–261.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica 57*(3), 571–587.

Seuanez-Salgado, M. (2006). Choosing to have less choice. *FEEM Working Paper* (37).

von Neumann, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior.* Princeton University Press.

Wald, A. (1949). Statistical decision functions. *Annals of Mathematical Statistics 20*(2), 165–205.

APPENDIX

Lemma 1

Proof. Let $i \in I, s_i \in S_i, \sigma^*_{-i} \in \Delta(S_{-i})$. Payoff observability implies the section of payoff function U_{s_i,θ_i} must be $\mathcal{F}_{s_i}$-measurable: that is, any payoff value corresponds to a measurable set in partition $\mathcal{F}_{s_i,\theta_i}$. This measurability implies that if two distributions $\widehat{\pi}(\theta_0, \theta_{-i}, s_{-i})$ and $q_0(\theta_0)q_{-i}(\theta_{-i})\sigma^*_{-i}(s_{-i}|\theta_{-i})$ have the same $\mathcal{F}_{s_i}$ measure, they will induce the same conditional distribution (i.e., conditional on $\mathcal{F}_{s_i}$, 5$^{\text{th}}$ equality below). Formally, for any $\widehat{\pi} \in \widehat{\Sigma}_{-i,\theta_i}$,

$$
\begin{aligned}
U_i(s_i, \theta_i, \widehat{\pi}) &= \sum_{(\theta_0,\theta_{-i},s_{-i})\in\Theta_0\times\Theta_{-i}\times S_{-i}} U_i(s_i, \theta, s_{-i})\widehat{\pi}(\theta_0, \theta_{-i}, s_{-i}) \\
&= \int_{\Theta_0\times\Theta_{-i}\times S_{-i}} U_i(s_i, \theta, s_{-i})\ \widehat{\pi}(\mathrm{d}\theta_0, \mathrm{d}\theta_{-i}, \mathrm{d}s_{-i}) \\
&= \int_{\Theta_0\times\Theta_{-i}\times S_{-i}} U_{s_i,\theta_i}\ \widehat{\pi}(\mathrm{d}\theta_0, \mathrm{d}\theta_{-i}, \mathrm{d}s_{-i}) \\
&= \int_{\Theta_0\times\Theta_{-i}\times S_{-i}} U_{s_i,\theta_i}\ \widehat{\pi}(\mathrm{d}\theta_0, \mathrm{d}\theta_{-i}, \mathrm{d}s_{-i})|\mathcal{F}_{s_i,\theta_i} \\
&= \int_{\Theta_0\times\Theta_{-i}\times S_{-i}} U_{s_i,\theta_i,q_{-i}}\ \sigma^*_{-i}(\mathrm{d}s_{-i}|\theta_{-i})q_0(\mathrm{d}\theta_0)q_{-i}(\mathrm{d}\theta_{-i})|\mathcal{F}_{s_i,\theta_i} \\
&= \int_{\Theta_{-i}}\int_{\Theta_0}\int_{S_{-i}} U_{s_i,\theta_i,q_{-i}}\sigma^*_{-i}(\mathrm{d}s_{-i}|\theta_{-i})q_0(\mathrm{d}\theta_0)q_{-i}(\mathrm{d}\theta_{-i}) \\
&= \sum_{(\theta_0,\theta_{-i},s_{-i})\in\Theta_0\times\Theta_{-i}\times S_{-i}} U_i(s_i, \theta_i, q_{-i}, s_{-i})\sigma^*_{-i}(s_{-i}|\theta_{-i})q_0(\theta_0)q_{-i}(\theta_{-i}) \\
&= U_i(s_i, \theta_i, q_{-i}, \sigma^*_{-i})
\end{aligned}
$$

□

Lemma 2

Proof. I show how to adapt the standard equivalence result on second-order stochastic dominance to a framework with subjective beliefs, rather than lotteries. I prove the "only if" part: as the random variables are shown to satisfy SOSD for all concave functions in the sense of Rothschild & Stiglitz (1970), the reverse implication holds due to their Theorem 2. Let $s \in S$ denote a state of the world and Σ the σ-algebra defined on $\Delta(S)$. Let $U_{s_i,\theta_i} : \Delta(\Theta_0 \times \Theta_{-i} \times S_{-i}) \to \mathbb{R}$ be (the (s_i, θ_i)-section of) a Bernoulli utility function defined on the space of objective lotteries. To simplify notation, we simply write U_i in place of U_{s_i,θ_i}. Let φ_i be a strictly increasing and weakly concave function, and ν^i, μ^i be two subjective beliefs such that μ^i is an MPS of ν^i on the space of expected utility values. This immediately implies $\int_\Sigma U_i(\widehat{\pi}(s))\nu^i\,(\mathrm{d}\widehat{\pi}) = \int_\Sigma U_i(\widehat{\pi}(s))\mu^i\,(\mathrm{d}\widehat{\pi})$. An objective probability law on unknown states $\widehat{\pi}$ and the utility function jointly induce objective expected utility (a lottery) with value

$$u := U_i(\widehat{\pi}) \in \mathbb{U} := \mathrm{Im}U_i = \left[\min_{\pi\in\Sigma} U_i(\pi), \max_{\pi\in\Sigma} U_i(\pi)\right] \subseteq \mathbb{R}$$

where the minimum and maximum are well-defined by compact-valuedness of Σ. Since when payoffs are observable U_{s_i} is always $\mathcal{F}_{s_i}$-measurable, we can write — with a slight abuse of notation, i.e. identifying a belief μ on objective lotteries with its push-forward $\mu \circ U^{-1}$ on the space of expected utility values

$$\int_\Sigma \varphi\left(U(\widehat{\pi}(s))\right)\nu\,(\mathrm{d}\widehat{\pi}) = \int_\mathbb{U} \varphi\,(u)\,\nu\,(\mathrm{d}u)$$

Since μ is a mean-preserving spread of ν, there exists some random variable z such that, for $v \sim \nu$ and $u \sim \mu$, $u \overset{D}{=} v + z$ and $\mathbb{E}(z|v) = 0$ for every $v \in \mathbb{U}$. By the law of iterated expectations and the MPS relation: $\mathbb{E}_{\mu^i}(\varphi_i(u)) = \mathbb{E}_{\nu^i}(\varphi_i(u)) = \mathbb{E}_{\nu^i}(\varphi_i(v+\varepsilon)) = \mathbb{E}_{\nu^i}\left[\mathbb{E}(\varphi_i(v+\varepsilon)|v)\right]$. Therefore we may write

$$\begin{aligned} \int \varphi_i(u)\mu^i(\mathrm{d}u) &= \int \mathbb{E}\left[\varphi_i(v+\varepsilon)|v\right]\nu^i(\mathrm{d}v) \\ &\le \int \varphi_i\left(\mathbb{E}\left[v+\varepsilon|v\right]\right)\nu^i(\mathrm{d}v) \\ &= \int \varphi_i(\underbrace{\mathbb{E}\left[v|v\right]}_{=v} + \underbrace{\mathbb{E}\left[\varepsilon|v\right]}_{=0})\nu^i(\mathrm{d}v) \\ &= \int \varphi_i(v)\nu^i(\mathrm{d}v) \end{aligned}$$

The inequality is an application of Jensen's inequality valid for any weakly concave function, and we can separate the sum in the expected value by linearity of the operator. Hence, ν^i second-order stochastically dominates μ^i under φ_i. $\square$

Lemma 3

Proof. I prove that μ is an MPS of Σ-measurable function (probability distribution) ν on the space of expected utility values if

1. $\int_\Sigma \hat{\pi}(s)\nu\left(\mathrm{d}\hat{\pi}\right) = \int_\Sigma \hat{\pi}(s)\mu\left(\mathrm{d}\hat{\pi}\right)$
2. conv supp$\nu \subset$ conv suppμ

Step 1: Condition (1) is equivalent to $\int_\Sigma U(\hat{\pi}(s))\nu\left(\mathrm{d}\hat{\pi}\right) = \int_\Sigma U(\hat{\pi}(s))\mu\left(\mathrm{d}\hat{\pi}\right)$. The two measures are necessarily absolutely continuous (with respect to the Lebesgue measure), so that we can invoke the Radon-Nykodym theorem and, together with the results of Lemma 2 restate condition 1 as

$$\int_{\mathbb{U}} u\ \nu\left(u\right)\mathrm{d}u = \int_{\mathbb{U}} u\ \mu\left(u\right)\mathrm{d}u$$

That is, $\mathbb{E}_\nu(u) = \mathbb{E}_\mu(u)$.

Step 2: Let $g = \mu - \nu$ denote the function which describes the probability spread between the two distributions , with $G(u) = \int_{-\infty}^{u} g(t)\mathrm{d}t$. We need to prove that $T(y) := \int_{-\infty}^{y} G(u)\mathrm{d}u \ge 0\ \forall y \in \mathbb{R}$ with $T(\infty) = 0$. By Lemma 1* in Deb & Seo (2011), Lemma 2 and Theorem 2 in Rothschild & Stiglitz (1970), this proves that μ is a MPS of ν.

Let $\underline{u} :=$ inf suppμ and $\overline{u} :=$ sup suppμ, $\underline{v} :=$ inf suppν and $\overline{v} :=$ sup suppν. Condition (2) guarantees $\underline{u} < \underline{v}$ or $\overline{u} > \overline{v}$. Consider the first case; the other is symmetric. Clearly, then, $\int_{-\infty}^{\underline{u}} \mu(t)\mathrm{d}t = \int_{-\infty}^{\underline{u}} \nu(t)\mathrm{d}t = 0$ and $\int_{-\infty}^{\overline{u}} \mu(t)\mathrm{d}t = \int_{-\infty}^{\overline{u}} \nu(t)\mathrm{d}t = 1$, so that $g(u) = G(u) = 0\ \forall u \notin [\underline{u}, \overline{u}]$. This automatically grants $T(\infty) = yG(y)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} ug(u)\mathrm{d}u = 0$ (since in this case 0 is rapidly reached within g on a bounded subset of the real line, the limit is not indeterminate). Additionally, $u \in [\underline{u}, \underline{v}) \Rightarrow g(u) > 0 \Rightarrow G(u) > 0$.

There is a unique point $c \ge \underline{v}$ such that $G(u) \ge 0\ \forall u \le c$ and $G(u) \le 0\ \forall u > c$. Assume by way of contradiction this were not the case, i.e. G changes sign more than once. Then there exists (at least one) point $c' > c$ such that

- $G(u) \ge 0$ if $u \le c$;
- $G(u) \le 0$ if $u \in (c, c')$;
- $G(c') > 0$.

But this cannot be: either (i) $G(u) \geq 0$ for every $u > c'$, in which case μ conserves more probability density until the end of the distribution, contradicting $T(\infty) = 0$, or (ii) the latter is satisfied, which requires $G(c'') < 0$ for some other $c'' \in (c', \overline{v}]$, in which case the mean-preservation assumption is necessarily contradicted, as probability mass or density has been moved towards the lower end of the distribution on net. In accordance with this, it follows that $T(y) \geq 0 \ \forall y \leq c$. For $y > c$, note that

$$T(y) = \underbrace{T(c)}_{\geq 0} + \int_c^y G(u)\mathrm{d}u$$

Since $\forall u > c, G(u) \leq 0$, $T(y)$ must be nonincreasing for all $y > c$. Since also $T(y) = 0 \ \forall y \geq \overline{u}$, it follows that $T(y) \geq 0 \ \forall y$, which completes the proof. □

Proposition 1

Proof. If $SSCE(BG, \varphi, \nu) = \emptyset$, the statement vacuously holds. Assume $SSCE(BG, \varphi, \nu) \neq \emptyset$. Let $\sigma^* \in SSCE(BG, \varphi, \nu)$ and consider some θ_i and $s_i^* \in \mathrm{supp}\sigma^*_{\theta_i}$. The game is assumed to satisfy Def. 1; by Lemma 1,

$$V^{\varphi_i}_{\theta_i}(s_i^*, \nu^i_{s_i,\theta_i}) = V^{\varphi_i}_{\theta_i}(s_i^*, \mu^i_{s_i,\theta_i}) = U_i(s_i^*, \theta_i, \hat{\pi}) = U_i(s_i^*, \theta_i, q_{-i}, \sigma^*_{-i}) \ \forall \hat{\pi} \in \hat{\Sigma}_{-i,\theta_i}(s_i^*, q_{-i}, \sigma^*_{-i})$$

By Lemma 2, if φ_i is weakly concave and $\mu^i_{s_i,\theta_i}$ is a mean-preserving spread of $\nu^i_{s_i,\theta_i}$, for every $s_i \in S_i$,

$$\begin{aligned} V^{\varphi_i}_{\theta_i}(s_i^*, \mu^i_{s_i,\theta_i}) &= U_i(s_i^*, \theta_i, q_{-i}, \sigma^*_{-i}) \\ &= V^{\varphi_i}_{\theta_i}(s_i^*, \nu^i_{s_i,\theta_i}) \geq V^{\varphi_i}_{\theta_i}(s_i, \nu^i_{s_i,\theta_i}) \\ &= \int \varphi_i\left(U_i(s_i, \theta_i, \hat{\pi})\right) \nu^i_{s_i,\theta_i}(\mathrm{d}\hat{\pi}) \\ &\geq \int \varphi_i\left(U_i(s_i, \theta_i, \hat{\pi})\right) \mu^i_{s_i,\theta_i}(\mathrm{d}\hat{\pi}) = V^{\varphi_i}_{\theta_i}(s_i, \mu^i_{s_i,\theta_i}) \end{aligned}$$

Since the support of $\mu^i_{s_i,\theta_i}$ is contained in $\hat{\Sigma}_{-i,\theta_i}(s_i^*, q_{-i}, \sigma^*_{-i})$ by assumption, the statement holds. □

Theorem 1

Proof. First note that the second inclusion always holds by construction and, if $\varepsilon = 0$, then $SSCE(BG, \varphi, \varepsilon) = SCE(BG, F)$. If $\gamma = 0$, the whole statement vacuously holds. Consider then the generic case $\varepsilon, \gamma > 0$, and a profile of beliefs $\mu \in R_\varepsilon$. Either $\mu \in R^*_\varepsilon$ or $\mu \in R_\varepsilon \backslash R^*_\varepsilon$. In the first case, $\mu \in R^\gamma_\varepsilon$, hence the equilibrium set for this belief is preserved. Assume it is the nontrivial case (otherwise, the belief is irrelevant) where $SSCE(BG, \varphi, \mu) \neq \emptyset$. As mentioned in the text, ignorance of the matching process ensures convexity of $\hat{\Sigma}_{-i,\theta_i}$ for any strategy profile. Let $\mu \in R_\varepsilon \backslash R^*_\varepsilon$. Then by definition there exist $i \in I, \theta_i \in \Theta_i, \sigma^* \in SSCE(BG, \varphi, \mu)$ and $s_i \in \mathrm{supp}\sigma^*_i$ such that we can construct a new profile $\tilde{\mu}$ as follows.

Let $\underline{u} := \inf_{\hat{\pi} \in \mathrm{supp}\mu^i_{s_i,\theta_i}} U_{s_i,\theta_i}(\hat{\pi})$ and $\overline{u} := \sup_{\hat{\pi} \in \mathrm{supp}\mu^i_{s_i,\theta_i}} U_{s_i,\theta_i}(\hat{\pi})$. These utility values can be attained within the support of $\mu^i_{s_i,\theta_i}$ as the game is finite, hence the suprema are maxima. Denote the expected utility profile by $\hat{u} := \int_{\underline{u}}^{\overline{u}} u \ \mu^i_{s_i,\theta_i}(\mathrm{d}u)$. Then there exists $\beta \in (\underline{u}, \overline{u})$ such that $\int_{\underline{u}}^{\beta} u \ \mu^i_{s_i,\theta_i}(\mathrm{d}u) = \int_{\beta}^{\overline{u}} u \ \mu^i_{s_i,\theta_i}(\mathrm{d}u) = \hat{u}/2$. Let $a = \varepsilon - (\overline{u} - \underline{u}) \in [0, \varepsilon]$, and $t \in (a, a + \gamma/2]$. Then, define

$$\tilde{\mu}^i_{s_i,\theta_i}(u) := \begin{cases} \mu^i_{s_i,\theta_i}(u+t) & u \in [\underline{u} - t, \beta - t] \\ \mu^i_{s_i,\theta_i}(u-t) & u \in [\beta + t, \overline{u} + t] \\ 0 & \text{otherwise} \end{cases}$$

For every other strategy, type and player, assume $\tilde{\mu}$ and μ coincide. Then, $\tilde{\mu} \in (R_{\varepsilon+\gamma} \backslash R_\varepsilon)$. By convexity of $\hat{\Sigma}_{-i,\theta_i}$, and again by definition of $R^*_\varepsilon \not\ni \mu$, an appropriate choice of i, θ_i and s_i ensures that for $t \to a^+$, $\tilde{\mu}^i_{s_i,\theta_i} \in \Delta\left(\hat{\Sigma}_{-i,\theta_i}(s_i, \sigma^*_{-i})\right)$. It is immediate to verify that $\int_{\underline{u}-t}^{\beta-t} u\ \tilde{\mu}^i_{s_i,\theta_i}(\mathrm{d}u) = \frac{\hat{u}-t}{2}$, $\int_{\beta+t}^{\overline{u}+t} u\ \tilde{\mu}^i_{s_i,\theta_i}(\mathrm{d}u) = \frac{\hat{u}+t}{2}$, so that

$$\begin{aligned}
\int_{\mathrm{supp}\tilde{\mu}^i_{s_i,\theta_i}} u\ \tilde{\mu}^i_{\theta_i}(\mathrm{d}u) &= \int_{\underline{u}-t}^{\overline{u}+t} u\ \tilde{\mu}^i_{s_i,\theta_i}(\mathrm{d}u) \\
&= \int_{\underline{u}-t}^{\beta-t} u\ \tilde{\mu}^i_{s_i,\theta_i}(\mathrm{d}u) + \underbrace{\int_{\beta-t}^{\beta+t} u\ \tilde{\mu}^i_{s_i,\theta_i}(\mathrm{d}u)}_{=0} + \int_{\beta+t}^{\overline{u}+t} u\ \tilde{\mu}^i_{s_i,\theta_i}(\mathrm{d}u) \\
&= \frac{\hat{u}}{2} - \frac{t}{2} + \frac{\hat{u}}{2} + \frac{t}{2} = \hat{u}
\end{aligned}$$

By Lemma 3 and Proposition 1, $SSCE(BG, \varphi, \mu) \subseteq SSCE(BG, \varphi, \tilde{\mu})$. Since μ was arbitrary, for any $\mu \notin R^*_\varepsilon$ we can always find some $\tilde{\mu} \in (R_{\varepsilon+\gamma} \backslash R_\varepsilon) \subseteq R^\gamma_\varepsilon$ as above. □

Lemma 4

Proof. By Proposition 1 in BCMM, we know that if a game satisfies observable payoffs then $MSCE \supseteq SSCE$, thus including those determined by any exogenous restriction on priors. Analogously to their setting, it is sufficient to show that every MSCE is a Bayesian Equilibrium to prove that also every SSCE is a Bayesian Equilibrium, regardless of the restrictions we put on beliefs as long as we allow at least singleton support priors – which is the case since $C_1 \subset R_\varepsilon\ \forall \varepsilon > 0$ – considering that $SSCE(BG, \varphi, C_1) = SCE(BG, f)$. Hence we simply extend their proof to the payoff uncertain case. Let σ^* be a MSCE of (BG, φ) in accordance with Definition 4. By Lemma 1, observability of payoffs implies that for any strategy s_i of any θ_i, $U_i(s_i, \theta_i, \hat{\pi}) = U_i(s_i, \theta_i, q_{-i}, \sigma^*_{-i})$ for every $\hat{\pi} \in \hat{\Sigma}_{-i,\theta_i}(s_i, q_{-i}, \sigma^*_{-i})$. Consider then some θ_i and $s^*_i \in \mathrm{supp}\sigma^*_{\theta_i}$; OSI implies that $\mathcal{F}_{s^*_i,\theta_i} = \mathcal{F}_{s_i,\theta_i}$ for every s_i that is a best reply to some conjecture. We therefore have

$$\begin{aligned}
&\hat{\Sigma}_{-i,\theta_i}(s_i, q_{-i}, \sigma^*_{-i}) = \\
&\qquad = \left\{\hat{\pi} \in \Delta(\Theta_0 \times \Theta_{-i} \times S_{-i}) : \hat{\pi}(\theta_0, \theta_{-i}, s_{-i})|\mathcal{F}_{s_i,\theta_i} = q_0(\theta_0) q_{-i}(\theta_{-i}) \sigma^*_{-i}(s_{-i}|\theta_{-i})|\mathcal{F}_{s_i,\theta_i}\right\} \\
&\qquad = \left\{\hat{\pi} \in \Delta(\Theta_0 \times \Theta_{-i} \times S_{-i}) : \hat{\pi}(\theta_0, \theta_{-i}, s_{-i})|\mathcal{F}_{s^*_i,\theta_i} = q_0(\theta_0) q_{-i}(\theta_{-i}) \sigma^*_{-i}(s_{-i}|\theta_{-i})|\mathcal{F}_{s^*_i,\theta_i}\right\} \\
&\qquad = \hat{\Sigma}_{-i,\theta_i}(s^*_i, q_{-i}, \sigma^*_{-i})
\end{aligned}$$

Hence we can prove that σ^* is a Bayesian equilibrium:

$$\begin{aligned}
U_i(s^*_i, \theta_i, q_{-i}, \sigma^*_{-i}) &= \min_{\hat{\pi} \in \hat{\Sigma}_{-i,\theta_i}(s^*_i, q_{-i}, \sigma^*_{-i})} U_i(s^*_i, \theta_i, \hat{\pi}) \geq \min_{\hat{\pi} \in \hat{\Sigma}_{-i,\theta_i}(s_i, q_{-i}, \sigma^*_{-i})} U_i(s^*_i, \theta_i, \hat{\pi}) \\
&= \min_{\hat{\pi} \in \hat{\Sigma}_{-i,\theta_i}(s_i, q_{-i}, \sigma^*_{-i})} U_i(s_i, \theta_i, \hat{\pi}) = U_i(s_i, \theta_i, q_{-i}, \sigma^*_{-i})
\end{aligned}$$

□

INNOVATIONS AND WORK: ASSESSING THE IMPACT OF AUTOMATION ON LABOR OUTCOMES THROUGH A CROSS-COUNTRY AND CROSS-INDUSTRY ANALYSIS

Sharmista Lakkineni[†]

DOI: 10.82029/2025015

Abstract. This paper examines the relationship between technological change and labor market outcomes using cross-country and cross-industry data spanning 25 years (1995–2020). Specifically, it investigates the impact of automation, proxied by total factor productivity (TFP) growth, on two key labor market indicators: aggregate employment and aggregate labor share of value added. The theoretical framework of this paper derives from Autor and Salomons (2018) and delineates four channels—comprising one direct effect and three indirect effects—by which automation influences labor market outcomes. This paper extends A&S' analysis by 13 years and makes methodological changes by revising the lag structure of TFP growth to account for longer, more variable innovation-to-productivity effects and incorporating previously omitted controls to capture the final demand effects of automation. Therefore, the primary goal of this paper is to reassess Autor and Salomons (2018)'s analysis by employing methodologies better suited to the new data. The theoretical framework in this paper is based on the core idea that, while certain technological innovations may displace labor, countervailing responses within the economy can mitigate downward shifts in aggregate labor demand, making it crucial to estimate both the direct and indirect effects of automation. Consistent with Autor and Salomons (2018), this paper finds that for both employment and labor share of value added, there is a negative direct effect of automation on labor outcomes in the industry where the innovation occurs. For employment, this negative direct effect is offset by countervailing forces elsewhere in the economy — including upstream and downstream linkages, final demand effects, and compositional changes — yielding a net positive effect of automation. In contrast, the net effect on labor share of value added is negative, reflecting the broader trend of declining labor shares observed in developed countries over recent decades. However, our quantitative analysis also reveals a significant discrepancy compared to A&S' findings: the size of the net positive impact of automation on aggregate employment is notably diminished. This disparity suggests that recent technological advancements, such as advanced artificial intelligence and robotics introduced in the last decade, may differ fundamentally in their labor market effects from older technologies. Consequently, these newer technologies may exert a less favorable influence on labor outcomes, underscoring the need for nuanced understanding and strategic adaptation to technological change in contemporary economies.

JEL classification: J21, J24

Keywords: Automation, Labor Outcomes, Macro–Micro Linkages.

1. INTRODUCTION

Automation can be defined as an expansion of the set of tasks where capital can substitute for labor (Acemoglu and Restrepo (2018a)). The last two decades have witnessed major technological

[†]Lakkineni: sharmista.lakkineni@studbocconi.it. I am grateful for the guidance provided by Professor Natalie Chen, Professor James Fenske, and Dr Samuel Obeng from the University of Warwick. I thank the editor and two anonymous referees for useful comments and feedback.

advances, particularly in artificial intelligence (AI) and Robotics, that have the potential to reshape future lifestyles and workplaces by automating various human activities. A recent US survey (Smith and Anderson (2017)) indicates that, despite some optimism, there is widespread public unease about the societal impacts of these innovations, particularly their potential to displace entire job categories. Despite such growing public sentiment, economic literature is far from providing a consensus as to whether automation, particularly new technologies like AI and robotics, will ultimately displace labor.

Acemoglu and Restrepo (2018b) highlight that the trend of declining labor share of value added in developed economies has often been cited in Economics as evidence to claim that advancements in technologies like robotics and AI have contributed to, at the very least, a relative decline in workers' compensation. However, they stress the inadequacy of existing frameworks in fully accounting for both the direct impacts of automation and the potential countervailing effects. This deficiency underscores the need for a comprehensive framework that can capture the complexities of technological change in the labor market.

This paper extends the research conducted by David Autor and Anna Salomons (2018) (henceforth, A&S), refining and applying their theoretical framework comprising the direct and indirect effects of industry-level automation. It utilizes the latest sample of the EUKLEMS & INTANprod cross-country and cross-industry data (1995–2020) to estimate the impact of automation on labor market outcomes. Labor displacement due to automation can either take the form of employment displacement or erosion of labor's share of value added in the economy. Similar to Autor and Salomons (2018), employment is measured as both the number of workers and hours worked, while automation is measured as other-country, industry-level total factor productivity (TFP) growth.

The theoretical basis of this paper is constructed upon a framework that outlines four channels through which industry-level automation influences aggregate labor market outcomes. The first of these channels is the direct effect of automation on employment and labor shares within the industries where the automation occurs. However, productivity growth due to automation will not just be confined in the industry where it originates but will also affect employment and labor shares elsewhere in the economy. The remaining three channels measure the indirect and potentially countervailing effects of automation across the broader economy. The interplay of these four channels determines the net impact of automation, and whether it ultimately proves labor-displacing (Autor and Salomons (2018)).

Diverging from alternative approaches in this domain, this paper, akin to Autor and Salomons (2018), employs TFP growth as the primary measure of technological advancement, as opposed to more focused metrics like robotics (Graetz and Michaels (2018)) or routine task replacement. While this "omnibus" or "all-encompassing" measure might fall short in reflecting particular types of technological progress, the main advantage of using TFP growth to measure automation is that it eliminates the heterogeneity in innovation across sectors and time periods. Moreover, all margins of technological progress ultimately lead to an increase in TFP by increasing labor or capital productivity or reallocating tasks between capital and labor (A&S, 2018).

This paper contributes to the literature in several ways. Firstly, it extends the analysis of the impact of industry-specific TFP growth on labor market outcomes using Autor and Salomons (2018)'s theoretical framework. While their original analysis spanned from 1970 to 2007, significant advancements in robotics and AI have occurred since 2007. By examining data from 1995 to 2020, this study aims to evaluate the impact of this new wave of automation, driven by technologies such as AI and robotics, and ascertain whether the conclusions drawn by Autor and Salomons (2018) remain applicable in the current technological landscape or if this era of automation represents a departure from previous waves of technological progress. Secondly, this paper makes methodological changes to the regression specifications used by Autor and Salomons (2018). Primarily, the lag structure of the explanatory variable (TFP growth) is changed in line with the demands of the new data and based on literature examining the persistence of contemporaneous TFP shocks, which suggest longer and more variable lags between innovation and productivity growth than accounted for by A&S (Gort and Klepper (1982); Foster et al. (2018)). Moreover, several additional controls that were omitted in A&S' analysis are added to Specification 3 to measure the final demand effects of automation.

A preliminary summary of the results of the paper is as follows: Automation as embodied in TFP growth has been slightly employment-augmenting but significantly labor-share-displacing across the 25 years in the data sample (1995–2020). The results of this paper are qualitatively similar to those of Autor and Salomons (2018) who also find the net effect on employment to be positive but that on labor share to be negative. However, quantitatively, the net effects of TFP growth on aggregate employment and hours worked are much smaller in this paper than those of Autor and Salomons (2018). This discrepancy in results for aggregate employment and hours worked can be attributed to the following: (1) differences in lag structures employed in the regression specifications; (2) the inclusion of additional controls in specification 3 of this paper, which significantly reduce the positive productivity effects found in A&S' analysis; and (3) a potentially fundamental difference in the way new technologies of the last two decades affect employment compared to technologies of the past. For labor share of value added, the results of this paper are both qualitatively and quantitatively similar to those of A&S. The appendix also details a series a robustness checks for the estimates of this paper, which show the stability of the results.

2. LITERATURE REVIEW

The economic literature on the impact of automation on the future of work is vast and varied. Acemoglu and Restrepo (2019) point out that the debate on this issue in economics is characterized by a "false dichotomy": one side contends that the rise in automation will lead to the elimination of labor-intensive and cognitively demanding work, leaving an increasingly dwindling set of activities where labor can add value; while a contradictory view claims that automation will actually increase labor demand and wages like other waves of technological innovations have done in the past. Aghion et al. (2022) also highlight the existence of a dichotomy, classifying economic literature regarding the impact of automation on labor into the "old view of negative direct effects and positive indirect effects" and the "new view of positive direct effects and negative indirect effects."

The "negative direct effects and positive indirect effects" view (old view) suggests that the direct effect of automation is to displace employment, reduce labor shares, and suppress wages. However, there exist countervailing forces that increase the demand for labor and limit the wage decline induced by automation. There are varying views within this strand as to what the countervailing forces actually are.

Acemoglu and Restrepo (2019) highlight several countervailing forces like reduced costs of production; increased productivity in previously automated tasks, i.e., deepening of automation; and induced capital accumulation. However, they argue that these forces are generally incomplete and by themselves cannot balance out the direct negative effects of automation. Despite this, previous waves of technological innovation and automation have not induced a secular downward trend in labor shares. They posit that this is because of an even powerful countervailing force — the creation of new labor-intensive jobs by automation. However, Aghion et al. (2017) suggest a different counterbalancing force, pointing to the well-known "Baumol Cost Disease" effect. Baumol (1967) suggests that "economic growth is constrained not by what we do well but instead what is essential and hard to improve," which is why sectors with rapid productivity like agriculture and industry see declines in GDP shares, whereas relatively slower productivity growth sectors like services see increases. In the context of automation and labor outcomes, because labor tasks are the "weak link," i.e., they are essential but expensive, labor shares remain elevated because of the Baumol force (Aghion et al. (2017)).

The old-view analyses, which are primarily run at industry or national levels, show mixed results regarding the impact of automation but generally lean more towards the view emphasizing the negative impact of automation on labor outcomes. Within the broad framework of the old-view analyses, varying measures of automation have been used to conduct empirical analysis. Early papers (Krueger (1993); Autor et al. (1998); Bresnahan et al. (2002)) use computers or IT as proxies for automation. A newer measure of automation involves using automation-related patents. Research based on this method provides mixed evidence — Mann and Püttmann (2017) find negative effects of automation on employment, while Webb (2019) finds positive effects. Autor and Salomons (2018) use industry-level movements in TFP as a measure of automation and find net positive effects for employment and net negative effects for labor share of value added. Recently, with the provision of data on the deployment of robots at the country and industry level by the International Federation of Robotics (IFR), several papers have used exposure to robots as a measure of automation (Autor and Dorn (2013); Acemoglu and Restrepo (2020); Cheng et al. (2019); Dauth et al. (2021); Graetz and Michaels (2018)). In these papers, the job-destruction effect of automation dominates.

The “positive direct effects and negative indirect effects” view (new view) suggests that automation increases employment at the firm level because the increase in productivity induced by automation allows the firms that adopt automation to offer better quality-adjusted prices compared to opponent firms, thereby expanding their market size, and increasing labor demand. However, there might be an overall (indirect) negative effect if firms adopting automation cause substantial decreases in employment for non-automating firms, leading the latter to exit the market (Aghion et al. (2022)).

The new-view analyses are based on firm-level research, which has been made plausible with the availability of firm-level microdata on technology adoption in recent years. These studies conducted using such firm-level data from a wide range of countries, including the United States (Acemoglu et al. (2022)), the United Kingdom (Webb (2019)), and France (Aghion et al. (2021); Acemoglu et al. (2020)), generally predict a positive effect of automation on employment in firms adopting the automation. Therefore, this relatively new line of research on automation at the firm level is seemingly at odds with the old strand of literature that supports a more pessimistic view of automation (Aghion et al. (2022)).

3. DATA

The main analysis of this paper draws from the 2023 release of the EU KLEMS & INTANprod database, which is an industry-level panel dataset providing information on output, productivity, employment, and capital formation for developed countries for the period 1995–2020. The analysis in this paper is limited to 12 developed countries: Austria, Belgium, Germany, Denmark, Spain, Finland, France, Italy, Netherlands, Sweden, the United Kingdom, and the United States of America. Moreover, the analysis focuses on 27 market industries out of the 42 industries available in the database. It drops non-market sectors such as public administration, defense, and the private household sector. The 27 market industries can be grouped into five broad sector groups: (1) mining, utilities, and construction; (2) manufacturing; (3) high-tech services; (4) low-tech services; and (5) health and education (See Appendix Table A1).

Specification 2 of this paper requires calculating supplier and customer weights of industries. These are calculated from the input-output coefficients given in the World input-output Database (Timmer et al. (2015)).

Each specification controls for country-level business cycle effects. Data on business cycle indicators are acquired from OECD Composite Leading Indicators: Reference Turning Points and Component Series.

3.1. **Summary Statistics**

Table 1 summarizes trends in aggregate hours worked and labor share of value added by country. It shows that growth of log hours worked has been positive without exception; that is, employment as measured by aggregate labor hours has been rising in all countries in the last three decades. However, the rate of growth of labor hours has been declining, with the most rapid growth seen in the 1990s and the lowest growth during the 2010s. The growth patterns of labor share of value added are more varied across countries and over time: on average, labor share is rising in the 1990s but falling in subsequent decades, with the sharpest decline in the 2000s.

Tables 2 and 3 summarize the trends in employment, hours, labor share, and TFP by decades and by five broad industry sector groups, respectively. They show results from baseline regressions of decade and sector group dummy variables on the main variables of interest. All regression models in this paper are weighted by time-averaged shares of the relevant weighting variable (employment, hours worked, value added), multiplied by time-varying country shares in the total annual value of the weighting variable. Employment growth (the number of workers and hours) is positive in all decades but is lowest in the 2000s. It is negative in manufacturing; strongly positive in services and health and education; and modestly positive in mining, utilities, and construction. Labor share is strongly negative in manufacturing and modestly positive or negative in others. TFP growth is strongly positive in manufacturing. The baseline regression results, therefore, paint the following broad picture: sectors where productivity growth is the highest (e.g., manufacturing) show the largest declines in labor outcomes (the number of workers, hours worked, value added shares).

Table 1: Trends in Hours Worked and Labor Shares of Value Added

Country	**100 × Δ log hours worked**			**100 × Δ log labor share**		
	1990s	**2000s**	**2010s**	**1990s**	**2000s**	**2010s**
Austria	2.124	1.043	0.115	-0.766	-0.318	0.385
Belgium	3.215	2.611	1.196	0.392	-0.167	-0.773
Germany	1.106	1.127	0.782	0.286	-0.251	0.519
Denmark	2.932	0.908	0.633	-0.806	0.351	-0.174
Spain	5.363	2.581	-0.96	1.217	-1.439	0.105
Finland	2.675	1.38	0.831	-0.852	0.108	-0.935
France	2.36	1.777	0.429	-0.639	0.014	0.102
Italy	2.168	1.436	2.003	-0.539	-0.194	-2.178
Netherlands	3.652	1.466	1.557	0.08	-0.124	0.701
Sweden	1.518	1.291	1.262	0.666	0.393	0.349
UK	2.198	1.659	1.109	0.936	1.525	0.146
USA	5.367	0.156	1.162	2.661	-1.711	0.587
Average	2.89	1.453	0.843	0.22	-0.153	-0.097

4. METHODOLOGY

Following Autor and Salomons (2018), this paper makes use of an accounting framework to estimate the aggregate impact of automation on labor outcomes. The specifications below are taken from Autor and Salomons (2018). However, methodological changes have been made to some of the specifications (details below) to both refine A&S' analysis and also to ensure it is consistent with the new data being used in this paper.

The key outcome variables measured in this paper are (1) employment as measured by "log change in labor hours worked" and "log change in number of people employed" and (2) "log change

Table 2: Trends in Key Variables (By Decade)

Decade	Emp	Hours	Lab Share	TFP
1990s	2.230***	2.172***	0.890	0.399
	(0.251)	(0.267)	(0.676)	(0.245)
2000s	0.577***	0.075	-0.229	0.140
	(0.222)	(0.240)	(0.157)	(0.228)
2010s	1.197***	1.168***	-0.018	0.385***
	(0.135)	(0.145)	(0.137)	(0.128)

Notes. Standard errors clustered by country-industry in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Trends in Key Variables (By Broad Sector)

Broad Sector	Emp	Hours	Lab Share	TFP
Health and Education	1.760***	1.571***	0.772	-0.960***
	(0.132)	(0.169)	(0.475)	(0.069)
High-Tech Services	1.743***	1.516***	0.109	0.478
	(0.251)	(0.171)	(0.131)	(0.467)
Low-Tech Services	0.769***	0.289*	0.033	0.158
	(0.196)	(0.147)	(0.147)	(0.299)
Manufacturing	-1.033***	-1.306***	-0.320***	1.188***
	(0.181)	(0.175)	(0.085)	(0.237)
Mining, utilities, and construction	0.462*	0.275	-0.117	-0.687***
	(0.252)	(0.298)	(0.179)	(0.094)

Notes. Standard errors clustered by country-industry in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

in labor shares." The main measures of automation are industry-level changes in "TFP."

4.1. Addressing Endogeneity Concerns and Timing Issues

Before moving on to the main methodological framework, it is useful to highlight how this paper overcomes endogeneity and other concerns that arise in the data.

The primary endogeneity concerns relate to the use of TFP changes as a measure of automation. TFP estimates can be confounded with business cycle effects, trends within industries, and cross-industry differences in cyclical sensitivity. Moreover, a simultaneity issue arises because labor share of value added (dependent variable) is used in the calculation of TFP growth (independent variable), inducing a mechanical correlation between the two.

$$\Delta lnTFP_j = \Delta lnV_j - \bar{v}_{L,j}\Delta lnL_j - \bar{v}_{K,j}\Delta lnK_j \tag{1}$$

In equation (1), $\Delta \ln \mathrm{TFP}_j$ denotes the log change in TFP for industry j; $\Delta \ln V_j$ is the log change in value added; $\Delta \ln L_j$ and $\Delta \ln K_j$ are the log changes in labor and capital inputs, respectively; and $\bar{v}_{L,j}$ and $\bar{v}_{K,j}$ are the labor and capital factor shares used as weights, respectively.

To overcome the simultaneity issue, "Leave-Out Mean of Industry-Level TFP growth" in all other countries is used as a proxy for "Own-country Industry-Level TFP growth." This implies that, for a given industry and country under consideration, the TFP growth for that pair is left out of the calculation. The average or mean value is then taken from the TFP growth rates of the same industry in all the other countries in the sample. This approach assumes that movements in the technological frontier in a particular industry are common among industrialized economies.

Table 4 confirms the utility of using "Leave-Out Mean industry-level TFP growth" by showing that other-country, same-industry TFP growth is a strong predictor of own-country-industry TFP growth. The regression results below are robust to country, year, sector, and business cycle main effects.

TABLE 4: Relationship between Leave-Out Mean and Own-Country-Industry TFP growth

	dlnTFP_I ***(Own-country)***							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
dlnTFP_OT ***(Leave-Out Mean)***	0.566***	0.558***	0.554***	0.552***	0.555***	0.526***	0.441***	0.524***
	(0.115)	(0.127)	(0.124)	(0.123)	(0.125)	(0.119)	(0.117)	(0.117)
Fixed effects								
Country	NO	YES	YES	YES	YES	YES	YES	YES
Year	NO	NO	YES	YES	YES	YES	YES	YES
Country × Time Trend	NO	NO	NO	YES	NO	NO	NO	NO
Country × Business Cycle	NO	NO	NO	NO	YES	NO	NO	NO
Country × Year	NO	NO	NO	NO	YES	YES	NO	NO
Sector group	NO	NO	NO	NO	NO	YES	NO	YES
Industry	NO	NO	NO	NO	NO	NO	YES	YES
R^2	0.114	0.121	0.128	0.138	0.165	0.170	0.204	0.143
N	6 854	6 854	6 854	6 854	6 854	6 854	6 854	6 854

NOTES. Standard errors in parentheses are clustered by country-industry. All models weighted by industry value added shares within countries, multiplied by time-varying country shares in total value added. * p<0.10; ** p<0.05; *** p<0.01.

A second issue arises due to timing. Contemporaneous productivity innovations are unlikely to introduce steady-state effects immediately. In the original paper by A&S, a lag structure for log change in TFP that comprises contemporaneous and five distributed lags is employed. They justify their lag structure by using projection models that involve regressing a series of first differences of increasing length of the outcome variables on TFP growth (Òscar Jordà (2005)). They find that the effects of TFP growth on the outcome variables plateau after 3 years and, therefore, argue that not more than four or five lags of TFP growth are needed to capture the impulse response of a contemporaneous TFP shock. However, from the perspective of previous economic literature (Gort and Klepper (1982); Foster et al. (2018)) as well as comments on the original paper by Autor and Salomons (2018), a 5-year lag specification is likely insufficient to capture the dynamic effects of TFP growth. In fact, in the sample used by this paper, the Oscar Jorda projection models predict that the effects of contemporaneous TFP shocks last much longer than 3 years and only plateau after a period of 7–8 years. Therefore, the regression specifications in this paper employ a lag structure of contemporaneous and seven distributed lags to ensure completeness and also use as many observations in the data as possible (See Appendix A.4. for projection model results and further explanation).

Lastly, several fixed effects are employed in the regression specifications to control for country effects, time trends, business cycle effects, and sector group effects. Additionally, since this paper estimates "first-difference" models, industry-country effects are implicitly eliminated.

4.2. **Methodological Framework: Direct and Indirect Effects of Industry-Level Innovations**

The aggregate impact of automation on labor market outcomes is the net total of four smaller effects.

4.2.1. *Direct industry-level effects.* The first main specification estimates the within-industry "direct" effects of TFP growth on own-industry outcomes.

$$\Delta lnY_{i,c,t} = \beta_0 + \sum_{k=0}^{7} \beta_1^k \Delta lnTFP_{i,c\neq c(i),t-k} + \alpha_c + \delta_t + \alpha_c \times t + \alpha_c \times (t = peak) \\ + \alpha_c \times (t = trough) + \epsilon_{i,c,t} \tag{2}$$

Here, i indexes industry, c indexes country, and t indexes year. $\Delta lnY_{i,c,t}$ is the outcome of interest and $\Delta lnTFP_{i,c\neq c(i),t-k}$ is the leave-out mean of TFP growth. This is a first-difference specification estimated at the industry-country-time level, and so implicitly eliminates industry-country effects. α_c represents country trends; δ_t represents time trends; $\alpha_c \times t$ represents country-time interaction terms, which allow country trends to accelerate or decelerate over the sample interval; $\alpha_c \times (t = peak)$ and $\alpha_c \times (t = trough)$ represent country-specific cyclical peak and trough indicators interacted with country indicators, respectively, to take into account country-specific business cycles. Results for this specification are reported in Table 5.

TABLE 5: Results from Specification 1 — Direct Within-Industry Effects

	$\Delta \ln Y_{i,c,t}$ (Employment & Hours)					
	Employment			Hours		
	(1)	(2)	(3)	(1)	(2)	(3)
$\sum \Delta$ lnTFP$_{i,c,t-k}$	–1.271***	–0.707***	–0.691***	–1.127***	–0.482***	–0.482***
	(0.241)	(0.201)	(0.193)	(0.255)	(0.191)	(0.192)
Fixed Effects						
Country	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Sector group	No	Yes	Yes	No	Yes	Yes
Country×Time Trend	Yes	Yes	Yes	Yes	Yes	No
Country×Business Cycle	Yes	Yes	Yes	Yes	Yes	No
Country×Year	No	No	Yes	No	No	Yes
R^2	0.366	0.414	0.518	0.447	0.484	0.560
N	4866	4866	4866	4869	4869	4869

	$\Delta \ln Y_{i,c,t}$ (Labor Share)		
	(1)	(2)	(3)
$\sum \Delta \ln\text{TFP}_{i,c,t-k}$	–0.870***	–1.013***	–0.998***
	(0.214)	(0.247)	(0.253)
Fixed Effects			
Country	Yes	Yes	Yes
Year	Yes	Yes	Yes
Sector group	No	Yes	Yes
Country×Time Trend	Yes	Yes	No
Country×Business Cycle	Yes	Yes	No
Country×Year	No	No	Yes
R^2	0.070	0.071	0.112
N	4869	4869	4869

NOTES. * p<0.10; ** p<0.05; *** p<0.01. Standard errors in parentheses.

Table 5 shows the results for the effect of lagged TFP growth (leave-out mean) on three within-industry outcome variables: log number of workers (Employment), log hours of labor input (Hours), and log labor share of value added (labor share). Point estimates in each column are the sum of the six β_1^k coefficients.

Industries experiencing relative gains in productivity exhibit relative declines in employment. The coefficient in column 1 of employment can be interpreted as follows: an increase in 1 SD of TFP growth predicts a fall in own-industry employment by ≈ 1.3 log points. Inclusion of sector-group fixed effects reduces the estimate from -1.3 to -0.70. This indicates that TFP innovations may spill over across industries within a sector. These spillovers are modelled in equation 2. Estimates are analogous for labor hours worked. However, inclusion of sector-group fixed effects decreases both the significance and absolute value of the estimates. Similar to employment and hours worked, a negative relationship between TFP growth and labor shares can also be seen: A rise in TFP of 1 SD predicts a fall in own-industry labor shares by about -0.87 percentage points over a 7-year horizon.

This paper tests the robustness of these estimates using a variety of checks: (1) weighting all countries equally instead of by country size; (2) excluding contemporaneous TFP terms in specification 1 and only including lags of TFP; (3) eliminating the group of self-employed workers from the sample; and (4) imputing zeros for negative reported values of the TFP measure. The results of the robustness checks are reported in the appendix (Table A.3). The results show that the estimations in Table 5 are largely stable even after including the above robustness checks.

Therefore, the results from Table 5 indicate that industries experiencing rising productivity show negative labor outcomes within the industry where the automation occurs. However, it would be erroneous to conclude using just the results in Table 5 that productivity growth leads to labor displacement on the aggregate. Relative employment declines in industries with rising productivity do not imply that aggregate employment falls as productivity rises. The first specification estimated within-industry effects and did not incorporate potentially countervailing effects operating through other channels.

4.2.2. *Indirect Effect of Consumer and Supplier-linked sectors.* The productivity growth resulting from automation is likely to extend beyond the sector where the automation initially occurs, impacting both customer and supplier industries associated with the originating firm. For instance, industries with more efficient suppliers may experience an uptick in purchases, while those with more efficient customers may encounter fluctuations in output demands. These input-output linkages are considered by adding two new terms to the regression specification seen above.

$$\Delta lnY_{i,c,t} = \beta_0 + \sum_{k=0}^{7} \beta_1^k \Delta lnTFP_{i,c\neq c(i),t-k} + \alpha_c + \delta_t + \alpha_c \times t + \alpha_c \times (t = peak)$$
$$+\alpha_c \times (t = trough) + \sum_{k=0}^{7} \beta_1^k \Delta ln\widetilde{\text{TFP}}^{SUP}_{i,c\neq c(i),t-k} + \sum_{k=0}^{7} \beta_1^k \Delta ln\widetilde{\text{TFP}}^{CONS}_{i,c\neq c(i),t-k} + \epsilon_{i,c,t} \tag{3}$$

The two additional TFP terms measure the weighted sum of TFP growth in all other domestic industries $j \neq i$ that are customer/suppliers of industry i.

The customer and supplier weights are calculated using the 2016 version of the World Input-Output Database (Timmer et al. (2015)). The supplier weights represent the proportion of value added from each domestic supplier industry j in relation to the value added of industry i. Similarly, the customer weights denote the proportions of value added from each industry i utilized in the final products of domestic industry j. Just as with the within-industry TFP growth terms, for the supplier and customer industries, industry-level, leave-out means of TFP growth for all other countries in the sample are used.

Results from Equation 2, which includes terms indicating TFP growth in customer and supplier industries, are given in Table 6. It is in this specification that this paper's results differ significantly from those of Autor and Salomons (2018). In their paper, productivity growth emanating from supplier industries predicts steep increases in employment, labor hours and wage bill, although not in labor share of value added. They point out that the positive effect on labor outcomes of supplier industry productivity growth reveals a first channel by which the negative within-industry effects of productivity on labor outcomes are countervailed.

Table 6: Results from Specification 2 — Customer & Supplier Linkages

	Industry Effects: Employment & Hours & Wage Bill		
	(1) Emp	(2) Hours	(3) Wage Bill
Own-Industry TFP growth	–0.722**	–0.500***	–0.271
	(0.215)	(0.203)	(0.233)
Supplier-Industry TFP growth	0.293	0.320	0.319
	(0.431)	(0.476)	(0.368)
Customer-Industry TFP growth	–0.024	–0.049	–0.022
	(0.271)	(0.307)	(0.326)
Fixed Effects			
Country	Yes	Yes	Yes
Year	Yes	Yes	Yes
Sector group	Yes	Yes	Yes
Country×Time Trend	Yes	Yes	Yes
Country×Business Cycle	Yes	Yes	Yes
Country×Year	Yes	Yes	Yes

The estimates from this paper's data sample show that supplier- and customer-industry TFP growth does not have a significant effect on employment or hours worked. However, the effect on labor share of value added of customer industry TFP growth is negative and statistically significant. This indicates that when there is a positive technology shock in customer industries, while there are no upstream effects in terms of employment for supplier industries, the labor share of value added of supplier industries decreases.

The differences in these results relative to Autor and Salomons (2018) could potentially indicate

Industry Effects: Nom. Value Added & Real Value Added & Labor Share			
	(4) Nom. VA	(5) Real VA	(6) Labour Share
Own-Industry TFP growth	0.730***	2.101***	–0.991***
	(0.228)	(0.459)	(0.207)
Supplier-Industry TFP growth	0.841	0.644	–0.769
	(0.544)	(0.446)	(0.494)
Customer-Industry TFP growth	1.197***	–0.779	–1.418***
	(0.242)	(0.514)	(0.195)
Fixed Effects			
Country	Yes	Yes	Yes
Year	Yes	Yes	Yes
Sector group	Yes	Yes	Yes
Country×Time Trend	Yes	Yes	Yes
Country×Business Cycle	Yes	Yes	Yes
Country×Year	Yes	Yes	Yes

NOTES. * p<0.10; ** p<0.05; *** p<0.01. Standard errors in parentheses.

that newer technologies of the last two decades have different effects on labor than the technologies of the past.

4.2.3. *Final Demand Effects Through the Effect of Productivity.* Productivity in any one industry augments aggregate income and raises final demand, implying that productivity growth in any one sector can lead to increased labor demand across all other sectors (this is because labor demand is a derived demand). The regression specification below estimates the relationship between country-specific aggregate economic growth and industry-specific inputs.

$$\Delta lnY_{i,c,t} = \lambda_0 + \sum_{k=0}^{7} \lambda_1^k \Delta lnValueAdded_{j\neq i,c,t-k} + \alpha_s + \epsilon_{i,c,t} \tag{4}$$

$\Delta lnValueAdded_{j\neq i,c,t-k}$ represents the growth of own-country real or nominal value added, where own-industry output is excluded from the measure to prevent mechanical correlation with industry outcomes.

This specification is estimated in two different ways in this paper. Firstly, the regression given above is estimated in line with A&S (2018), where identification arises from sector-group fixed effects (α_s), and the indicator variables for country, year, and business cycles from specifications 1 and 2 are dropped. Additionally, this paper also estimates equation (5) including country, year, and business cycle fixed effects consistent with specifications 1 and 2.

$$\begin{aligned}\Delta lnY_{i,c,t} = \lambda_0 + \sum_{k=0}^{7} \lambda_1^k \Delta lnValueAdded_{j\neq i,c,t-k} + \alpha_s + \alpha_c \times t \\ + \alpha_c \times (t = peak) + \alpha_c \times (t = trough) + \epsilon_{i,c,t}\end{aligned} \tag{5}$$

Regression Table 7 indicates a potential second countervailing effect to the negative within-industry effects of productivity on labor outcomes.

When a similar specification to A&S (2018) is estimated, using only sector-group fixed effects to estimate the final demand effects of industry-level TFP growth, the results of this paper are similar to theirs. The results show that each log point increase in the country-level real value added predicts a ≈ 0.3 log point and ≈ 0.5 log point increase in same-country, other-industry employment and hours worked, respectively. This could potentially indicate that TFP growth emanating from

TABLE 7: Results from Specification 3 — Final Demand Effects (Using Limited Controls)

	(1) Employment	(2) Hours Worked	(3) Real Value Added
Aggregate Real Value Added	0.261***	0.518***	0.930***
	(0.092)	(0.097)	(0.147)
Sector-Group Fixed Effects	Yes	Yes	Yes

	(4) Wage Bill	(5) Nominal Value Added	(6) Labor Share
Aggregate Nominal Value Added	0.852***	1.013***	–0.147
	(0.068)	(0.112)	(0.109)
Sector-Group Fixed Effects	Yes	Yes	Yes

NOTES. * p<0.10; ** p<0.05; *** p<0.01. Standard errors in parentheses.

any one sector raises final demand in the economy, implying that each industry's productivity growth contributes to aggregate labor demand across all sectors.

TABLE 8: Results from Specification 3 — Final Demand Effects (Using Full Controls)

	(1) Employment	(2) Hours Worked	(3) Real Value Added
Aggregate Real Value Added	–0.470	0.125	0.109
	(0.300)	(0.352)	(0.362)
Fixed Effects			
Country	Yes	Yes	Yes
Year	Yes	Yes	Yes
Sector group	Yes	Yes	Yes
Country×Time Trend	Yes	Yes	Yes
Country×Business Cycle	Yes	Yes	Yes
Country×Year	Yes	Yes	Yes

	(4) Wage Bill	(5) Nominal Value Added	(6) Labor Share
Aggregate Nominal Value Added	0.946***	0.430	0.413
	(0.352)	(0.349)	(0.320)
Fixed Effects			
Country	Yes	Yes	Yes
Year	Yes	Yes	Yes
Sector group	Yes	Yes	Yes
Country×Time Trend	Yes	Yes	Yes
Country×Business Cycle	Yes	Yes	Yes
Country×Year	Yes	Yes	Yes

NOTES. * p<0.10; ** p<0.05; *** p<0.01. Standard errors in parentheses.

However, as shown in Table 8 below, these estimates are NOT robust to the inclusion of a full set of country, year, business cycle, and sector-group fixed effects. The inclusion of these fixed effects, which were omitted by Autor and Salomons (2018) in their paper, results in the estimates for employment, hours worked, and labor shares to be insignificant. This is in contrast with the significant and positive final demand effects on employment found in table 7.

The results from table 8 indicate that the final demand effects through the effect of productivity do not necessarily act as a countervailing force to the negative within-industry effects of TFP growth, as implied by Autor and Salomons (2018).

4.2.4. *Compositional Between-Sector Effects.* Uneven productivity growth across industries can shift the aggregate labor share through changes in relative sector sizes.

To quantify the importance of within-industry vs. between-industry shifts in labor shares, a simple shift-share decomposition can be estimated:

$$\Delta \bar{L}_{c,t} = \sum_i \bar{\omega}_{i,c,t} \Delta l_{i,c,t} + \sum_i \Delta \omega_{i,c,t} \bar{l}_{i,c,t} \tag{6}$$

The term $\Delta \bar{L}_{c,t}$ on the left indicates the change in aggregate log labor share in country c over time interval τ. The first term on the right is the contribution of within-industry changes in labor share and the second term is the contribution of between-industry shifts in labor share to aggregate changes in labor share. Here, $l_{i,c,t}$ is the log labor shares in each industry i; $\omega_{i,c,t}$ is the weight corresponding to industry i's share of value added in country c at time τ; $L_{c,t}$ therefore is the weighted sum of log labor shares of industry i: $L_{c,t} = \sum_i \omega_{i,c,t} l_{i,c,t}$.

Estimating the second term in the decomposition will give us the contribution of between-industry shifts in aggregate labor shares. If technological progress leads to an increase in the relative size (weight $\omega_{i,c,t}$) of industries with low average labor shares ($\bar{l}_{i,c,t}$), i.e., capital-intensive industries like manufacturing and mining, then this will indirectly have a negative effect on the aggregate labor share. However, if technological progress increases the relative importance of labor-intensive industries, like health and education, then this will indirectly raise the aggregate labor share.

Column 4 of Table 6 shows that an increase in own-industry TFP growth predicts an increase in industry-level nominal value added with an elasticity of 0.730. Therefore, sectors experiencing technological growth are expanding in the economy as a share of nominal value added. Since the relative sizes of industries experiencing technological growth are increasing, if these industries are mainly capital-intensive, then the aggregate labor share will fall. Conversely, if these industries are mainly labor-intensive, then the aggregate labor share will rise.

Table 9 shows that "Health and Education" is the sector group with the highest average labor share (most labor-intensive), while "Manufacturing" and "Mining, Utilities, and Construction" are the sector groups with the lowest average labor shares (most capital-intensive).

Table 9: Average Log Labor Shares by Industry Sector

Broad Industry Sector	Average Labor Share
Health and Education	86.46774
High-Tech Services	63.11317
Low-Tech Services	61.74819
Manufacturing	61.63176
Mining, utilities, and construction	50.25314

The summary statistics from Table 10 (snippet below) show that labor-intensive industries Health and Education have the lowest (negative) TFP growth. Manufacturing, a capital-intensive industry, has the highest TFP growth. Although mining and construction, which are also capital-intensive industries, do not show such positive TFP growth, overall, technological progress seems to be more concentrated in capital-intensive industries than in labor-intensive industries.

Table 10: Snippet from Table 3 (Summary Statistics)

Broad Industry Sector	100× Mean Annual Log Change TFP
Health and Education	-0.960***
	(0.069)
High-Tech Services	0.478
	(0.467)
Low-Tech Services	0.158
	(0.299)
Manufacturing	1.188***
	(0.237)
Mining, utilities, and construction	-0.687***
	(0.094)

The contribution of the between-sector shifts in labor shares to aggregate changes in labor shares is shown in the figures below with the final aggregate results.

4.3. **Aggregate Effects**

4.3.1. *Estimating Aggregate Effects.* Using estimates from all the individual specifications, this paper now quantifies the implied contribution of TFP growth on the evolution of **aggregate** employment and labor shares through all four channels outlined above.

For all outcome variables (employment, hours worked, and labor share of value added), the contributions of the first three effects (direct; supplier and customer; final demand) are estimated. The fourth compositional effect can only be estimated for labor share of value added.

The contribution of direct, within-industry effects on aggregate labor outcomes ($\Delta \ln Y_{c,t}^{\text{OWN}}$):

$$\Delta \ln Y_{c,t}^{\text{OWN}} \equiv \frac{\partial \ln Y_{c,t}}{\partial \ln \text{TFP}_{i,c \neq c(i),t}^{\text{OWN}}} = \sum_{k=0}^{7} \beta_1^k \times \sum_{i=1}^{I} \omega_{i,c} \times \Delta \ln \text{TFP}_{i,c \neq c(i),t}^{\text{OWN}} . \tag{7}$$

Here, $\ln Y_{c,t}$ is the log of the outcome variable (employment, hours, and labor shares) in country c at time t; $\sum_{k=0}^{7} \beta_1^k$ is the sum of the seven contemporaneous and lagged coefficients in Specification 1; $\omega_{i,c}$ is the average outcome-variable share in industry i in country c; and $\Delta \ln TFP_{i,c \neq c(i),t}^{\text{OWN}}$ is the own-industry TFP growth. Therefore, the direct-effect contribution can be calculated as the sum of the β_1^k's in Specification 1 multiplied by the corresponding weighted $\Delta \ln TFP_{i,c \neq c(i),t}^{\text{OWN}}$ term.

The contribution of supplier- and customer-industry effects ($\Delta \ln Y_{c,t}^{L}$) can be calculated in an analogous way:

$$\Delta \ln Y_{c,t}^{L} \equiv \frac{\partial \ln Y_{c,t}}{\partial \ln \widetilde{\text{TFP}}_{j \neq i,\, c,\, t}^{L}} = \sum_{k=0}^{7} \beta_1^k \times \sum_{i=1}^{I} \omega_{i,c} \times \Delta \ln \widetilde{\text{TFP}}_{j \neq i,\, c,\, t}^{L} , \tag{8}$$

$$L \in \{\text{SUP, CUST}\}.$$

The contribution of the final demand effect on aggregate outcomes ($\Delta \ln Y_{c,t}^{\text{FD}}$) will be the product of four terms:

1. The effect of TFP growth in industry i on the real value added in i ($\sum_{k=0}^{7} \beta_{1,\text{VA}}^k \rightarrow$ estimate from Specification 2; Table 6, column 6).

2. The effect of real value added growth in i on total real value added in the economy ($\phi_{i,c} \rightarrow$ can be calculated as the average value-added share of industry i in country c).

3. The effect of growth in real value added on employment, hours, and labor shares in each industry ($\sum_{k=0}^{7} \lambda_1^k \rightarrow$ estimate from Specification 3; Table 8).

4. The size of industry i relative to overall employment, hours, and labor shares in the economy ($\omega_{i,c}$).

$$\Delta \ln Y_{c,t}^{\mathrm{FD}} \equiv \frac{\partial \ln Y_{c,t}}{\partial \ln VA_{c,t}} \times \frac{\partial \ln VA_{c,t}}{\partial \ln \mathrm{TFP}_{i,c\neq c(i),t}} = \sum_{k=0}^{7} \lambda_1^k \times \sum_{k=0}^{7} \beta_{1,\mathrm{VA}}^k \times \sum_{i=1}^{I} \omega_{i,c} \times \phi_{i,c} \,. \tag{9}$$

Focusing only on the labor-share of value added, the contribution of compositional (between-industry) shifts is

$$\Delta \ln Y_{c,t}^{\mathrm{COMP}} = \sum_{i}^{I} \big(\Delta\widehat{\omega}_{i,c}\, \bar{\ell}_{i,c}\big), \tag{10}$$

where $\Delta\widehat{\omega}_{i,c}$ is the predicted change in the value added share of industry i in country c (measures how the relative importance of an industry changes); $\bar{\ell}_{i,c}$ is the average log labor share in industry i and country c.

4.3.2. *Aggregate Results.* The bars in the figures 1 and 2 show how each of the direct and three indirect effects of TFP growth affect employment and hours worked, on average, over the full 25-year period (1995–2020), respectively. The final fifth bar estimates the net effect on aggregate employment, summing the direct and indirect components, over the outcome period.

FIGURE 1: Direct and Indirect Effects of TFP growth on Employment

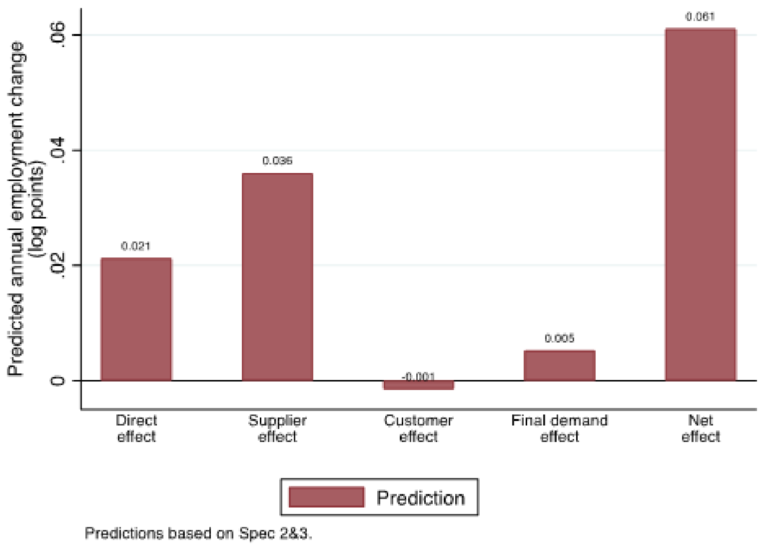


Figure 2: Direct and Indirect Effects of TFP growth on Hours Worked

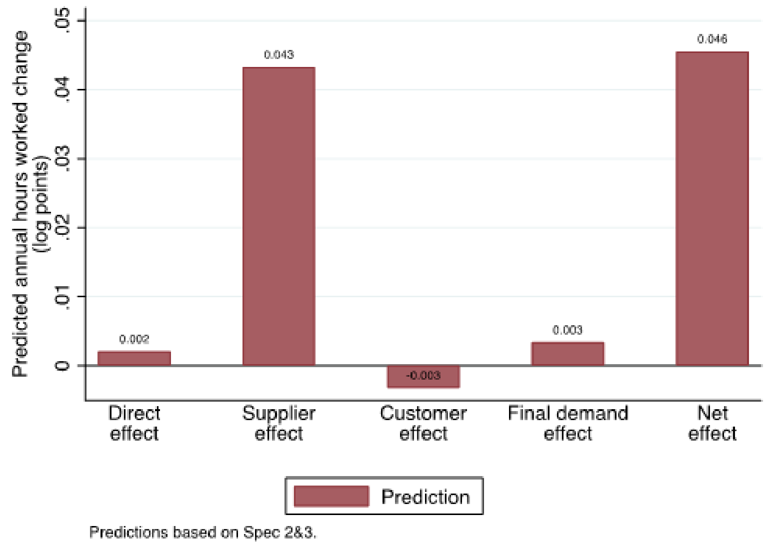
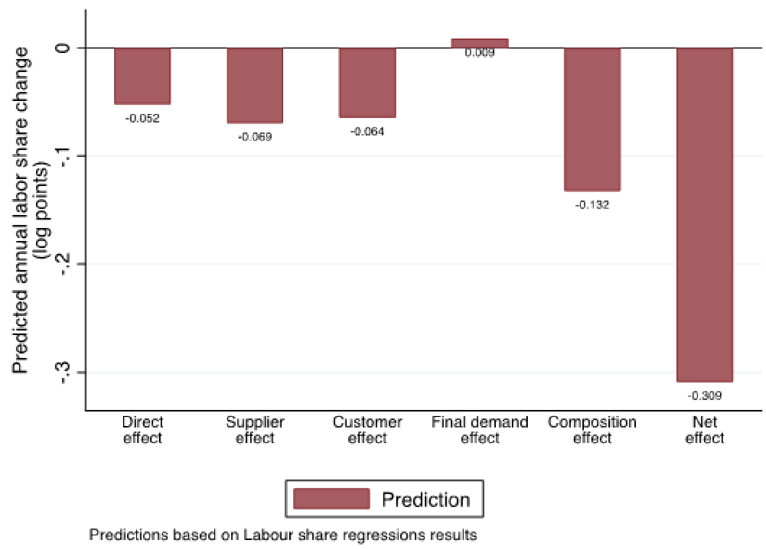


Figure 3: Direct and Indirect Effects of TFP growth on Labor Share of Value Added



A similar diagram is produced for labor share of value added (Figure 3). However, the fourth channel of between-sector compositional effects is also added.

The main conclusion from the diagrams is as follows: Automation as embodied in TFP growth has been slightly employment-augmenting but significantly labor-share displacing. For employment and hours worked, the positive indirect supplier, customer, and final demand effects seemingly offset the negative within-industry effect of productivity growth seen in Table 5. However, for labor share of value added, the positive offsetting indirect effects do not exist. In fact, the additional channels of interindustry linkages and final demand effects contribute negatively to aggregate labor share.

Therefore, these results imply that while technological progress, on the aggregate, has not led to workers being displaced from their jobs, the contribution of labor to the output produced in the economy has declined over the course of the 25 years in the sample.

5. LIMITATIONS

There are certain limitations of this paper that future work should aim to address. Firstly, while TFP growth is an omnibus measure that encompasses all types of technological progress, it might not adequately reflect automation, which is a particular type of technological progress. Therefore, to isolate components of technological growth that are more closely related to automation, future work should look at more direct measures of automation, including ICT-specific and AI-specific technological progress. Secondly, the leave-out mean TFP approach relies on the assumption that the technological frontier at the industry level is similar across countries. However, there is evidence to show that productivity growth at the industry-level varies across countries (ICT revolution in the 1990s was more concentrated in the US than other developed countries). Lastly, this paper shows that the direct within-industry TFP effect on employment is negative. However, firm-level studies find that there is strong positive correlation between TFP and firm-level employment growth. This discrepancy is because industry-level fluctuations in productivity reflect not only the within-firm innovations but also the between-firm innovations. These can only be considered if the industry-level data used in this paper are further refined with firm-level microdata.

6. CONCLUSION

Following the theoretical framework given by Autor and Salomons (2018), this paper uses cross-country, cross-industry data over a period of 25 years (1995–2020) to explore the relationship between industry-level changes in TFP and labor market outcomes through four channels: (1) the direct effects of changes in TFP on labor outcomes within the same industry; (2) the indirect effect of TFP changes in supplier and customer industries on own-industry labor outcomes; (3) the productivity effect of industry-level technological changes on aggregate labor demand; and (4) the compositional effect of TFP changes on shifts in labor shares between industries. This paper, therefore, contributes to not only the literature on the effects of new technologies on labor market outcomes but also the literature on macro–micro linkages that analyze how small shocks in one part of the economy are amplified and propagated throughout the economy through various direct and indirect channels (Acemoglu et al., 2016). This paper finds that for both employment and labor share of value added, there is a negative direct effect of automation on labor outcomes in the industry where the innovation occurs. For employment, this negative direct effect is offset by countervailing forces elsewhere in the economy, while for labor share of value added, it is not. These findings are in line with trends seen across developed countries of positive employment growth but declining contribution of labor in the economy.

REFERENCES

Acemoglu, D., U. Akcigit, and W. Kerr (2016). Networks and the Macroeconomy: An Empirical Exploration. *NBER Macroeconomics Annual 30*(1), 273–335.

Acemoglu, D., G. Anderson, D. Beede, C. Buffington, E. Childress, E. Dinlersoz, L. Foster, N. Goldschlag, J. Haltiwanger, Z. Kroff, P. Restrepo, and N. Zolas (2022). Automation and the Workforce: A Firm-Level View from the 2019 Annual Business Survey. Social Science Research Network. [Accessed 27 Nov. 2023].

Acemoglu, D., C. Lelarge, and P. Restrepo (2020). Competing with Robots: Firm-Level Evidence from France. *AEA Papers and Proceedings 110*, 383–388.

Acemoglu, D. and P. Restrepo (2018a). Modeling Automation. *AEA Papers and Proceedings 108*, 48–53.

Acemoglu, D. and P. Restrepo (2018b). The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review 108*(6), 1488–1542.

Acemoglu, D. and P. Restrepo (2019). Artificial Intelligence, Automation, and Work. National Bureau of Economic Research Working Paper.

Acemoglu, D. and P. Restrepo (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy 128*(6), 2188–2244.

Aghion, P., C. Antonin, S. Bunel, and M. Jaravel (2021). What Are the Labor and Product Market Effects of Automation? New Evidence from France. Harvard University.

Aghion, P., C. Antonin, S. Bunel, and M. Jaravel (2022). The Direct and Indirect Effects of Automation on Employment: A Survey of the Recent Literature. Harvard University. [Accessed 23 Sep. 2022].

Aghion, P., B. F. Jones, and C. I. Jones (2017). Artificial Intelligence and Economic Growth. SSRN.

Autor, D. H. and D. Dorn (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review 103*(5), 1553–1597.

Autor, D. H., L. F. Katz, and A. B. Krueger (1998). Computing Inequality: Have Computers Changed the Labor Market? *The Quarterly Journal of Economics 113*(4), 1169–1213.

Autor, D. H. and A. Salomons (2018). Is Automation Labor Share-Displacing? Productivity Growth, Employment, and the Labor Share. *Brookings Papers on Economic Activity 2018*(1), 1–87.

Baumol, W. J. (1967). Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis, journal = American Economic Review. *57*(3), 415–426.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt (2002). Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *The Quarterly Journal of Economics 117*(1), 339–376.

Cheng, H., R. Jia, D. Li, and H. Li (2019). The Rise of Robots in China. *Journal of Economic Perspectives 33*(2), 71–88.

Dauth, W., S. Findeisen, J. Suedekum, and N. Woessner (2021). The Adjustment of Labor Markets to Robots. *Journal of the European Economic Association 19*(6).

Foster, L., M. Klepper, J. Haltiwanger, and et al. (2018). Innovation, Productivity Dispersion, and Productivity Growth. National Bureau of Economic Research Working Paper w24420.

Gort, M. and S. Klepper (1982). Time Paths in the Diffusion of Product Innovations. *The Economic Journal 92*(367), 630.

Graetz, G. and G. Michaels (2018). Robots at Work. *The Review of Economics and Statistics 100*(5), 753–768.

Krueger, A. B. (1993). How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984–1989. *The Quarterly Journal of Economics 108*(1), 33–60.

Mann, K. and L. Püttmann (2017). Benign Effects of Automation: New Evidence from Patent Texts. SSRN. [Accessed 27 Nov. 2023].

Smith, A. and J. Anderson (2017). Automation in Everyday Life. Pew Research Center Internet, Science & Tech.

Timmer, M. P., E. Dietzenbacher, B. Los, R. Stehrer, and G. J. de Vries (2015). An Illustrated User Guide to the World Input–Output Database: The Case of Global Automotive Production. *Review of International Economics 23*, 575–605.

Webb, M. (2019). The Impact of Artificial Intelligence on the Labor Market. SSRN. [Accessed 27 Nov. 2023].

Òscar Jordà (2005). Estimation and Inference of Impulse Responses by Local Projections, journal = American Economic Review. *95*(1), 161–182.

A. Appendix: Further Visualisation of Social Network Analysis

A.1. List of Industries

Industry Code (nace_r2_code)	Broad Sector Grouping	Industry Name (nace_r2_name)
B	Mining, Utilities, and Construction	Mining and Quarrying
C10–C12	Manufacturing	Manufacture of food products; beverages and tobacco products
C13–C15	Manufacturing	Manufacture of textiles, wearing apparel, leather and related products
C16–C18	Manufacturing	Manufacture of wood, paper, printing, and reproduction
C19	Manufacturing	Manufacture of coke and refined petroleum products
C20	Manufacturing	Manufacture of chemicals and chemical products
C21	Manufacturing	Manufacture of basic pharmaceutical products and pharmaceutical preparations
C22–C23	Manufacturing	Manufacture of rubber and plastic products and other non-metallic mineral products
C24–C25	Manufacturing	Manufacture of basic metals and fabricated metal products, except machinery and equipment
C26–C27	Manufacturing	Computer, electronic, optical products; electrical equipment
C28	Manufacturing	Manufacture of machinery and equipment
C29–C30	Manufacturing	Manufacture of motor vehicles, trailers, semi-trailers and of other transport equipment
C31–C33	Manufacturing	Manufacture of furniture; jewelry, musical instruments, toys; repair and installation of machinery and equipment
D–E	Mining, Utilities, and Construction	Electricity, gas, steam; water supply, sewerage, waste management
F	Mining, Utilities, and Construction	Construction
G45	Low-Tech Services	Wholesale and retail trade and repair of motor vehicles and motorcycles
G46	Low-Tech Services	Wholesale trade, except of motor vehicles and motorcycles
G47	Low-Tech Services	Retail trade, except of motor vehicles and motorcycles
H	Low-Tech Services	Transportation and storage
I	Low-Tech Services	Accommodation and food service activities
J	High-Tech Services	Information and communication
K	High-Tech Services	Financial and insurance activities
L	High-Tech Services	Real estate activities
M–N	High-Tech Services	Professional, scientific, and technical activities; administrative and support service activities
P	Health and Education	Education
Q	Health and Education	Human health and social work activities
R–S	Low-Tech Services	Arts, entertainment, recreation; other services and service activities, etc.

A.2. **Variable List — EUKLEMS data**

Variable	Variable label
industry	Industry Code
country	Country Code
industry_name	Industry Name
country_name	Country Name
year	Year
sectorgroup	Broad Industry Grouping
ind_hier	Industry identifier variable
country_hier	Country identifier variable
cross	Group (industry × country)
COMP	Compensation of employees, current prices, millions of national currency
EMP	Number of persons employed, th.
EMPE	Number of employees, th.
GO_CP	Gross output, current prices, millions of national currency
GO_PI	Gross output, price indexes (2015)
GO_Q	Gross output, chained-link volumes (2015), millions of national currency
H_EMP	Total hours worked by persons engaged
H_EMPE	Total hours worked by employees
II_CP	Intermediate inputs, current prices, millions of national currency
II_PI	Intermediate inputs, price indexes (2015)
II_Q	Intermediate inputs, chained-link volumes (2015), millions of national currency
VA_CP	GROSS VALUE ADDED, current prices, millions of national currency
VA_PI	GROSS VALUE ADDED, price indexes (2015)
VA_Q	GROSS VALUE ADDED, chained-link volumes (2015), millions of national currency
CAP	Capital compensation, millions of national currency
LAB	Labor compensation, millions of national currency
VATFP_I	TFP index, 2015=100 – contributions to value added growth
bcycle	Business cycle indicators
lnEMP	Natural log of EMP
lnH_EMP	Natural log of H_EMP
lnVATFP_I	Natural log of VATFP_I
lnLAB	Natural log of LAB
lnVA_CP	Natural log of VA_CP
lnVA_Q	Natural log of VA_Q
hwwage	Mean hourly wage
lnhwage	Log of mean hourly wage
laborshare	Wage Bill as a share of value added
lnlaborshare	Log of Labor Share
ln_VA_PI	Natural Log of VA_PI
weight_emp_i	Industry employment share within countries, averaged over time
weight_h_i	Industry hours worked share within countries, averaged over time

weight_VA_i	Industry VALUE ADDED share within countries, averaged over time
weight_emp_c	Share of country's EMP in total EMP by year
weight_h_c	Share of country's hours worked in total hours worked by year
weight_VA_c	Share of country's VA in total VA by year
weight_emp_i_c	weight_emp_i × Share of country's EMP in total EMP by year
weight_h_i_c	weight_h_i × Share of country's H EMP by year
weight_VA_i_c	weight_VA_i × Share of country's VA in total VA by year
D1_lnhwage	100 × annual change in log of mean hourly wage
D1_lnLAB	100 × annual change in log total wage bill
D1_lnEMP	100 × annual change in log employment
D1_ln H_EMP	100 × annual change in log hours worked
D1_ln VATFP_I	100 × annual change in log value added based TFP by country
D1_ln_VA_CP	100 × annual change in log nominal VA
D1_ln_VA_Q	100 × annual change in log VA (volume)
D1_ln_VA_PI	100 × annual change in log VA (price indices)
D1_laborshare	Annual percentage point change in labor share
D1_lnlaborshare	100 × annual log change in labor share
D1_ln VATFP_OT	Country leave-out TFP growth
VA_CP_TOT	(sum) VA_CP
VA_PI_TOT	(sum) VA_PI
VA_PYP_TOT	(sum) VA_PYP
VA_Q_TOT	(sum) VA_Q
LOVA_CP_TOT	Value added (current prices) leaving out own-industry
lnLOVA_CP_TOT	log LOVA_CP_TOT
D1_lnVACP_LO	100 × annual log change in LOVA_CP_TOT
LOVA_Q_TOT	Value added (volume) leaving out own-industry
lnLOVA_Q_TOT	log LOVA_Q_TOT
D1_lnVAQ_LO	100 × annual log change in LOVA_Q_TOT
LOVA_PI_TOT	Value added (price indices) leaving out own-industry
lnLOVA_PI_TOT	log LOVA_PI_TOT
D1_lnVAPI_LO	100 × annual log change in LOVA_PI_TOT
ct	Group (country year)
ic	Group (industry country)
it	Group (industry year)

A.3. **Robustness Tests for Table 4 (Specification 1)**

	Employment	**Hours Worked**	**LaborShare**
Test 1: All countries given equal weight			
$\Sigma\Delta \ln \mathrm{TFP}_{i,c,t-k}$	-0.752^{**}	-0.378^{**}	-0.803^{***}
	(0.149)	(0.164)	(0.200)
No. of Observations	4,866	4,869	4,869
Test 2: Excluding Contemporaneous TFP Effects			
$\Sigma\Delta \ln \mathrm{TFP}_{i,c,t-k}$	-0.901^{***}	-0.859^{***}	-0.789^{**}
	(0.178)	(0.183)	(0.333)
No. of Observations	4,866	4,869	4,869
Test 3: Setting Negative TFP growth to zero			
$\Sigma\Delta \ln \mathrm{TFP}_{i,c,t-k}$	-0.982^{***}	-0.651^{***}	-0.658
	(0.216)	(0.239)	(0.408)
No. of Observations	4,866	4,869	4,869
Test 4: Excluding Self-Employed from Employed			
$\Sigma\Delta \ln \mathrm{TFP}_{i,c,t-k}$	-0.594^{***}	-0.522^{**}	-1.075^{***}
	(0.229)	(0.205)	(0.247)
No. of Observations	4,866	4,869	4,862
Fixed Effects			
Country	Yes	Yes	Yes
Year	Yes	Yes	Yes
Sector group	Yes	Yes	Yes
Country×Time Trend	No	No	No
Country×Business Cycle	No	No	No
Country×Year	Yes	Yes	Yes

A.4. **Test for Determining Lag Structure**

Ideal lag structure for regressions can be tested using Òscar Jordà (2005) local projection models.

$$\ln Y_{i,c,t+K} - \ln Y_{i,c,t-1} = \beta_0 + \beta_1\, \Delta\mathrm{TFP}_{i,c\neq c(i),\,t-1} + \sum_{k=0}^{K} \beta_2^k\, \Delta \ln \mathrm{TFP}_{i,c\neq c(i),\,k} + \beta_3\, \Delta \ln \mathrm{TFP}_{i,c\neq c(i),\,t-2} + \beta_4\, \Delta \ln Y_{i,c,t-2} + \alpha_{c,t} + \gamma_s + \varepsilon_{i,c,t} \quad (11)$$

Since the steady-state effects of contemporaneous TFP shocks will not manifest immediately, a lag structure is used to estimate the relationship between TFP shocks and labor outcomes.

$\ln Y_{i,c,t+K}$ denotes the outcome variable in industry i, country c, and year t, and K is the time horizon for the local projection. $\ln Y_{i,c,t+K} - \ln Y_{i,c,t-1}$ represents the change in the outcome variable from the base year (t–1) to year t+K.

Impulse variable is $\Delta\mathrm{TFP}_{i,c\neq c(i),\,t-1}$: log change in leave-out mean TFP in base year.

Lagged history of TFP growth and the outcome variable are used as controls:

$$\Delta \ln \mathrm{TFP}_{i,c\neq c(i),\,t-2}, \quad \Delta \ln Y_{i,c,\,t-2}.$$

The models also control for country–year and sector-group fixed effects: $\alpha_{c,t}$, γ_s.

The figure below reports local projection estimates and confidence intervals for the relationship between a TFP shock and industry-level changes in the outcome variables. For almost all outcome variables, there are small or negligible contemporaneous effects. Outcome variables of interest only undergo changes in ensuing years after the TFP shock. In almost all cases, the effects of the TFP shock do *not* plateau until 7–8 years after the shock. To ensure both completeness and utilization of as many observations as possible in the dataset, seven lags are included in the main specifications along with the contemporaneous effect.