NBER WORKING PAPER SERIES

TAX INCIDENCE ANOMALIES

Youssef Benzarti

Working Paper 32819
http://www.nber.org/papers/w32819

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2024

I am grateful to Jarkko Harju, Wojciech Kopczuk, Henrik Kleven, Emmanuel Saez, Alisa Tazhitdinova and Danny Yagan for helpful comments. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Tax Incidence Anomalies
Youssef Benzarti
NBER Working Paper No. 32819
August 2024
JEL No. H0,H22

ABSTRACT

This paper reviews the literature on the incidence of consumption and labor taxes and focuses on the empirical results that show stark departures from the canonical model of tax incidence, which I refer to as anomalies. In particular, there is mounting evidence questioning three fundamental implications of the canonical model: (1) that statutory incidence is irrelevant for economic incidence, (2) that the relative magnitude of the demand and supply elasticities is a sufficient statistic for tax incidence, and (3) that incidence is symmetric for increases and decreases. I review this empirical evidence and draw implications for the canonical model's relevance.

Youssef Benzarti
Department of Economics
University of California
Santa Barbara, CA 93106
and NBER
youssef.benzarti@gmail.com

1. Introduction

Tax incidence is the study of the effects of tax policy on market prices and on the economic welfare of individuals. Practically, it answers the following question: what happens to prices when the government introduces a tax or changes an existing one. It is a form of positive analysis and is fundamental to optimal tax policy. Because tax policy is at the heart of redistribution, any attempt at solving income or wealth inequality using taxation requires a robust understanding of tax incidence. Consider, for example, a tax on capital. Since capital is primarily owned by the top income and wealth individuals (Piketty and Saez 2003; Saez and Zucman 2016), a capital tax should be progressive. However, if firms are able to shift the burden of taxation to workers via lower wages or to consumers via higher prices, then capital taxes might actually be regressive. While such shifting effects might sound odd to non-tax-economists, in theory, they are plausible under plausible assumptions. Indeed, in seminal work, Harberger (1962) shows theoretically that, once general equilibrium effects are accounted for, workers can bear anything from less than zero to over 100% of the burden of capital taxes, depending on the relative magnitude of different elasticities.

Consequently, investigating the incidence of taxation is crucial for our understanding of the distributional effects of tax policy. The most recent review of the tax incidence literature dates back to more than two decades ago (see Fullerton and Metcalf 2002; Kotlikoff and Summers 1987). While it covered some empirical work, it was predominantly theoretical. The past two decades have seen a burgeoning of empirical tax incidence research. In this review, I focus on tax incidence anomalies – I discuss empirical studies that provide evidence of striking departures from the canonical model of tax incidence.

Why do I focus on anomalies rather than write a more encompassing review of the tax incidence literature? First, many – if not the majority – of empirical tax incidence results from the past two decades are anomalous, in that they deviate from the canonical model of tax incidence to varying degrees. Second, recent theoretical work by Saez and Zucman (2023) shows that anomalous tax incidence results are more relevant for welfare than standard ones. This finding, which relies on insights from the optimal tax literature, is discussed in Section 5.

To a certain extent, it might not be surprising that the empirical tax incidence literature mostly finds deviations from the canonical incidence model. Models are inherently incomplete and, therefore, better datasets and more compelling identification strategies are likely to find results that contradict them.

This is another reason why focusing on anomalies is important: they help us build more realistic tax incidence models. This is especially important because these models are widely used by policy makers, for example, for scoring tax reforms, and so using the most accurate, realistic and parsimonious model can be crucial for tax policy.[1]

Some tax incidence anomalies can be (and have been) incorporated into the canonical tax incidence model. For example, by adding tax salience or evasion effects, the canonical tax incidence model can explain the importance of the remitting party in determining economic incidence. Incorporating these anomalies helps make sense of the public debate and ensuing laws that legislate statutory tax incidence (see Keen and Slemrod 2021). These laws, if viewed from the perspective of the canonical model alone, would be puzzling, since statutory incidence is irrelevant under the canonical assumptions.

As anomalies build up, a new tax incidence paradigm might be needed. This is especially true if some of the anomalies being considered cannot be parsimoniously incorporated into the canonical tax incidence model. Indeed, some of the empirical evidence I discuss in this paper, such as the price hysteresis effects of changes in value-added taxes, would require adding so many degrees of freedom to the canonical model that it may make more sense to use an alternative one that could account for such stark departures without requiring ad-hoc patches.

This review is organized by type of tax and by type of anomaly. I focus on consumption taxes and labor taxes and ignore capital taxes because most of the recent empirical evidence on tax incidence comes from these two types of taxes and also because capital taxes are conceptually more complex. Readers interested in the incidence of capital taxes should refer to Auerbach (2006).

Overall, there are three types of anomalies that apply to both labor and consumption taxes. First, there is mounting evidence that questions the usefulness of using the relative magnitude of the supply and demand elasticities as a sufficient statistic for tax incidence. Second, there is empirical evidence that tax incidence is asymmetric: market prices (and quantities) respond differently to increases and decreases in taxes. Third, statutory incidence appears to matter for economic incidence, which is an empirical finding that contradicts a fundamental implication of the canonical tax incidence model.

I start by discussing the incidence of consumption taxes with a particular focus on the VAT. Because VAT rates are substantially higher than sales tax rates and the VAT base is wider than the sales tax base

[1]See for example, Tax Policy Center (2022), US Congressional Budget Office (2018), US Treasury, Office of Tax Analysis (2019) and US Joint Committee on Taxation (2019).

it is often easier to find better identifying variation for VATs than sales taxes.[2] The main takeaway from the past decade of research on the incidence of VATs is that most of the empirical evidence appears to be at odds with the canonical model of tax incidence. First, while the common wisdom is that the incidence of consumption taxes likely falls on consumers (see Fullerton and Metcalf (2002)), most of the modern evidence on the incidence of VAT cuts suggests otherwise (Kosonen (2015), Benzarti and Carloni (2019)). On the other hand, the incidence of VAT increases falls predominantly on consumers (Benzarti et al. (2020) and Benzarti, Garriga, and Tortarolo (2024a)). When put together, these two findings imply that consumption tax incidence is asymmetric. This finding could easily be reconciled with the canonical model by adding a degree of freedom and assuming that elasticities are different for increases and decreases in taxes (although it is unclear what could micro found such an assumption). The real challenge to the canonical model is that this asymmetry appears to be very persistent, with some of the evidence showing that it still exists a decade later. The second anomaly is that the relative magnitude of the demand and supply elasticities is not a sufficient statistic for tax incidence. Unless one is willing to add many degrees of freedom and assume different elasticities for different contexts such as tax increases versus tax decreases, firms that are part of chains versus independent ones (Harju, Kosonen, and Skans (2018) and Benzarti, Garriga, and Tortarolo (2024a)), contexts with and without evasion (Kopczuk et al. (2016)), etc., inputting off-the-shelf demand and supply elasticity estimates into the canonical model is unlikely to predict accurate estimates of tax incidence. The third anomaly is that there is empirical evidence confirming the common non-economist wisdom that statutory incidence matters for economic incidence, at least in certain contexts, such as when taxes are not salient (Chetty, Looney, and Kroft (2009)) or when tax evasion is prevalent (Kopczuk et al. (2016)).

Next, I cover recent empirical evidence on the incidence of labor taxes. Because most of the evidence comes from payroll tax variation, the focus is mostly on those taxes. Similarly to VATs, payroll taxes are well suited for quasi-experimental designs because they offer better variation (payroll tax exemptions on young workers, by cohort, past certain earning caps, etc.). These sources of variation are seldom available for other labor taxes. There is a strong consensus among public and labor economists that the incidence of payroll taxes is likely borne by workers. The basis for this common wisdom is that labor demand is considered to be orders of magnitude more elastic than labor supply since competitive firms

[2] In fact, most of the evidence on sales taxes comes from variation in the tax rate of sin goods or gasoline, which are usually subject to high tax rates, with the exception of Poterba (1996) and Besley and Rosen (1999).

are expected to pay workers their marginal product of labor and thus should pass on any payroll taxes to workers. However, recent empirical evidence from the past two decades contradicts this common wisdom by showing that firms may actually bear most of the employer portion of payroll taxes and shows substantial employment effects of payroll tax cuts.[3]

It is likely that fairness norms are causing the canonical model's failure: paying two workers who are similarly productive a different wage because they are subject to different payroll tax rates appears to violate such fairness norms and may depress morale, which could be more costly for firms than simply bearing the burden of payroll taxes. While these fairness norms can be parsimoniously incorporated into the canonical model of tax incidence, they substantially reduce its predictive power and render it practically unusable. Indeed, because norms are context specific, blindly applying demand and supply elasticity estimates from prior research to the canonical model in order to predict the incidence of payroll taxes in a new context is not informative anymore, at least without substantial sociological investigations of this new environment. For this reason, the relative magnitude of the labor demand and supply elasticities is unlikely to be sufficient in the case of payroll tax incidence.

If fairness norms are important, then the statutory incidence of payroll taxes may have real effects. Saez, Matsaganis, and Tsakloglou (2012) shows that employers adjust total wages so as to offset perceptions of inequality in posted wages (which are net of employer payroll taxes but inclusive of employee payroll taxes). This implies that the way the payroll tax is legally split between workers and firms eventually matters for its economic incidence.

Finally, payroll tax changes lead to asymmetric employment effects. Saez, Schoefer, and Seim (2021) shows that temporary payroll tax cuts can lead to higher employment, even after they are repealed. This asymmetry also helps rationalize the finding that payroll tax cuts have employment effects while minimum wage increases do not, even though they are both changes in total labor costs (only one is an increase and the other one a decrease).

To conclude this review of the empirical literature, I discuss how one could incorporate some of these anomalies in the canonical model, and the challenges that would entail. Two ingredients seem particularly important: (1) incorporating and, perhaps more importantly, estimating the general equilibrium effects of tax incidence, and, (2) incorporating dynamic effects/sluggishness. Finally, I

[3]One important tax incidence channel that has not been explored yet is whether consumers bear some of the incidence of payroll taxes. This is difficult because it requires linking price data to firm-level data.

discuss the welfare implications of these anomalies in light of the theoretical insights of Saez and Zucman (2023).

2. Canonical Tax Incidence Model

2.1. Model

Consider a partial equilibrium model with a single good and assume that the government levies a per-unit tax dt. Let p denote the pre-tax price, so that $p + dt$ represents the tax inclusive price. Denote by $S(p)$ and $D(p)$ the supply and demand functions, respectively. Assuming there is no tax, the equilibrium condition is given by $S(p) = D(p)$. If the tax is levied on consumers (as is typically the case for consumption taxes), producers will charge $p + dp$, while consumers will pay $p + dp + dt$, of which dt will be remitted to the government. If, on the other hand, the tax is levied on producers (as is the case for employer payroll taxes), the consumers will pay $p + dp + dt$ to the producers, but producers will keep $p + dp$ and remit dt to the government. In both cases, the equilibrium condition becomes $S(p + dp) = D(p + dp + dt)$, which can be rewritten as:

$$\frac{dp}{dt} = \frac{D'(p)}{S'(p) - D'(p)}. \tag{1}$$

Denote by $\epsilon_D = \frac{q}{D}\frac{dD}{dq}$ the price elasticity of demand and $\epsilon_S = \frac{p}{S}\frac{dS}{dp}$ the price elasticity of supply. Equation (1) can be rewritten as

$$\frac{dp}{dt} = \frac{\epsilon_D}{\epsilon_S - \epsilon_D}. \tag{2}$$

The model shows that the price response to taxes entirely depends on the relative magnitudes of the supply and demand elasticities.

The presented model can be easily modified to consider labor markets: in such a model, the wage response to a personal income, employee or employer payroll tax would depend on the elasticities of labor supply and labor demand.

2.2. Implications

This simple canonical tax incidence model makes three sharp and empirically testable predictions.

(1) The relative magnitudes of the supply and demand elasticities is a sufficient statistic for determining tax incidence. Equation (2) shows that the incidence split between consumers and producers is entirely dependent on the relative magnitude of the demand and supply elasticities. If supply is very elastic ($\epsilon_S = \infty$) and/or demand is very inelastic ($\epsilon_D = 0$) then $\frac{dp}{dt}$ is very small, hence the effect of the tax on producers' prices is negligible. This means that producers will bear very little of the tax, which will be entirely shifted to consumer prices. Empirically, we would observe the consumer price increasing by dt. Conversely, if supply is very inelastic and/or demand is very elastic, the incidence will be borne by firms. Empirically, the consumer price will decrease by dt and the producer price will remain constant. In intermediate cases, the incidence will be shared: the party that is relatively more inelastic will bear more of the incidence of the tax.

The common wisdom in public finance is that consumers and workers tend to be less elastic than firms. Thus, it is often believed that consumption taxes are likely to be borne by consumers (see Fullerton and Metcalf 2002). Similarly, economists believe that labor demand is more elastic than labor supply, implying that income and employee/employer payroll taxes are likely to be born by workers.

(2) Tax incidence should be symmetric for increases and decreases in taxes. Equation (2) is silent about the direction of tax changes. Therefore, tax increases and tax decreases should yield similar pass-through rates, as long as they are of equal sizes.

Note that the magnitude of tax changes may matter according to this equation, i.e., the pass-through rates of small tax changes might be different from that of large tax changes. In fact, Equation (2) is derived using small tax changes and so it may not be well-defined for large tax changes. This is an issue that is often overlooked but it may matter especially given that many of the recent empirical tax incidence papers use large tax changes for identification.

(3) Statutory incidence is irrelevant for economic incidence. Whether the tax is levied on demand or supply (or who collects and remits a tax) is irrelevant to its economic incidence according to this model. Imposing a tax on the supply or demand will lead to a response of market prices and quantities of equal size that depends solely on the relative magnitude of the supply and demand elasticities.

Researchers have considered several modifications of this simplest model, including relaxing the perfect competition assumptions (see Fullerton and Metcalf (2002)), or assuming that taxes are not

perfectly salient (Chetty, Looney, and Kroft 2009). To the best of my knowledge, none of these additions are able to fully explain some of the anomalies I describe next.

3. Consumption Taxes

Consumption taxes raise the most revenue around the world. This has become especially true with the quasi-universal adoption of the VAT. In this section, I discuss recent empirical evidence on consumption tax incidence, which has mostly focused on VATs. The majority of studies have found that VAT cuts are mostly borne by firms while, conversely, VAT increases are mostly borne by consumers. This leads to an asymmetry in the pass-through of VAT changes to prices. There is compelling evidence that this asymmetry is persistent and I discuss how this hysteresis might be a common feature of the way prices tend to respond to policy shocks. Second, I discuss evidence from sales and gasoline taxes that shows that statutory incidence matters for the economic incidence of consumption taxes when salience and evasion effects are prevalent. Finally, I argue that, because of so many departures from the canonical model, it is unlikely that the relative magnitude of the demand and supply elasticities is a sufficient statistic for consumption tax incidence.

3.1. Anomaly 1: Prices Respond Asymmetrically to Changes in VAT Rates

VAT Cuts. VAT cuts have become ubiquitous around the world, especially in the aftermath of the Covid-19 pandemic. They have been used to stimulate sluggish demand (after the Great Recession in the UK in 2009), support struggling industries (Hotel and Restaurant services in several OECD countries in 2020-2021) and dampen the effects of inflation.[4] These policy goals cannot be all consistent at the same time: either incidence is borne by consumers, in which case prices should fall, or it is borne by firms, in which case prices should remain stable. Of course, incidence could be shared and the policy goals might have some "leakage", and only partly benefit their intended recipients. Despite these expectations, the empirical literature on the incidence of VAT cuts mostly finds that they benefit firms (a notable exception being Gaarder (2018)).

[4]Here are some examples of countries that have recently cut the VAT rate on foodstuffs to 0%: Peru, Poland, Portugal, Spain, Bulgaria, North Macedonia, Lithuania, Cyprus, Uruguay, Fiji, Oman, and Togo. Bosnia cut its rate from 17% to 5%, Croatia from 13% to 5%, Latvia from 21% to 5%, Turkey from 8% to 1%, DR Congo from 16% to 8%, Costa Rica from 13% to 1%, Romania from 9% to 5%, and Greece from 24% to 13%. Italy, Germany, Belgium, Austria, Slovakia, Estonia, Angola, and the Netherlands are currently considering cutting the VAT rate on foodstuffs.

Kosonen (2015) estimates the effect of a 14 percentage point plausibly exogenous VAT cut implemented for the hairdresser industry in Finland in 2007 and compares it to a control group made of beauty salons to estimate its effect on prices.[5] Using this reform and a difference-in-differences approach, Kosonen (2015) estimates a pass-through rate of the VAT cut to prices of roughly 50%.

Benzarti and Carloni (2019) uses another VAT cut, which was also plausibly exogenous to the underlying economic conditions at the time: a 9 p.p. cut in the VAT rate applied to sit-down restaurants. The restaurant industry provides a richer setting than in Kosonen (2015) since hairdressers, in Finland, are mostly owner-operated. This allows a more detailed incidence analysis that goes beyond simply estimating price effects but also estimating the burden of the VAT cut on workers (wages and employment) and sellers of intermediate products. The goal of the policy was for the VAT cut to be split equally between firms, workers and consumers. Moreover the French Government urged firms to reinvest the extra profits into their restaurants. The actual impact was substantially different: Benzarti and Carloni (2019) finds that consumers benefited the least from the policy: the incidence on consumers was approximately 13%, firms captured approximately 55% of the VAT with no evidence of an increase in investments. The rest was split between workers and suppliers of intermediate inputs (who raised their prices).

At the time, Benzarti and Carloni (2019) and Kosonen (2015) provided evidence that was in stark contrast with the common wisdom in Public Finance that consumers bear most of the incidence of consumption taxes. However surprising, this finding could still be rationalized using the canonical tax incidence model and assuming that firms were relatively less elastic than consumers, at least in the case of restaurants and hairdressers.

Most of the evidence on the effect of VAT cuts on prices comes from EU member countries, with the exception of Benzarti, Garriga, and Tortarolo (2024a) which considers a 21 percentage point temporary VAT cut in Argentina. This setting is noteworthy because Argentina was experiencing high levels of inflation at the time ($\approx$ 50% annual inflation rate). While inflation is not explicitly modeled in the canonical tax incidence model, prices are likely to be more flexible in high-inflation regimes thus leading to higher pass-through rates. Consistently, Benzarti, Garriga, and Tortarolo (2024a) finds substantially higher pass-through rates to prices of the VAT cut but significantly short of full pass-through ($\approx$ 50%).

Finally, Benzarti et al. (2020) collects data on all VAT cuts (and VAT increases) that have occurred in

[5]This specific VAT cut was implemented because the European Commission was interested in experimenting with lower VAT rates in order to assess their effect on the economy, and it chose a few small sectors to experiment with. This implies that the reform was plausibly exogenous to the underlying economic conditions and their effects on prices.

the European Union from 1996 to 2015 and estimate their effect on prices. Consistent with the evidence discussed above, Benzarti et al. (2020) finds that VAT cuts had very limited effects on prices. The average pass-through rate to prices of the different VAT cuts considered varies between 5 and 10%. Note that the estimated pass-through rates in Benzarti et al. (2020) are substantially smaller than in other papers, both for VAT increases and decreases. This might be due to several factors: (1) attenuation bias due to measurement error caused by averaging over thousands of VAT changes; (2) focus on smaller VAT changes (2 to 3 percentage points on average); and, (3) the fact that most VAT changes affected several commodities at the same time because they were implemented on the general or reduced VAT rates, whereas many other studies exploited sector-specific reforms.

Overall, most of the empirical evidence on VAT cuts shows that their incidence is predominantly borne by firms and only partially benefits consumers, irrespective of the policy goals used to justify them.

VAT Increases. Most of the evidence on the pass-through of VAT increases comes from Benzarti et al. (2020), which collected data on all VAT increases in the European Union from 1996 to 2015 and estimated their effect on prices using a two-way fixed effects specification. Overall, the paper relies on more than 2,000 VAT changes across all commodities and EU-member states. The paper consistently finds that the pass-through of VAT increases to prices is 4 to 5 times larger than that of VAT cuts. This holds true for different sub-samples of the reforms being considered, including in times of growth and recessions, different sectors of the economy, sector-specific reforms and more general ones, etc. While the paper is not able to estimate the effect of these VAT changes on other outcomes besides prices, it zeros in on one specific reform for which administrative firm-level data is available. This data allows for a more comprehensive incidence analysis by using a large VAT cut on the hairdresser industry in Finland, the first leg of which was analyzed by Kosonen (2015). First, because the reform was plausibly exogenous to the underlying economic conditions, using it ensures that the asymmetric pass-through estimated using the rest of the EU VAT changes is not caused by changes in economic conditions. Second, the firm-level data allows to observe how the VAT changes affect the balance sheet of firms and, in particular, how they affect profits. Benzarti et al. (2020) finds that the VAT cut led to a large and sustained increase in firm profits and markups, but the VAT increase barely affected firms, since it was mostly passed through to prices and thus borne by consumers. Benzarti et al. (2020) also shows that this asymmetry is more

pronounced for firms with tighter profit margins.

Price Hysteresis. One important feature of the asymmetric pass-through of VATs to prices is that the asymmetry tends to persist over time and is still present several years after the VAT is reinstated to its original level. Benzarti et al. (2020) show that the repeal of the temporary VAT cut on hairdressers leads to price levels that are significantly higher than their pre-VAT cut counterfactual. This effect implies that the VAT cut-then-repeal policy resulted in higher equilibrium prices despite the tax rate reverting to its original pre-reform level. Moreover, this asymmetry persists for at least ten years. Figure A1 plots the price response to the VAT cut and its repeal in the control and treatment groups considered in Benzarti et al. (2020). It shows some convergence of the two price series four to five years after the repeal of the VAT cut, but a persistent price wedge of 3 to 5 percentage points remains after that and does not seem to be fading, more than 10 years after the repeal of the VAT cut. This is also reflected in the firm-level data: firm profits and markups are higher post-VAT-cut repeal relative to their pre-VAT cut levels and remain high for several years with no evidence of convergence towards symmetry. These findings imply that the government paid to increase profits through a temporary VAT cut, and consumers are subsequently paying for permanently higher firm profits via higher prices, once the VAT cut is repealed. Benzarti et al. (2020) estimates very similar patterns on prices when using the remaining EU VAT changes.

Relatedly, Benzarti, Garriga, and Tortarolo (2024a) finds similar hysteresis, in a different context (grocery stores in Argentina) that does not appear to be specific to VAT changes: prices exhibit hysteresis as a response to other policies as well. Benzarti, Garriga, and Tortarolo (2024a) analyze the effect of price controls and show that after these policies are repealed, prices for goods that were subject to price controls tend to remain at a lower level than those that were not. This hysteresis needs to be incorporated into the canonical incidence model and macroeconomic pricing models and accounted for by policy makers prior to implementing any (temporary) policies that may distort prices.

Evidence of Symmetry. A few papers show that the pass-through of consumption taxes to prices can be symmetric as well. Doyle and Samphantharak (2008) estimate the pass-through of a moratorium on the 5% sales tax on gasoline in Illinois and Indiana in 2000. They find that the cut and its repeal are almost fully passed through to prices. There are several reasons why prices may be more likely to respond symmetrically in this setting. First, the effects might be different for gasoline taxes and VATs. Second,

these cuts were passed as part of an election promise and so gasoline retailers may have been under strong political pressure to adjust prices. Third, gas prices might be much more flexible than other prices. Similarly Gaarder (2018) estimates the pass-through of a temporary VAT cut to prices in Norway and finds compelling evidence that both the VAT cut and its repeal are almost fully passed through to prices. These differences further highlight the shortcomings of the standard incidence model which often ignores any institutional specificities (such as political pressure) and instead focuses solely on the relative magnitude of the elasticities of demand and supply.

Incorporating Insights From Monetary Economics?. Ultimately, the study of tax incidence is really the study of prices. Another area of economics that is concerned with market prices is Monetary Economics (see Nakamura and Steinsson (2013) for a review of that literature). One important difference in the way Public Economists and Macro Economists model prices is in incorporating dynamics: the canonical tax incidence model features immediate adjustments, whereas workhorse macro models (menu cost models, Calvo pricing models) are sluggish.

The asymmetric pass-through and resulting price hysteresis are intrinsically a dynamic feature. And so, perhaps, it is unsurprising that it would not fit the canonical tax incidence model, which features immediate adjustments. One possible avenue for explaining these would be to incorporate insights from macro pricing models. In particular, adding pricing frictions in the form of menu costs seems like a natural extension of the canonical model. Benzarti, Garriga, and Tortarolo (2024b) shows that such menu costs can indeed lead to some degree of asymmetric pass through but are unlikely to lead to medium- or long-run price hysteresis. Future research needs to: (1) empirically investigate whether asymmetric pass through and hysteresis is a defining feature of prices and in what settings it is more pronounced/more likely to hold; and, (2) new theoretical insights need to be incorporated in the standard tax incidence (and macro pricing) models to account for this feature. Some candidates are models that assume that competing firms may fail to coordinate on prices (see Cooper and John (1988)).

3.2. Anomaly 2: Does Statutory Incidence Matter?

An important implication of the standard incidence model introduced in Section 2 is the so-called tax neutrality result: statutory incidence (who legally pays for/remits the tax) is irrelevant for economic incidence (who bears the economic burden of the tax as a result of market price changes). Despite the

clarity of this theoretical prediction, policymakers and the general public appear to believe otherwise given how much importance they give to statutory incidence. For example, several US States legislate who is legally liable to pay for sales taxes (Keen and Slemrod (2021)). Similarly, governments often announce how a specific tax cut should be split, as was the case for the 2009 French VAT cut on sit down restaurants analyzed in Benzarti and Carloni (2019), where then French President Sarkozy instructed restaurants to equally split the windfall from the VAT cut among firm owners, workers and consumers.

There are very few direct tests of whether statutory incidence matters in the case of consumption taxes. Testing this effect using a quasi-experimental design would require additional variation relative to the standard empirical tax incidence analysis, which is usually unavailable.

One notable exception is Kopczuk et al. (2016), which provides evidence that statutory incidence affects prices by showing that the pass through of US state fuel diesel taxes to consumer prices is smaller when the point of tax collection is at the retailer level rather than upstream in the supply chain. The most likely explanation for this finding is that evasion is more prevalent at the retailer level. Kopczuk et al. (2016) augment the canonical tax incidence model by adding evasion considerations and show that higher levels of tax evasion predict lower levels of pass-through rates to consumers. Intuitively, a remitting party that does not remit a tax will pass through less of it to prices than a remitting party that fully remits it.

Chetty, Looney, and Kroft (2009) provides indirect evidence that statutory incidence likely matters for economic incidence in the case of sales taxes. Using a randomized control trial (RCT) and observational data, Chetty, Looney, and Kroft (2009) shows that consumers tend to under-react to sales taxes, by buying more, when they are not included in prices. Chetty, Looney, and Kroft (2009) theoretically derive the implications of this finding for tax incidence and shows that it would cause statutory incidence to matter for economic incidence. Intuitively, if consumers under-react to sales taxes, then producers face less pressure to reduce the consumer price to compensate them for the existing tax. But this theoretical prediction has not yet been tested empirically. Testing it is difficult because it requires an exogenous change in salience and in tax rates. Whether this issue is relevant for Value-Added Taxes is an open question, but since VATs are included in prices, presumably they should be salient (although VAT *changes* may not be).

3.3. Anomaly 3: Elasticities are Not a Sufficient Statistic

An important implication of the canonical tax incidence model is that it is sufficient (a la Chetty (2009)) to know the relative magnitude of demand and supply elasticities in order to derive tax incidence. This is a powerful implication of the canonical model, since it allows researchers and policy makers to input elasticity estimates from different contexts. However, the evidence covered in Sections 3.1 and 3.2 show that tax incidence can be highly context dependent. Indeed, we saw how prices respond differently to increases and decreases in VAT rates as well as how salience and evasion affect incidence.

In this section, I discuss an additional form of context-dependence: tax incidence appears to be sensitive to whether a firm is part of a chain or is independent. Harju, Kosonen, and Skans (2018) estimates the effect of a large VAT cut on restaurants in Finland (9 p.p.) and in Sweden (13 p.p.) and finds dramatically different levels of pass-through rates in chain than in independent restaurants. Chains tend to pass-through 100% of the VAT cut immediately after it is implemented while independent restaurants pass-through 0% of it. These patterns change after a few months: chain restaurants slowly adjust prices to converge towards smaller levels of pass-through rates, while independent restaurants keep pass-through rates at similar levels of 0%.

Benzarti, Garriga, and Tortarolo (2024a) finds similar effects for chain and independent supermarkets in Argentina. The authors estimate that the pass-through rate of a 21 p.p. VAT cut on groceries is approximately 80% in chain supermarket, and 30% in independent stores. While the difference between store types are not as dramatic as in Harju, Kosonen, and Skans (2018), there is no evidence of convergence even after a period of five months. On the opposite, it appears as if independent supermarkets further trend towards 0% pass-through in later periods. While these findings are novel in Public Finance, differential pricing strategies by different store types have been explored in Industrial Organization more extensively. DellaVigna and Gentzkow (2019), for example, provides evidence that national and regional chains tend to respond to demand shocks differently. The authors that such differences are likely driven by access to different sources of information regarding local demand shocks. Such insights should be incorporated, to the extent possible, in our canonical incidence model.

In light of this evidence and the one discussed in the previous sections, one would need to estimate a four by four matrix of elasticities (direction of tax change, salience, evasion and type of firm) in order to predict tax incidence using the canonical model. As more evidence of the sensitivity of the canonical

model to different contexts emerges, this elasticity matrix may become too large to be practical.

4. Labor Taxes

Labor taxes are taxes imposed on the labor input of a firm's production function. The payroll tax is one example of such a tax. Similarly the portion of the income tax imposed on labor earnings can also be considered a labor tax. For the purposes of this section, I focus mostly on the payroll tax both because it is a "clean" example of a labor tax and also because it is the labor tax for which we have the most empirical evidence.[6]

The common wisdom in Public Finance is that the incidence of payroll taxes predominantly falls on workers (Fullerton and Metcalf 2002). The primary reason being that, in a competitive market, firms pay workers their tax-adjusted marginal products of labor. As a result, under the canonical model outlined in Section 2, most of the burden of payroll taxes is borne by labor supply, and any changes to labor taxes must be offset with one-to-one changes in net-of-tax wages.

The earlier empirical tax incidence literature found evidence that was broadly consistent with workers bearing most of the burden of payroll taxes, thus reinforcing the common wisdom (Gruber 1994, 1997). However, empirical evidence from the last decade, which employs quasi-experimental research designs and administrative data, has consistently found evidence that deviates from the canonical incidence model. In this section, I review the main anomalies researchers have uncovered and discuss the attempts to incorporate them into the canonical incidence model.

4.1. Anomaly 1: the Relative Magnitude of Elasticities is Not a Sufficient Statistic

This anomaly is a consequence of two findings: (1) fairness considerations restrict wage adjustments thus limiting the predictive power of the relative magnitude of the labor demand and supply elasticities; and, (2) there is evidence showing that the labor demand elasticity might be endogenous to changes in the payroll taxes. I discuss both in this section.

[6]There is very little research on the incidence of income taxes, perhaps because finding a compelling control group can be challenging. One notable exception is Lehmann, Marical, and Rioux (2013).

4.1.1. Fairness Effects Matter More Than Elasticities

Saez, Schoefer, and Seim (2019) studies the consequences of a 16 percentage point cut (from 31% to 15%) of the employer-portion of the payroll tax in Sweden applied to employees younger than 26. The prediction of the canonical model is straightforward: because employees that are slightly older than 26 should be similar to the employees that are slightly younger, their marginal products of labor should be the same. Furthermore, since employers only care about the tax-inclusive wages (wages inclusive of all payroll taxes) and not about the posted wages (which are net of employer payroll taxes but inclusive of employee payroll taxes), a payroll tax cut should result in a decrease of the tax-inclusive wage. This would lead firms to raise posted wages until the tax-inclusive wages are equalized for similar workers, resulting in an increase in the posted and net-of-tax wages of the under-26 workers. In other words, the tax cut would be borne by workers, as is commonly believed.

Instead, Saez, Schoefer, and Seim (2019) shows stark departures from these predictions. First, the authors find that net-of-tax wages do not respond to the payroll tax cut, and, conversely, total wages decrease substantially. Therefore, at first glance, the payroll tax cut is fully borne by firms, since it results in lower labor costs and does not lead to higher wages. Second, they estimate substantial employment effects, amounting to 2 to 3 percentage point employment increases for workers younger than 26.[7] These effects appear to be mostly driven by fewer separations rather than more new hires.

Interpreting these results through the lens of the canonical framework, one would conclude that labor supply is much more elastic than labor demand. In other words, workers are very sensitive to changes in wages, which greatly affects the numbers of hours they want to work and their decision to work altogether. Conversely, firms are not very concerned with wages in their decision to hire workers.

Instead, Saez, Schoefer, and Seim (2019) offers an alternative incidence model, one where firms are subject to a wage equity constraint within firms. Because of this constraint, firms cannot pay a lower wage to younger workers, even though their marginal product of labor is lower than that of workers who are significantly older (perhaps because younger workers lack experience or on-the-job training).[8] This leads to an inefficiently low level of hiring of young workers. When the payroll tax is cut, the

[7] Using variation in payroll taxes caused by unemployment insurance variation across U.S. states, Guo (2023) shows qualitatively similar employment responses to payroll taxes.

[8] While the tax variation used for identification assumes that workers slightly younger than twenty six are similar to workers who are slightly older, the wage equity constraints apply to workers with age differences of a substantially larger magnitude.

wedge between labor costs and the marginal product of labor due to the wage equity constraints shrinks, resulting in a reduction in the inefficient unemployment of younger workers. One key implication of this model is that the effect of targeted payroll tax cuts should spillover onto workers who were not originally targeted by them via collective incidence effects. Saez, Schoefer, and Seim (2019) finds evidence consistent with this prediction: the payroll tax cut led to an increase in wages across all workers in firms that were more intensely treated by this cut (because they had a higher share of younger workers).

Saez, Schoefer, and Seim (2019) offers a radical departure from the canonical model of tax incidence, which solely relies on the relative magnitudes of the supply and demand elasticities, by showing that the canonical model's wage prediction power is severely restricted by fairness constraints. [9] If this fairness mechanism indeed explains these findings, then the relative magnitude of the demand and supply elasticities is not a sufficient statistic for tax incidence since fairness effects dominate. Realistically, one would need to incorporate insights from sociology on fairness and when and how it applies into the canonical model, in order to understand and predict the incidence of payroll taxes.

4.1.2. Labor Demand Elasticities Might Be Endogenous to Payroll Taxes

Benzarti and Harju (2021b) analyze a discontinuity in the employer portion of the payroll tax in Finland: firms with capital depreciation levels in excess of a certain threshold are subject to a higher employer payroll tax rate than firms whose capital depreciation remains below the threshold.[10] Benzarti and Harju (2021b) shows that the payroll tax discontinuity was mostly borne by firms and resulted in significant employment effects with negligible impacts on worker earnings. Furthermore, the study shows that payroll tax changes affected the production function of firms, leading them to substitute low-skilled workers with high-skilled ones and lowering overall investments. In contrast to Saez, Schoefer, and Seim (2019), this payroll tax change affected all employees in the treated firms rather than a subgroup of them, thus shutting down the equity channel discussed earlier.

The findings from Benzarti and Harju (2021b) have two implications. First, even in the absence of fairness concerns, the employer portion of payroll taxes is borne by firms.[11] This implies that labor

[9] In addition, Saez, Schoefer, and Seim (2019) shows that payroll tax cuts can be used as a policy tool to stimulate firm activity since they result in a slew of firm level effects such as increases in capital, sales, value added and profits (in addition to the aforementioned employment effects).

[10] In recent work, Lobel (2024) finds similar results to Benzarti and Harju (2021b) and shows that they can be explained by market power.

[11] Benzarti and Harju (2021a) uses yet another form for payroll tax variation: a cut in the employer portion of the payroll

demand is not as elastic as common wisdom suggests, or that other factors may be at play. Second, the labor demand elasticity is endogenous to changes in payroll taxes due to firms adjusting their production functions in response to payroll tax changes. This means that the labor demand elasticity is not a sufficient statistic (in the spirit of Chetty 2009). This latter fact highlights the limitation of the partial equilibrium framework of the canonical model. Since labor demand is an input in production and modern labor taxes are high, firms may find it optimal to adjust production functions in response to labor tax changes, thus necessitating a more dynamic incidence model, a la Harberger (1962). Both (1) and (2) therefore raise serious concerns with using the canonical tax incidence model of Section 2 to predict the incidence of payroll tax changes.

4.2. Anomaly 2: Statutory Incidence Matters

Most of the recent evidence on the incidence of payroll taxes relies on variation of the employer portion of the payroll tax. According to the canonical model of tax incidence, and its tax neutrality result, one could simply take the estimates for the employer portion of the payroll tax and apply them to the employee portion. In light of evidence discussed in Section 4.1, we would expect the incidence of the employee portion of the payroll tax to fall on firms as well.

The results from Saez, Matsaganis, and Tsakloglou (2012) imply otherwise. This paper uses a cohort-based discontinuity in the payroll tax cap (above which the payroll tax rate is zero) in Greece and estimates its effect on tax-inclusive wages, posted wages and net-of-payroll-tax wages. The reform affected individuals who entered the labor force after a certain cutoff date. As a result of the reform such employees faced permanently higher payroll tax caps, thus leading to an increase in their marginal payroll tax rate on those earnings between the old and the new cap. Using a regression discontinuity design, Saez, Matsaganis, and Tsakloglou (2012) compare individuals who entered the labor force just before and just after the cutoff date, and find that employers compensated above-the-cutoff workers for the increase in the *employer* portion of the payroll tax only but not for the increase in the employee portion. In other words, employers adjusted wages so as to keep posted wages (which are net of employer payroll taxes but inclusive of employee payroll taxes) similar across workers above and below the cutoff date. As a consequence, tax-inclusive wages increased while net-of-tax wages decreased.

tax targeted at all firms but only in certain regions of Finland. The empirical results are broadly consistent with Benzarti and Harju (2021b): firms that are treated by the payroll tax cut tend to fare better, especially during periods of recession, suggesting that the payroll tax cut mostly benefited firms.

The likely explanation for this finding is that economic incidence is constrained by fairness considerations, similarly to Saez, Schoefer, and Seim (2019): employers strive to ensure fairness in posted wages and so compensate new cohorts of workers just enough to satisfy this social norm. Why do employers target posted wages rather tax-inclusive or net-of-tax wages? Perhaps, because posted wages are most salient and the easiest to observe in the labor markets.

One important policy implication of the findings from Saez, Matsaganis, and Tsakloglou (2012) is that changing the statutory incidence of payroll taxes may have important equilibrium effects, even when payroll tax rates or caps remain unchanged. Note that while some taxes and benefits are administered directly by the government (for example, the Earned Income Tax Credit and other tax credits in the U.S. are administered via tax returns), others are administered via employers (for example, employee payroll taxes and a large share of incomes taxes in the U.S. are collected via payroll). What would happen if employers acted as intermediaries between the government and tax credit recipients, paying the tax credits directly with paychecks?[12] Garriga and Tortarolo (2021) tackle this question empirically by using a change in the way a child tax credit is disbursed to workers in Argentina. Prior to 2003, employers acted as intermediaries between the government and workers, disbursing the credit together with paychecks. After 2003, the government started sending the credit directly to the bank account of recipients. If the tax neutrality result holds, this reform should have had no effect on wages. Instead, Garriga and Tortarolo (2021) shows that firms were actually capturing 6 to 14% of the credit via lower wages when they were disbursing it. A possible explanation for this finding is that employers were "bundling" the credit with wages, thus artificially inflating the earnings workers would receive to make the hiring package appear more attractive. This interpretation once again highlights the importance of salience for tax incidence, already discussed in Section 3.2 for consumption taxes: if taxes (or credits) are not salient, then who remits them matters for economic incidence.[13]

4.3. Anomaly 3: Is the Effect of Payroll Taxes Asymmetric?

Due to the lack of available tax variation, the payroll tax incidence literature has mostly focused on payroll tax *decreases*. In contrast, Saez, Schoefer, and Seim (2021) estimate the consequences of a tax *increase* – the repeal of the payroll tax cut they had analyzed in Saez, Schoefer, and Seim (2019). They use

[12] This was the spirit of the Advance EITC, which was established in 1975, and later repealed in 2010 (see Jones 2010).
[13] Relatedly, see Rothstein (2010) for evidence on the incidence of the EITC.

two sources of variation: (1) the repeal of the payroll tax cut which affected all workers younger than 26, (2) the natural aging out of the payroll tax cut, which happens to workers as they age. Both of these sources of variation show limited (dis-)employment effects of an increase in payroll taxes. These results contrast the findings of Saez, Schoefer, and Seim (2019) discussed in Section 4.1.1, and suggest that the employment effects of payroll tax increases and decreases are likely to be asymmetric. If this result is confirmed with more direct tests, then this would mean that the canonical model of tax incidence fails to predict the incidence of labor taxes in yet another way.[14]

The presence of such asymmetric responses would allow us to reconcile the facts that payroll tax changes have been shown to have substantial employment effects, while minimum wage changes have been shown to seemingly have no employment effects (see Cengiz et al. 2019). Note that both payroll tax changes (especially those affecting the employer portion) and minimum wage changes lead to changes of employer labor costs, given that the incidence of (the employer-portion) of payroll taxes appears to fall on firms. Therefore, we would expect them to have similar employment effects, yet the effects diverge.[15] The asymmetry provides a plausible explanation: the payroll tax literature has mostly focused on payroll tax *decreases* (which result in more employment via fewer separations) whereas the minimum wage literature typically studied minimum wage *increases* (which have limited employment effects).[16]

4.4. Incorporating (Perceptions of) Benefit Linkage

For some payroll taxes, there is a direct link between the taxes paid and the benefits received by worker. In the U.S., for example, there is a formula that determines how much pension a given worker will receive given their past Social Security contributions. This benefit linkage, however, is less clear in other cases: some contributions may simply be assigned to a general fund that is not tied to a particular benefit.

In theory, there are strong reasons to believe that benefit linkage and the value of those benefits should matter for tax incidence (Musgrave and Musgrave 1959; Summers 1989). Intuitively, if taxpayers

[14]Note that these asymmetric effects in employment might not be related to the to asymmetric responses of prices to consumption taxes because the payroll tax increase does not lead to a fall in posted wages (see Figure 1 of Saez, Schoefer, and Seim (2021)).

[15]One difference between the two is that minimum wages affect the lowest earners, whereas payroll taxes affect all employees. Could that be the reason why the employment effects are different? Probably not: the payroll tax cut in Saez, Schoefer, and Seim (2019) only affected workers younger than 26, including many of the lowest earners.

[16]The explanation put forward by Saez, Schoefer, and Seim (2021) for the null employment effect is that employers exposed to younger workers as a result of the payroll tax cut, questioned their pre-conceived notions about younger workers and reduced discrimination against them, thus increasing younger workers' employment even after the payroll tax cut is repealed.

value certain benefits of taxation, it is as if those benefits raise their net wages and thus lead to a smaller shift in the labor supply curve caused by the payroll tax. In the most extreme case, if workers value the benefit of a one dollar payroll tax at one dollar, the effect of this tax on their labor supply is fully offset as if there was no tax. In this case, workers bear the full burden of the tax.

Bozio, Breda, and Grenet (2019) provide compelling evidence that such effects matter in the context of payroll taxation in France. The authors show that workers bear the full burden of payroll tax changes with strong benefit linkage, while they bear very little of this burden for taxes with weak benefit linkage. They also provide a meta analysis of the payroll tax incidence literature, showing that studies that rely on payroll tax variation with strong benefit linkage tend to find substantial pass-through of payroll taxes to wages, while those with weak benefit linkage find limited wage effects.

It is arguable whether this is actually an anomaly or simply a missing feature in the canonical model. Incorporating such benefit-linkage effects in a theoretical framework is relatively straightforward (Summers 1989), but empirically estimating how much taxpayers value these benefits is challenging but important because of its strong effect on tax incidence.

4.5. Capital and Labor Shares

On the one hand, there is broad consensus that the capital to labor ratio of national income is similar across countries, (see Piketty 2014). On the other hand, there is large variation in payroll tax rates, payroll ceilings and the statutory incidence of payroll taxes across different countries in the world. France, for example, imposes a total payroll tax rate of approximately 40%, while it is approximately 14% in the US. Moreover, there is growing evidence showing that firms bear a substantial portion of the burden of (the employer portion of) payroll taxes. Assuming this finding is true then we should expect the labor and capital shares of national income to be affected by this wide variation in payroll tax rates. In other words, if payroll taxes are borne by firms, in countries with high payroll taxes, we should expect higher labor shares, in spite of the evidence that they appear to be constant across countries.

How can we make sense of this apparent contradiction? An obvious (but very hard to test) candidate explanation is that there are general equilibrium effects at play that are strong enough to cancel out the partial equilibrium effects of payroll taxes on labor shares. Because the studies I have discussed predominantly use a quasi-experimental approach, which is inherently unable to capture general

equilibrium effects, a new empirical paradigm might be needed to test this explanation and explain this paradox.

4.6. Implications for Estimating Labor Supply Elasticities Using Tax Variation

There is a large body of research (partially reviewed by Chetty 2012) that estimates labor supply elasticities using tax variation. The typical approach employs some form of labor tax variation (changes in taxes for certain workers but not others, or kink/notch points in the tax schedule) in order to estimate labor responses. Labor supply elasticities are then estimated as the ratio of the estimated quantity response over the tax change being considered. By assigning the entirety of the observed labor response to labor supply, these studies implicitly assume that the incidence of labor taxes is fully borne by workers.

While this approach made sense under the common wisdom that labor supply bears most of the incidence of labor taxes, evidence discussed in this section shows that it is not always the case, and the estimated quantity responses are in fact a mix of demand and supply responses. In the extreme case, when the incidence is fully borne by firms, as in Saez, Schoefer, and Seim (2019) and Benzarti and Harju (2021b), net-of-tax wages should not change, workers should not respond, and consequently the entirety of the labor response should be attributed to labor demand. Some papers in the literature have recognized the dual response of supply and demand and have provided evidence of "firm responses" (Chetty et al. 2011; Tazhitdinova 2020; Gudgeon and Trenkle 2024) confounding labor supply responses.

Therefore, to obtain structural labor supply elasticity estimates using tax variation, it is important to estimate incidence as well. If doing so is not feasible, researchers should clearly state the assumptions made about the incidence of the tax change they are studying.

5. Fixing the Canonical Model

The canonical model of tax incidence is powerful because of its simplicity, tight connection to empirical parameters (elasticities) and ability to precisely predict incidence. In this section, I categorize the evidence covered above into anomalies that are relatively easy to incorporate into the canonical model and those anomalies that do not easily fit into its framework and would require a more substantial overhaul.

Most anomalies are relatively easy to incorporate into the canonical model by adding one or two

degrees of freedom to it. In fact, many of the papers covered above already incorporate these into their models: such as in the case of evasion (Kopczuk et al. 2016), salience (Chetty, Looney, and Kroft 2009), benefit-linkage (Bozio, Breda, and Grenet 2019) and asymmetric employment effects for payroll tax changes (Saez, Schoefer, and Seim 2021).

While these anomalies are easy to incorporate, they require additional degrees of freedom, which usually entail estimating additional elasticities. As discussed in Section 3.3, many of these anomalies apply across contexts and so using the canonical model to predict incidence would require a large matrix of elasticity estimates, which may become quickly impractical.

This issue is particularly problematic when it comes to fairness and its effect on the incidence of payroll taxes. In spite of a large labor economics literature estimating labor supply and demand elasticities, some of the research I covered in this review makes it clear that we cannot simply input these elasticity estimates into the canonical model to predict the incidence of payroll taxes, because fairness effects seem to dominate market effects. Fairness considerations are relatively easy to incorporate into the canonical model, in fact Saez, Schoefer, and Seim (2019) offers one way of doing so. However, a canonical model that incorporates fairness, while more precise and better suited for explaining ex-post results, is not very practical for predicting where incidence will fall, since fairness norms are highly context dependent.

There are two general additional issues that are hard to fit into the canonical model, at least in its current form: (1) general equilibrium effects, and, (2) issues caused by dynamic effects. The first issue might seem obvious: the canonical model is a partial equilibrium model, and so it is unsurprising that it would not account for general equilibrium effects. It is also a very common issue in applied micro research, which, in its modern form, often relies on quasi-experimental approaches that are not well-suited for estimating general equilibrium effects. Can we simply ignore these effects and focus on partial equilibrium effects, especially given that there is no empirical evidence that they matter for tax incidence? This lack of evidence is mostly due to the fact that we are not well-equipped for detecting them. And while there may be good theoretical reasons for why general equilibrium effects may not matter in other applied micro fields, it is not the case for tax incidence: Harberger (1962) shows that there are compelling theoretical reasons why general equilibrium effects can fully shift incidence from one party to another. The main issue here is perhaps not one of modelling, since tax incidence models with general equilibrium effects already exists, but one of empirics: estimating such effects would require a

new empirical paradigm.

The second general issue with the canonical model is that it features immediate adjustments rather than being more sluggish, despite the mounting evidence that dynamics matter for tax incidence. In fact, price dynamics have been extensively investigated by macro economists and incorporating some of their insights and findings might be a reasonable way forward (see Nakamura and Steinsson (2013)). It is also the case that, because empirical tax incidence research mostly relies on quasi-experimental methods and high-quality data, new empirical findings on price dynamics have been uncovered by tax-incidence economists that can be useful to discipline macro price models. While this is a less daunting endeavour than estimating general equilibrium effects, incorporating dynamics/sluggishness is perhaps more pressing because of its sizeable effects on tax incidence.

Overall, anomalies are important for improving (or replacing) the canonical model of tax incidence. Recent research shows that they might also be crucial for welfare. Saez and Zucman (2023) shows that classical tax incidence analysis a la Harberger (Harberger 1962, 1964)) is not normatively relevant. Intuitively, any tax incidence effects can be easily offset by readjusting the tax system at no cost.[17] For example, if the economic incidence of a corporate tax is borne by workers (via a reduction in wages) then the rental rate of capital should be higher. By taxing these extra profits, and redistributing the revenue to workers, the incidence effect of the corporate tax on labor could be fully offset. However, as Saez and Zucman (2023) emphasize, if tax incidence is non-classical as discussed in this survey, tax incidence cannot be offset by readjusting the tax system, and therefore tax incidence becomes normatively relevant, and has to be taken into account when thinking about optimal tax policy.

[17]This idea is similar to Atkinson and Stiglitz (1976)'s insight (applied to consumption taxes): adding progressivity to a consumption tax is not needed since their regressive effects can be corrected using a non-linear income tax.

References

Atkinson, Anthony Barnes, and Joseph E Stiglitz. 1976. "The design of tax structure: direct versus indirect taxation." *Journal of public Economics* 6 (1-2): 55–75.

Auerbach, Alan J. 2006. "Who bears the corporate tax? A review of what we know." *Tax policy and the economy* 20: 1–40.

Benzarti, Youssef, and Dorian Carloni. 2019. "Who Really Benefits From Consumption Tax Cuts? Evidence From a Large VAT Reform in France." *American Economic Journal: Economic Policy* 11 (1): 38–63.

Benzarti, Youssef, Dorian Carloni, Jarkko Harju, and Tuomas Kosonen. 2020. "What goes up may not come down: asymmetric incidence of value-added taxes." *Journal of Political Economy* 128 (12): 4438–4474.

Benzarti, Youssef, Santiago Garriga, and Darío Tortarolo. 2024a. "Can VAT Cuts and Anti-Profiteering Measures Dampen the Effects of Food Price Inflation?" *NBER Working Paper*.

Benzarti, Youssef, Santiago Garriga, and Dario Tortarolo. 2024b. "Using VAT Changes to Test Macro Price Models." *Working Paper*.

Benzarti, Youssef, and Jarkko Harju. 2021a. "Can payroll tax cuts help firms during recessions?" *Journal of Public Economics* 200: 104472.

Benzarti, Youssef, and Jarkko Harju. 2021b. "Using payroll tax variation to unpack the black box of firm-level production." *Journal of the European Economic Association* 19 (5): 2737–2764.

Besley, Timothy J, and Harvey S Rosen. 1999. "Sales Taxes and Prices: An Empirical Analysis." *National Tax Journal*: 157–178.

Bozio, Antoine, Thomas Breda, and Julien Grenet. 2019. "Does Tax-Benefit Linkage Matter for the Incidence of Social Security Contributions?" *Working Paper*.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The effect of minimum wages on low-wage jobs." *The Quarterly Journal of Economics* 134 (3): 1405–1454.

Chetty, Raj. 2009. "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods." *Annu. Rev. Econ.* 1 (1): 451–488.

Chetty, Raj. 2012. "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply." *Econometrica* 80 (3): 969–1018.

Chetty, Raj, John N Friedman, Tore Olsen, and Luigi Pistaferri. 2011. "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records." *The quarterly journal of economics* 126 (2): 749–804.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *The American Economic Review* 99 (4): 1145–1177.

Cooper, Russell, and Andrew John. 1988. "Coordinating coordination failures in Keynesian models." *The Quarterly Journal of Economics* 103 (3): 441–463.

DellaVigna, Stefano, and Matthew Gentzkow. 2019. "Uniform pricing in us retail chains." *The Quarterly Journal of Economics* 134 (4): 2011–2084.

Doyle, Joseph J, and Krislert Samphantharak. 2008. "$2.00 Gas! Studying the Effects of a Gas Tax Moratorium." *Journal of Public Economics* 92 (3): 869–884.

Fullerton, Don, and Gilbert E Metcalf. 2002. "Tax Incidence." *Handbook of Public Economics* 4: 1787–1872.

Gaarder, Ingvil. 2018. "Incidence and Distributional Effects of Value Added Taxes." *The Economic Journal* 129 (618): 853–876.

Garriga, Santiago, and Dario Tortarolo. 2021. "Wage effects of employer-mediated transfers."

Gruber, Jonathan. 1994. "The incidence of mandated maternity benefits." *The American economic review*: 622–641.

Gruber, Jonathan. 1997. "The incidence of payroll taxation: evidence from Chile." *Journal of labor economics* 15 (S3): S72–S101.

Gudgeon, Matthew, and Simon Trenkle. 2024. "The speed of earnings responses to taxation and the role of firm

labor demand." *Journal of Labor Economics* 42 (3): 000–000.

Guo, Audrey. 2023. "Payroll Tax Incidence: Evidence from Unemployment Insurance." *arXiv preprint arXiv:2304.05605*.

Harberger, Arnold C. 1962. "The incidence of the corporation income tax." *The Journal of Political Economy*: 215–240.

Harberger, Arnold C. 1964. "The measurement of waste." *The American Economic Review* 54 (3): 58–76.

Harju, Jarkko, Tuomas Kosonen, and Oskar Nordström Skans. 2018. "Firm types, price-setting strategies, and consumption-tax incidence." *Journal of Public Economics* 165: 48–72.

Jones, Damon. 2010. "Information, preferences, and public benefit participation: Experimental evidence from the advance EITC and 401 (k) savings." *American Economic Journal: Applied Economics* 2 (2): 147–163.

Keen, Michael, and Joel Slemrod. 2021. *Rebellion, rascals, and revenue: Tax follies and wisdom through the ages*.: Princeton University Press.

Kopczuk, Wojciech, Justin Marion, Erich Muehlegger, and Joel Slemrod. 2016. "Does Tax-Collection Invariance Hold? Evasion and the Pass-Through of State Diesel Taxes." *American Economic Journal: Economic Policy* 8 (2): 251–86.

Kosonen, Tuomas. 2015. "More and Cheaper Haircuts After VAT Cut? On the Efficiency and Incidence of Service Sector Consumption Taxes." *Journal of Public Economics* 131: 87–100.

Kotlikoff, Laurence, and Lawrence Summers. 1987. "Tax Incidence." *Handbook of Public Economics* 2: 1043–1092.

Lehmann, Etienne, François Marical, and Laurence Rioux. 2013. "Labor income responds differently to income-tax and payroll-tax reforms." *Journal of Public Economics* 99: 66–84.

Lobel, Felipe. 2024. "Who Benefits from Payroll Tax Cuts? Market Power, Tax Incidence and Efficiency." *Working Paper*.

Musgrave, Richard Abel, and Richard A Musgrave. 1959. *The theory of public finance: a study in public economy*. vol. 658: McGraw-Hill New York.

Nakamura, Emi, and Jón Steinsson. 2013. "Price rigidity: Microeconomic evidence and macroeconomic implications." *Annu. Rev. Econ.* 5 (1): 133–163.

Piketty, Thomas. 2014. *Capital in the twenty-first century*.: Harvard University Press.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income inequality in the United States, 1913–1998." *The Quarterly journal of economics* 118 (1): 1–41.

Poterba, James M. 1996. "Retail price reactions to changes in state and local sales taxes." *National Tax Journal*: 165–176.

Rothstein, Jesse. 2010. "Is the EITC as good as an NIT? Conditional cash transfers and tax incidence." *American economic Journal: economic policy* 2 (1): 177–208.

Saez, Emmanuel, Manos Matsaganis, and Panos Tsakloglou. 2012. "Earnings determination and taxes: Evidence from a cohort-based payroll tax reform in Greece." *The Quarterly Journal of Economics* 127 (1): 493–533.

Saez, Emmanuel, Benjamin Schoefer, and David Seim. 2019. "Payroll taxes, firm behavior, and rent sharing: Evidence from a young workers' tax cut in Sweden." *American Economic Review* 109 (5): 1717–1763.

Saez, Emmanuel, Benjamin Schoefer, and David Seim. 2021. "Hysteresis from employer subsidies." *Journal of Public Economics* 200: 104459.

Saez, Emmanuel, and Gabriel Zucman. 2016. "Wealth inequality in the United States since 1913: Evidence from capitalized income tax data." *The Quarterly Journal of Economics* 131 (2): 519–578.

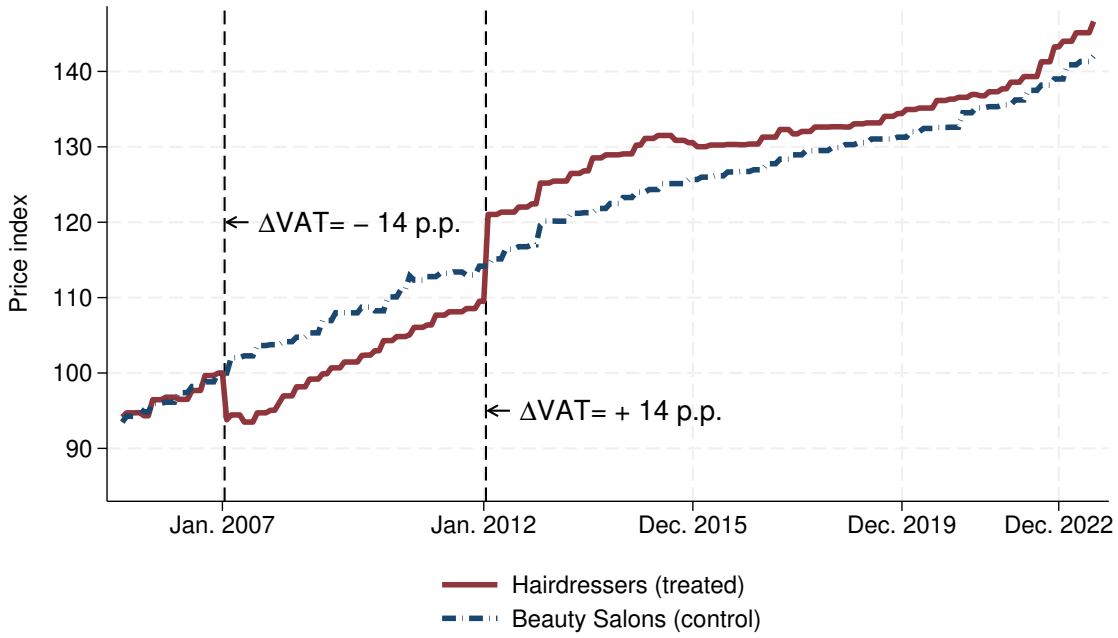
Saez, Emmanuel, and Gabriel Zucman. 2023. "Distributional Tax Analysis in Theory and Practice: Harberger Meets Diamond-Mirrlees."

Summers, Lawrence H. 1989. "Some simple economics of mandated benefits." *The American Economic Review* 79 (2): 177–183.

Tax Policy Center. 2022. "Brief Description of the Tax Model." *Washington DC. Link: https://www.taxpolicycenter.org/resources/brief-description-tax-model*.

Tazhitdinova, Alisa. 2020. "Do only tax incentives matter? Labor supply and demand responses to an unusually

large and salient tax break." *Journal of Public Economics* 184: 104162.
US Congressional Budget Office. 2018. "The Distribution of Household Income, 2016.." *Washington DC*.
US Joint Committee on Taxation. 2019. "Revenue Estimating Methodology." *Washington DC*.
US Treasury, Office of Tax Analysis. 2019. "Distributional Analysis of the US Tax System." *Washington DC*.



A. Price Hysteresis

FIGURE A1. Asymmetry Pass-Through

This figure extends the price series of Figure 1 in Benzarti et al. (2020) by seven years. It shows the pass-through of a 14 percentage point VAT cut and increase in the hairdresser industry in Finland compared to the beauty salon industry (which did not experience any VAT change). Even after the VAT is reinstated to its pre-January 2007 levels, prices in the hairdresser industry remain higher than in the control group for more than 10 years.