

Pricing when Customers Care about Fairness but Misperceive Markups

Erik Eyster, Kristof Madarasz, Pascal Michaillat*

September 7, 2017

This paper proposes a theory of price rigidity consistent with survey evidence that firms stabilize prices out of fairness to their consumers. The theory relies on two psychological assumptions. First, customers care about the fairness of prices: fixing the price of a good, consumers enjoy it more at a low markup than at a high markup. Second, customers underinfer marginal costs from prices: when prices rise due to an increase in marginal costs, customers underappreciate the increase in marginal costs and partially misattribute higher prices to higher markups. Firms anticipate customers' reaction and trim their price increases. Hence, the passthrough of marginal costs into prices falls short of one—prices are somewhat rigid. Embedded in a simple macroeconomic model, our pricing theory produces nonneutral monetary policy, a short-run Phillips curve that involves both past and future inflation rates, a hump-shaped impulse response of output to monetary policy, and a nonvertical long-run Phillips curve.

*Eyster: London School of Economics. Madarasz: London School of Economics. Michaillat: Brown University. We thank George Akerlof, Roland Benabou, Daniel Benjamin, Joaquin Blaum, Olivier Coibion, Stephane Dupraz, Gauti Eggertsson, John Friedman, Xavier Gabaix, Nicola Gennaioli, Yuriy Gorodnichenko, Shachar Kariv, David Laibson, John Leahy, Matthew Rabin, Ricardo Reis, David Romer, Klaus Schmidt, Jesse Shapiro, Andrei Shleifer, Silvana Tenreyro, and participants at seminars and conferences for helpful comments and discussions.

1. Introduction

Empirical evidence suggests that prices are somewhat rigid: neither are they fixed exactly, nor do they fully respond to marginal-cost shocks. A series of recent papers uses exogenous variations in marginal costs together with price data to measure the passthrough of marginal costs into prices. These variations in marginal costs arise in various contexts, from a broad variety of sources: changes in the value-added tax in several European countries, changes in unit labor costs in Swedish firms, changes in import tariffs in India, and fluctuations in exchange rates. In all these cases, the short-run passthrough is quite low, falling between 0.1 and 0.8, with a median estimate about 0.5. Moreover, those studies able to measure longer-run passthroughs also find incomplete passthrough after two years and more. This price rigidity has important implications: it determines the incidence and effects of taxes, the effects of tariffs and exchange-rate movements, the effects of changes in wages and commodity prices, and the influence of monetary policy on employment and output.

Numerous models have been developed to explain these facts. One channel that has received limited attention is the role of fairness. Yet a growing body of evidence (reviewed in Section 2, alongside evidence on passthrough rates) suggests that firms only reluctantly raise prices for fear of alienating customers, who are averse to paying prices that they regard as unfair. Because the welfare properties of the rigid-price models widely used for policy analysis depend upon their microfoundations, a model that conforms to the motivations of price setters may be beneficial.

In this paper, we develop a model of pricing that matches firms' view that fairness considerations play an important role in pricing. Our model rests upon two psychological assumptions. First, we assume that customers dislike paying prices above a fair markup on marginal costs. This assumption draws upon evidence from the seminal work of Kahneman, Knetsch, and Thaler (1986), who document that people find it acceptable for firms to raise prices in response to higher marginal costs but unfair for firms to raise prices in response to elevated demand. Several firm and consumer surveys, our own survey of French bakers, and religious and legal texts support the view that customers attend to markups and recoil at paying high markups, and that firms understand this. Because customers typically do not observe firms' costs, their fairness perceptions crucially depend upon their estimates of these costs. Here we assume that customers update their beliefs about firms' marginal costs less than rationally from available information—they form beliefs that lie somewhere between their priors and rational beliefs. Customers who underinfer firms' marginal costs from firms' prices partially misattribute higher prices to higher markups; they therefore conclude that the higher prices are less fair. This second psychological assumption about underinference draws on evidence from a number of different

contexts (discussed in Section 2) that people have a general tendency to infer less than they should about other people's private information from these other people's actions.

We begin our formal analysis in Section 3 by embedding these two psychological assumptions into a simple model of monopolistic pricing. When modeling customers' concern for fair prices, we assume that the utility derived from consuming a good depends on the perceived fairness of the transaction. Customers begin with some notional fair markup, K^f . When they buy a good at price P whose perceived marginal cost is MC^P , they deem its markup to be $K^P = P/MC^P$. Customers weight each unit of consumption by the factor $F = 2/[1 + (K^P/K^f)^\theta]$. The parameter θ describes fairness concerns: when $\theta = 0$, customers do not care about fairness, and the model reduces to a typical monopoly model; when $\theta > 0$, customers care about the fairness of prices. Demand decreases in price not only due to standard substitution effect, but also through the fairness channel: paying an unfairly high price lowers the marginal utility of consumption. Fairness concerns, operating through the fairness measure F , make demand more elastic than it would be otherwise.

Because customers do not directly observe firms' marginal costs, their inferences about marginal cost play a pivotal role in the model. We assume that customers misperceive the monopoly's true marginal cost MC given price P as $MC^P = (MC^b)^\gamma \times (P/K^b)^{1-\gamma}$. The parameter MC^b represents customers' prior expectation about the marginal cost, and the parameter K^b represents a perceived proportional markup rate. The parameter $\gamma \in [0, 1]$ measures customers' naivety when inferring marginal costs. When $\gamma = 0$, customers make inferences as if firms used the constant markup factor K^b . Given such inference, firms would indeed optimally employ a constant markup. In this case (whether or not K^b matches the equilibrium markup), fairness plays absolutely no role, and the marginal-cost passthrough is one. When $\gamma > 0$, customers underappreciate the extent to which changes in price reveal changes in marginal cost; in that case, fairness matters. Such customers do update their beliefs in the right direction from available information but stop short of rational inference: their beliefs move too little relative to their priors. As the customers underestimate the change in marginal cost that accompanies a price change, they misattribute part of that price change to a change in markup.

Fairness concerns combined with underinference generate price rigidity. After an increase in price spurred by higher marginal cost, customers underappreciate the increase in marginal cost; they conclude that the markup is higher, which they find unfair. This lower perceived transactional fairness increases the elasticity of their demand. Understanding this effect, the monopoly reduces its markup. In sum, rising marginal costs bring about price increases that are less than proportionate to the increase in marginal cost: the passthrough falls short of one.

In Section 4, we replace Calvo (1983) pricing with fairness pricing in a simple New Key-

nesian model to illustrate how our two psychological assumptions generate monetary-policy nonneutrality. In this dynamic model, consumers in period $t + 1$ form beliefs about current marginal costs that are a weighted average of their period- t beliefs and beliefs derived from period- $t + 1$ prices. The more sophisticated are consumers, the greater the weight that they attach to period- $t + 1$ prices. We find that in the short run, monetary policy affects output: monetary policy is nonneutral. Moreover, the short-run Phillips curve is not purely forward-looking: it links current employment and inflation not only to expected future inflation but also to past inflation. In addition, by calibrating parameters that measure fairness concerns and inferential naivety to match the empirical evidence on passthrough, we find that the model generates reasonable impulse responses to monetary-policy shocks. In particular, the impulse response of output is hump-shaped. Finally, the long-run Phillips curve is not vertical: higher steady-state inflation leads to higher steady-state employment. Although our model is not the only one that can generate these predictions, it differs from the rest of the literature by showing that these predictions follow naturally from a small set of behavioral assumptions backed by the literature and consistent with the views of real-world price-setters.

Related Literature. Rotemberg (2005) pioneered the study of the implications of fairness for price rigidity.¹ He assumes that customers care about firms' altruism—their taste for increasing customers welfare—which they re-evaluate following every price change. Customers buy a normal amount from the firm unless they can reject the hypothesis that the firm is altruistic, in which case they withhold all demand in order to lower the firm's profits. Firms react to the discontinuity in demand by refraining from passing on small cost increases, which leads to price stickiness.

In this paper, we retool the psychological assumption of Rotemberg (2005) that customers refuse to purchase from unfair firms by assuming that customers enjoy a good less the less fair they regard its price. Despite broad similarities, the two assumptions differ conceptually: unlike Rotemberg's, our assumption implies that customers would withhold demand from unfair firms even if doing so would not hurt the firms. This difference allows us to move away from Rotemberg's discontinuous, buy-normally-or-buy-nothing formulation to one in which customers continuously reduce demand as the unfairness of the transaction increases. The greater tractability of our continuous formulation allows us to obtain closed-form expressions for the markup and passthrough, as well as to embed the pricing model into a New Keynesian framework. It also allows us to clarify the role of inference about marginal costs in explaining

¹Rotemberg (2011) further explores the implications of fairness for pricing, focusing on other phenomena such as price discrimination.

price rigidity. We find that fairness is necessary but not sufficient to obtain price rigidity; only when fairness is combined with underinference about marginal costs do prices become rigid.

Our work also relates to other papers that introduce fairness considerations into otherwise standard models. For example, Akerlof (1982), Akerlof and Yellen (1990), and Benjamin (2015) introduce fairness into labor-market models to explain the prevalence of unemployment and wage rigidity. Rabin (1993), Fehr and Schmidt (1999), and Charness and Rabin (2002) add fairness to game-theoretic models to explain departures from pure self-interest observed in laboratory experiments, notably public-good and ultimatum games. Fehr, Klein, and Schmidt (2007) explore the implications of fairness for contract theory. A last example is Zajac (1985), who describes how principles of fairness can be incorporated into public-utility regulation.²

Much of the work on fairness, like Rotemberg (2005), uses social preferences, such as the models of Rabin (1993) or Fehr and Schmidt (1999). These preferences have the property that fairness considerations do not affect people's marginal rates of substitution amongst different goods or between labor and leisure. Consequently, people behave in general equilibrium as if they did not care about fairness (Dufwenberg et al. 2011). Our formulation of fairness has the advantage that it has effects even in general equilibrium. Customers who feel mistreated by firms withhold demand not to punish firms, as in models of social preferences, but instead because they derive less joy from consuming unfairly priced goods. Fairness perceptions affect marginal rates of substitution between goods, influencing the general equilibrium. We view our approach and the social-preference approach as complementary: while we fully agree with Schmidt (2011) that models of social preferences offer important insights on agency problems in organizational settings, we also believe that our preferences could help develop the role of fairness in other contexts, especially macroeconomics.

Finally, our pricing model relates to other models that rely on a nonconstant price elasticity of demand to create variations in markups after shocks. In international economics, these models have long been used to explain the behavior of exchange rates and prices (for example, Dornbusch 1985). More recent models include those by Bergin and Feenstra (2001), Atkeson and Burstein (2008), and Melitz and Ottaviano (2008). In macroeconomics, such models have been used to create real rigidities—in the sense of Ball and Romer (1990)—that amplify nominal rigidities. The precursor in this literature was Kimball (1995), with many studies building up on his model (for example, Dotsey and King 2005; Eichenbaum and Fisher 2007; Klenow and Willis 2016). There are two differences between our model and those, however. First, many of these models make reduced-form assumptions (either in the utility function or directly in the demand curve) to

²For surveys of the literatures bringing fairness into economics, see Fehr and Gächter (2000), Jones and Mann (2001), and Fehr, Goette, and Zehnder (2009).

obtain a nonconstant price elasticity of demand; our model provides a microfoundation for this property. Second, models of real rigidity cannot generate money nonneutrality alone: because the price elasticity of demand is increasing in the relative price charged by a firm, the real rigidity must be combined with a nominal rigidity (such as a menu cost) to generate nonneutrality. In contrast, in our model, the price elasticity of demand is increasing in the absolute price charged by a firm, so our mechanism alone generates nonneutrality.

2. Empirical Motivation

Our paper proposes an explanation of price rigidity based on customers' concern for the fairness of prices and underinference of marginal costs from prices. In this section, we present empirical evidence that motivates our theory. This evidence suggests that prices are rigid, that people care about the fairness of prices, and that people underinfer hidden information from observable actions. Our pricing model will be designed to explain the behavior of prices documented here; its assumptions will be designed to capture parsimoniously—although perhaps a bit coarsely—the evidence on fairness and underinference.

2.1. Price Rigidity

Here we present evidence from a broad range of contexts that prices are somewhat rigid: while prices respond immediately to fluctuations in marginal costs, they respond much less than one-for-one. We find that a median estimate of the marginal-cost passthrough is around 0.5.

First, the passthrough of value-added tax into prices is far below one. From a monopoly's perspective, a change in value-added tax acts exactly as a change in marginal cost, so the fact that the tax passthrough is incomplete indicates that the marginal-cost passthrough would also be incomplete.³ Recent reforms in value-added tax in European countries provide natural experiments that offer compelling evidence of low passthroughs. A first example comes from France, where Benzarti and Carloni (2016) study a 14-percentage-point cut of the value-added tax applied to sit-down restaurants (from 19.6% to 5.5%) in 2009. Using a difference-in-differences strategy comparing sit-down restaurants to non-restaurant market services and non-restaurant small firms, they find that restaurants prices adjusted immediately to the reform

³With a value-added tax τ , there is a wedge between the post-tax price P and the pretax price $\widehat{P} = P/(1 + \tau)$. The monopoly's profits are $Y^d(P) (\widehat{P} - MC) = Y^d(P) (P - (1 + \tau)MC)/(1 + \tau)$. Maximizing profits implies maximizing $Y^d(P) (P - (1 + \tau)MC)$. Hence, with a value-added tax, the monopoly behaves as if there was no tax but the marginal cost was $(1 + \tau)MC$. An increase in tax from τ_0 to τ_1 therefore triggers an increase in marginal cost by $(\tau_1 - \tau_0)/(1 + \tau_0) \times 100$ percent.

but only decreased by 2%. Their results imply a low passthrough of $2/(14.1/1.196) = 0.17$. Another example comes from Finland, where Kosonen (2015) studies the impact of a 14-percentage-point reduction in the value-added tax on hairdressing services (from 22% to 8%). He finds that hairdressers cut their prices by an amount that corresponds to a marginal-cost passthrough of 0.5. Finally, considering all changes to value-added taxes across all goods and all countries of the European Union between 1996 and 2015, Benzarti et al. (2017, Figures 1 and 2) provide systematic evidence of incomplete passthrough. For increases of the value-added tax the passthrough is estimated between 0.53 and 0.63, and for decreases between 0.06 and 0.17, depending on the empirical specification.

Second, the passthrough of shocks to marginal production costs into prices is much below one. The challenges to estimating passthrough are to isolate exogenous variations in marginal costs and measure both marginal costs and prices. A few studies have succeeded in doing that. Using matched data on product-level prices and producers' unit labor cost for Sweden, Carlsson and Skans (2012) find a moderate passthrough of idiosyncratic marginal-cost changes into prices: about 0.33. Next, De Loecker et al. (2016, Table 7) find that after trade liberalization in India, marginal costs fell significantly due to the import tariff liberalization, but prices failed to fall in step: they estimate passthroughs between 0.34 and 0.41, depending on the empirical specification.

Third, the passthrough of the exchange rate into import prices is well below one. Gopinath and Rigobon (2008) use microdata on US import prices at the dock for the period 1994–2005 to find that—even when conditioning on firms actually changing their prices—the exchange-rate passthrough into import prices is only around 0.2. Using more aggregated data on import prices for 8 developed economies, including the United States, Burstein and Gopinath (2014, Table 7.4) estimate a short-run exchange-rate passthrough into import prices between 0.13 and 0.75, depending on the country, with an average across countries of 0.45.

2.2. Fairness

The principal motivation for including fairness considerations into a pricing model is that price-setters identify it to be a major concern in price-setting. Starting with the pioneering survey by Blinder et al. (1998), researchers have interviewed managers at more than 11,000 firms across the US, Canada and Europe about their pricing practices. The typical study has managers evaluate the relevance of different pricing theories from the economics literature (for instance, menu costs) to explain their own pricing, in particular price rigidity. Amongst the theories that the managers deem most important, some version of fairness invariably appears, often called “implicit contracts” and described as follows: “firms tacitly agree to stabilize prices, perhaps

out of fairness to customers.” Table 1 summarizes these surveys.

Table 2 ranks the most popular theories of price rigidity from the surveys. Fairness appeals to price-setters more than any other theory, with a median rank of 1 and a mean rank of 2.1. The second most popular explanation for price rigidity takes the form of nominal contracts—prices do not change because they are fixed by contracts—which has a median rank of 3 and a mean rank of 2.8. Two common macroeconomic theories of price rigidity—menu costs and information delays—do not resonate at all with price setters, who consistently rank them amongst the least popular theories, with mean and median ranks around 10.

Given the evidence on price-setters’ beliefs, it should come as no surprise that consumers do indeed tend to regard price increases as unfair. In a survey conducted by Shiller (1997), 85% of respondents report that they dislike inflation because when they “go to the store and see that prices are higher”, they “feel a little angry at someone” (p. 21). The most common culprits include “manufacturers”, “store owners”, and “businesses”, and the most common causes include “greed” and “corporate profits” (p. 25). If firms aim to nurture customers’ goodwill, they will certainly account for customers’ aversion to price increases when setting prices.

To model fairness, we assume that people care about the fairness of firms’ markups. Religious and legal texts written over the ages suggest a long history of norms regarding markups. For example, Talmudic law specifies that the highest fair and allowable markup when trading “essential items” is 20% of the production cost, or one-sixth of the final price.⁴ The payer of any higher markup is entitled to a refund. Another example comes from 18th-century France, where local authorities fixed bread prices by publishing “fair” prices in official decrees. In the city of Rouen, for instance, the official bread prices took the costs of grain, rent, milling, wood, and labor into account, and granted a “modest profit” to the baker (Miller 1999, p. 36). Thus, officials fixed the markup that bakers could charge. Even today, French bakers attach such importance to convincing their customers of fair markups that their trade union decomposes into minute detail the cost of bread and the rationale for any price rise.⁵ A last example comes from the United States, where return-on-cost regulation for public utilities has been justified not only on efficiency grounds, but also on fairness grounds (see Okun 1981, p. 153 and Jones and Mann 2001, p. 153).

Our assumption that people care about the fairness of markups implies that they dislike price increases unjustified by cost increases, which entail a rise in markup. In a telephone survey of

⁴See the statement of Shmuel, p. 49b of *Bava Metzhia, Nezikin*, available at http://www.halakhah.com/pdf/nezikin/Baba_Metzia.pdf. Some “nonessential items” have maximum markups of 100%, while still others carry no limits. “Essential items” seem to cover food items, although there is debate about the exact boundaries of each category of goods (essential, nonessential, and other). Warhaftig (1987) discusses these rules.

⁵See <http://www.boulangerie.net/forums/bnweb/prixbaguette.php>.

Table 1. Description of Firm Surveys About Pricing

| Study | Country | Period | Sample size | Sales to customers |
|--|----------------|-----------|-------------|--------------------|
| Blinder et al. (1998) | United States | 1990–1992 | 200 | 85% |
| Hall, Walsh, and Yates (2000) | United Kingdom | 1995 | 654 | 59% |
| Apel, Friberg, and Hallsten (2005) | Sweden | 2000 | 626 | 86% |
| Amirault, Kwan, and Wilkinson (2006) | Canada | 2002–2003 | 170 | – |
| Kwapil, Baumgartner, and Scharler (2005) | Austria | 2004 | 873 | 81% |
| Aucremanne and Druant (2005) | Belgium | 2004 | 1,979 | 78% |
| Loupias and Ricart (2004) | France | 2004 | 1,662 | 54% |
| Lunnemann and Matha (2006) | Luxembourg | 2004 | 367 | 85% |
| Hoeberichts and Stokman (2006) | Netherlands | 2004 | 1,246 | – |
| Martins (2005) | Portugal | 2004 | 1,370 | 83% |
| Alvarez and Hernando (2005) | Spain | 2004 | 2,008 | 86% |

Table 2. Ranking of Theories Explaining Price Rigidity Across Surveys

| Theory | Country | | | | | | | | | | | Mean | Median |
|----------------------|---------|----|----|----|----|----|----|----|----|----|----|------|--------|
| | US | GB | SE | CA | AT | BE | FR | LU | NL | PT | ES | | |
| Implicit contracts | 4 | 5 | 1 | 2 | 1 | 1 | 4 | 1 | 2 | 1 | 1 | 2.1 | 1 |
| Nominal contracts | 5 | 1 | 3 | 3 | 2 | 2 | 3 | 3 | 1 | 5 | 3 | 2.8 | 3 |
| Coordination failure | 1 | 3 | 4 | 5 | 5 | 5 | 2 | 9 | 4 | 2 | 2 | 3.8 | 4 |
| Pricing points | 8 | 4 | 7 | – | 10 | 13 | 8 | 10 | 7 | 11 | 6 | 8.4 | 8 |
| Menu costs | 6 | 11 | 11 | 10 | 7 | 15 | 10 | 13 | 8 | 10 | 7 | 9.8 | 10 |
| Information delays | 11 | – | 13 | 11 | 6 | 14 | – | 15 | – | 8 | 9 | 10.9 | 11 |

Notes: Respondents to the surveys rated the relevance of several pricing theories in explaining price rigidity in their own firm. The table shows how common theories rank amongst the alternatives. Blinder et al. (1998, Table 5.1) describes the theories as follows (with wording varying slightly across surveys): “implicit contracts” stands for “firms tacitly agree to stabilize prices, perhaps out of fairness to customers”; “nominal contracts” stands for “prices are fixed by contracts”; “coordination failure” stands for two closely related theories, which are investigated in separate surveys: “firms hold back on price changes, waiting for other firms to go first” and “the price is sticky because the company loses many customers when it is raised, but gains only a few new ones when the price is reduced” (which is labeled “kinked demand curve”); “pricing points” stands for “certain prices (like \$9.99) have special psychological significance”; “menu costs” stands for “firms incur costs of changing prices”; “information delays” stands for two closely related theories, which are investigated in separate surveys: “hierarchical delays slow down decisions” and “the information used to review prices is available infrequently.” The rankings of the theories are reported in Table 5.2 in Blinder et al. (1998); Table 3 in Hall, Walsh, and Yates (2000); Table 4 in Apel, Friberg, and Hallsten (2005); Table 8 in Amirault, Kwan, and Wilkinson (2006); Table 5 in Kwapil, Baumgartner, and Scharler (2005); Table 18 in Aucremanne and Druant (2005); Table 6.1 in Loupias and Ricart (2004); Table 8 in Lunnemann and Matha (2006); Table 10 in Hoeberichts and Stokman (2006); Table 4 in Martins (2005); and Table 5 in Alvarez and Hernando (2005).

one-hundred Canadian residents, Kahneman, Knetsch, and Thaler (1986) document this pattern. They describe the following situation: “A hardware store has been selling snow shovels for \$15. The morning after a large snowstorm, the store raises the price to \$20.” Only 18% of subjects regard this pricing behavior as acceptable, whereas 82% regard it as unfair (p. 729). Conversely, our fairness assumption suggests that customers tolerate price increases following cost increases so long as the markup remains constant. Kahneman, Knetsch, and Thaler also identify this pattern: “Suppose that, due to a transportation mixup, there is a local shortage of lettuce and the wholesale price has increased. A local grocer has bought the usual quantity of lettuce at a price that is 30 cents per head higher than normal. The grocer raises the price of lettuce to customers by 30 cents per head.” 79% of subjects regard the grocer’s behavior as acceptable, and only 21% find it unfair (pp. 732–733).⁶

Numerous subsequent studies have confirmed and refined Kahneman, Knetsch, and Thaler’s results.⁷ For example, in a survey of 1,750 households in Switzerland and Germany, Frey and Pommerehne (1993, pp. 297–298) confirmed that customers dislike a price increase that involves an increase in markup; so too do Shiller, Boycko, and Korobov (1991, p. 389) in a comparative survey of 391 respondents in Russia and 361 in the United States. Using an online survey of 307 Dutch individuals, Gielissen, Dutilh, and Graafland (2008, Table 2) confirm that price increases that follow cost increases are fair, whereas those that follow demand increases are not. One natural concern about the snow-shovel-vignette evidence is that people may find the price increase unfair simply because it occurs during a period of hardship. To address this question, Maxwell (1995) ask 72 students at a Florida university about price increases following an ordinary increase in demand as well as those following a hardship. While fewer find price increases in the former environment than in the latter environment unfair (69% versus 86%), a substantial majority in each case perceive the price increase as unfair.

In our model, customers deem equally unfair for firms not to pass along cost decreases. The evidence on this assumption is weaker. Kahneman, Knetsch, and Thaler (1986) describe the following situation: “A small factory produces tables and sells all that it can make at \$200 each. Because of changes in the price of materials, the cost of making each table has recently decreased by \$20. The factory does not change its price of tables.” Only 47% of respondents find this unfair, despite the elevated markup (p. 734). However, subsequent studies challenge this finding by suggesting that people do expect the price to fall after the cost reduction. For instance, Kalapurakal, Dickson, and Urbany (1991) conducted a survey of 189 business students

⁶Strictly speaking, this evidence suggests that people care about additive markups, whereas our model assumes that people attend to multiplicative markups. We suspect that subjects would have volunteered the same preferences had they been posed questions about proportional markups.

⁷For a survey, see Xia, Monroe, and Cox (2004).

in the United States, and asked them to consider the following scenario: “A department store has been buying an oriental floor rug for \$100. The standard pricing practice used by department stores is to price floor rugs at double their cost so the selling price of the rug is \$200. This covers all the selling costs, overheads and includes profit. The department store can sell all of the rugs that it can buy. Suppose because of exchange rate changes the cost of the rug rises from \$100 to \$120 and the selling price is increased to \$220. As a result of another change in currency exchange rates, the cost of the rug falls by \$20 back to \$100.” Then two alternative scenarios were evaluated: “The department store continues to sell the rug for \$220” compared to “The department store reduces the price of the rug to \$200.” The scenario in which the department store reduces the price in response to the decrease in cost was considered significantly more fair: the fairness rating was +2.3 instead of -0.4 (where -3 is extremely unfair and +3 extremely fair). Similarly, using a survey with US respondents, Konow (2001, Table 6) finds that if a factory that sells a table at \$150 suddenly locates a supplier charging \$20 less for materials, then the average new fair price is \$138, well below \$150.⁸

Finally, in our model we assume that customers who find a transaction unfair derive lower utility from consuming the good, which reduces their propensity to purchase the good. There is some evidence suggesting that customers indeed reduce purchases when they feel unfairly treated. In a telephone survey of 40 US consumers, Urbany, Madden, and Dickson (1989) explore—by looking at a 25-cent ATM surcharge—whether a price increase justified by a cost increase is perceived as more fair than an unjustified one, and whether fairness perceptions affect customers’ propensity to buy. While 58% of respondents judge the introduction of the fee fair when justified by a cost increase, only 29% judge it fair when not justified (Table 1, panel B). Moreover, those people who find the surcharge unfair are indeed more likely to switch banks (52% versus 35%, see Table 1, panel C). Similarly, Piron and Fernandez (1995) present survey and laboratory evidence that customers who find a firm’s actions unfair are more likely to reduce their purchases with that firm.

We not only assume that customers bristle at unfair markups, but also that firms understand this. Blinder et al. (1998, p. 153, p. 157) find evidence that they do: 64% of firms say that customers do not tolerate price increases after demand increases; 71% of firms say that customers do tolerate price increase after cost increases. Apparently, the norm for fair pricing revolves around markups over marginal cost. Indeed, based on a survey of businessmen in the United

⁸Firms whose customers appraise prices relative to marginal costs have less incentive to innovate to cut marginal costs. In a survey of 1,530 cable-car customers in Switzerland, Bieger, Engeler, and Laesser (2010, Table 3) find that while an external, uncontrollable cost increase (for instance, from increased security requirements) gets perceived as a fair reason to raise prices, an internal, controllable cost increase (for instance, from higher marketing expenditures) gets perceived as a less fair reason for raising prices. Nevertheless, respondents find both types of price increase much fairer than an unexplained price increase.

Kingdom, Hall and Hitch (1939, p. 19) report that the “the ‘right’ price, the one which ‘ought’ to be charged” is widely perceived to be a markup (generally, 10%) over average cost. Okun (1975, p. 362) also observes in discussions with business people that “empirically, the typical standard of fairness involves cost-oriented pricing with a markup.”

To better understand how firms incorporate fairness into their pricing decisions, we interviewed 31 bakers in France in 2007. The French bread market makes a good case study because the market is large, bakers set their prices freely, and French people care enormously about bread.⁹ We sampled bakeries in cities and villages around Grenoble, Aix-en-Provence, Paimpol, and Paris. Overall, the interviews show that bakers’ efforts to preserve customer loyalty constrain price variation. Price adjustments are guided by norms of fairness to avoid antagonizing customers; in particular, cost-based pricing is widely used. Bakers raise the price of bread only in response to cost increases such as those for flour (generally only at the end of harvest in September), utilities, or wages. The bakers emphasized that prices increase only in response to cost increases, with any increase explained carefully. Bakers also refuse to increase prices in response to increased demand. Several bakers explained that they do not change prices during weekends (when more people shop at bakeries), during the holiday absences of local competitors (when their demand and market power rise), or during the summer tourist season (again, when demand rises) because it would be unfair, and hence anger and drive away customers.

2.3. Underinference

Customers do not directly observe firms’ marginal costs, so their perceptions as to how fairly firms price their goods depend upon their estimates of these costs. In modeling the inferences that customers draw about marginal costs, we assume that customers underappreciate the extent to which changes in prices reveal changes in marginal costs. In our model, a rational customer would understand that marginal costs move in proportion to prices. By contrast, we focus on customers who draw subproportional inferences about cost from price: upon observing a price change, they update their belief about the marginal cost in the right direction, but stop short of rational, proportional inference.

Our assumption of subproportional inference is motivated by numerous experimental studies

⁹In 2005, bakeries employed 148,000 workers, for a yearly turnover of 3.2 billion euros (Fraichard 2006). Since 1978, French bakers have been free to set their own prices, except during the inflationary period 1979–1987 when price ceilings and growth caps were imposed. For centuries, bread prices caused major social upheaval in France. Miller (1999, p. 35) explains that before the French Revolution, “affordable bread prices underlay any hopes for urban tranquility.” During the Flour War of 1775, mobs chanted “if the price of bread does not go down, we will exterminate the king and the blood of the Bourbons”; following these riots, “under intense pressure from irate and nervous demonstrators, the young governor of Versailles had ceded and fixed the price ‘in the King’s name’ at two sous per pound, the mythohistoric just price inscribed in the memory of the century” (Kaplan 1996, p. 12).

across a range of settings which establish that people underinfer other people's information from their actions. Samuelson and Bazerman (1985), Holt and Sherman (1994), and Carillo and Palfrey (2011), among others, provide evidence in the context of bilateral bargaining with asymmetric information that bargainers underappreciate the adverse selection in trade. The papers collected in Kagel and Levin (2002) present evidence that bidders underattend to the winner's curse in common-value auctions. In a metastudy of social-learning experiments, Weizsacker (2010) finds evidence that subjects behave as if they underinfer their predecessors' private information from their actions. In a voting experiment, Esponda and Vespa (2014) show that people underinfer others' private information from their votes.

Consumers who infer subproportionally succumb to money illusion because price contaminates their perceptions of markups: when the price is higher, households believe that the firm captures a larger markup; when the price is lower, households believe that the firm captures a smaller markup. Since their fairness perceptions depend upon the price level, people exhibit money illusion. Shafir, Diamond, and Tversky (1997) report evidence that indirectly supports our assumption using the following thought experiment: "Changes in the economy often have an effect on people's financial decisions. Imagine that the US experienced unusually high inflation which affected all sectors of the economy. Imagine that within a six-month period all benefits and salaries, as well as the prices of all goods and services, went up by approximately 25%. You now earn and spend 25% more than before. Six months ago, you were planning to buy a leather armchair whose price during the 6-month period went up from \$400 to \$500. Would you be more or less likely to buy the armchair now?" The higher prices were distinctly aversive: while 55% of respondents were as likely to buy as before and 7% were more likely to buy as before, 38% of respondents were less likely to buy than before (p. 355). Our model of underinference makes this prediction because some households perceive markups to be higher when prices are higher, which reduces the fairness of the transaction and households' willingness to pay for it.

Finally, our assumption of underinference resembles several recent models of limited attention. The "availability heuristic" documented by Tversky and Kahneman (1973) and formalized by Gennaioli and Shleifer (2010) posits that people infer information content by drawing upon a limited set of scenarios that come to mind: here higher prices suggest greed and increased markups, rather than higher marginal costs. Customers in our model are also "coarse thinkers" in the sense of Mullainathan, Schwartzstein, and Shleifer (2008) because they do not distinguish between scenarios where changes in price reflect changes in cost and those where they reflect changes in markup. Whereas we regard households' failure to infer marginal costs as a cognitive error, it might also result from economizing on attention costs along the lines proposed by Gabaix (2014, 2016). Lastly, the notion that to the extent that they infer about marginal costs,

consumers assume proportional markups matches evidence that people think proportionally. Bushong, Rabin, and Schwartzstein (2015) provide a model of such proportional thinking as well as summarize the supporting evidence.

3. Monopoly Model

We extend a simple model of monopoly pricing to include fairness considerations. Customers do not observe the monopoly’s marginal cost but attempt to infer it from the price. With rational or proportional inference, fairness plays no role. But with underinference, fairness affects the profit-maximizing markup in two important ways. First, it increases the price elasticity of demand and thus reduces the markup. Second, it makes the price elasticity of demand increase in the price. As a consequence, the markup falls after an increase in marginal cost, and the passthrough of marginal cost into price is less than one: the price is somewhat rigid. In Appendix B, we extend the monopoly model to allow the firm to credibly reveal its cost, and study how firms strategically reveal cost information to customers.

3.1. Assumptions

We consider a monopoly that sells a good to a representative customer for whom fairness matters. We assume that the firm cannot price-discriminate, so that each unit of the good gets sold at the same price P .

The customer assesses transactional fairness by comparing the purchase price to the perceived marginal cost of production. We assume that the firm’s marginal cost MC is unobservable to buyers—it is the firm’s private information. A buyer who purchases the firm’s good at price P makes an inference about the firm’s marginal cost of production, denoted by $MC^p(P)$; for simplicity, we restrict $MC^p(P)$ to be deterministic. We compare different inference processes below. Having inferred the marginal cost, the buyer deduces that the markup charged by the monopoly is

$$K^p(P) = \frac{P}{MC^p(P)}.$$

The perceived markup determines the transaction’s perceived fairness, which is measured by

$$(1) \quad F(K^p) = \frac{2}{1 + (K^p/K^f)^\theta}.$$

The parameter $\theta \geq 0$ governs the concern for fairness. When $\theta = 0$, the customer does not care about fairness: $F(K^p) = 1$ for any K^p . When $\theta > 0$, he does care about fairness, and the

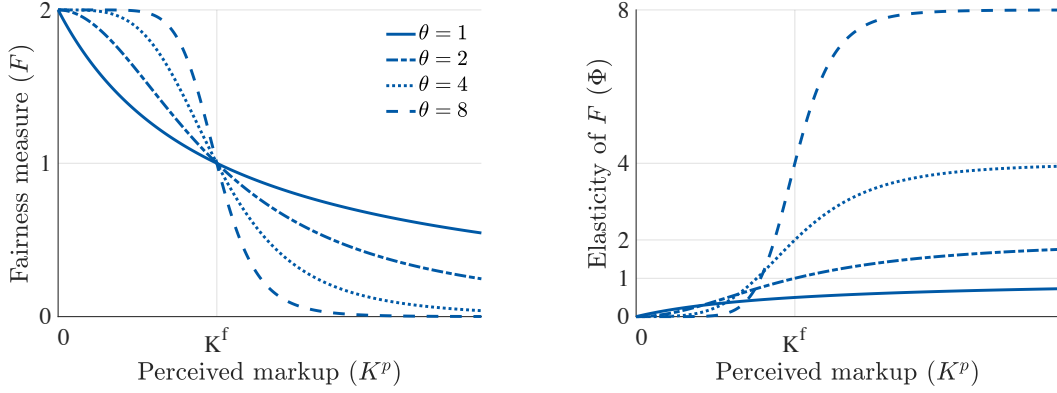


Figure 1. The Fairness Measure and Its Elasticity

higher the perceived markup the less fair the transaction: $F(K^p)$ decreases in K^p . The parameter $K^f > 0$ describes the notional fair markup and has the property that $F(K^f) = 1$. The fairness measure $F(K^p)$ is positive, bounded, decreasing in the perceived markup K^p , with $F(0) = 2$, $F(K^f) = 1$, and $F(\infty) = 0$. In absolute value, the elasticity of the fairness measure with respect to the perceived markup is

$$(2) \quad \Phi(K^p) \equiv -\frac{d \ln(F)}{d \ln(K^p)} = \theta \frac{(K^p/K^f)^\theta}{1 + (K^p/K^f)^\theta}.$$

The elasticity is increasing in the perceived markup, with $\Phi(0) = 0$, $\Phi(K^f) = \theta/2$, and $\Phi(\infty) = \theta$. A useful result is that the elasticity of Φ is $d \ln(\Phi)/d \ln(K^p) = \theta - \Phi$. The fairness measure and its elasticity are plotted in Figure 1. Two properties of the fairness measure F are central to our results: it is decreasing in perceived markup, and its absolute elasticity is increasing. Any fairness measure with these properties would yield the same results; we select the functional form (1) for its analytical tractability.¹⁰

A consumer who buys Y at price P enjoys the fairness-adjusted consumption

$$Z = F(K^p(P)) \cdot Y,$$

For $K^p > K^f$, $F(K^p) < 1$, and this formulation looks as if the customer lost a fraction $1 - F(K^p) > 0$ of each unit of the good due to unfair pricing. Analogously, when $K^p < K^f$, the fairness measure exceeds one, and the consumer enjoys heightened consumption. The higher is θ , the more customers become upset when consuming an overpriced item and content when

¹⁰While not every function $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ that is decreasing with the properties that $f(0) > 0$ and $\lim_{x \rightarrow \infty} f(x) = 0$ has an increasing elasticity as assumed, most of the examples that come to mind do satisfy property.

consuming an underpriced item.

The representative customer has quasilinear utility

$$\frac{\epsilon}{\epsilon - 1} Z^{(\epsilon-1)/\epsilon} + M,$$

which depends on fairness-adjusted consumption Z and money balances M . The parameter $\epsilon > 1$ determines the concavity of the utility function. The customer maximizes utility subject to the budget constraint

$$M + P \cdot Y = I,$$

where $I > 0$ is income.

Finally, the monopoly has constant marginal cost $MC > 0$. Taking marginal cost MC as given, the monopoly chooses the price P and output Y to maximize profits $V = (P - MC) \cdot Y$ subject to the customer's demand for its product.

3.2. Optimal Pricing

We determine the optimal, profit-maximizing price for the monopoly. Given the budget constraint and utility function, the customer chooses Y to maximize

$$\frac{\epsilon}{\epsilon - 1} (F \cdot Y)^{(\epsilon-1)/\epsilon} + I - P \cdot Y.$$

The first-order condition is

$$F \cdot (F \cdot Y)^{-1/\epsilon} = P,$$

which yields the demand curve:

$$(3) \quad Y^d(P) = F(K^P(P))^{\epsilon-1} \cdot P^{-\epsilon}.$$

The price affects demand through two channels. First, the typical substitution effect, captured by $P^{-\epsilon}$. Second, the fairness channel, captured by $F(K^P(P))^{\epsilon-1}$: the price influences the perceived markup and thus the perceived fairness of the transaction; this affects the marginal utility of consumption and hence demand.

Given the demand curve (3), the monopoly chooses P to maximize profits $V = (P - MC) \cdot Y^d(P)$. The first-order condition of the maximization is

$$Y + (P - MC) \cdot \frac{dY^d}{dP} = 0,$$

or equivalently,

$$P - (P - MC) \cdot \frac{-P}{Y} \cdot \frac{dY^d}{dP} = 0.$$

We denote by

$$E \equiv -\frac{d \ln(Y^d)}{d \ln(P)}$$

the price elasticity of demand, in absolute value. The first-order condition then gives that

$$(4) \quad P = \frac{E}{E - 1} MC.$$

Hence, to maximize profits, the monopoly sets its price at a markup $K = E/(E - 1)$ over marginal cost.

To characterize the profit-maximizing markup, we need to determine the price elasticity of demand, E . The firm takes into consideration how its price causes the consumer to substitute as well as how its price influences the customer's inference about markups. In particular, the firm understands that the perceived marginal cost MC^p depends upon P and enters the demand $Y^d(P)$. Using (3), we obtain the price elasticity of demand:

$$(5) \quad E = \epsilon + (\epsilon - 1)\Phi(K^p) \left(1 - \frac{d \ln(MC^p)}{d \ln(P)} \right),$$

where the elasticity $\Phi(K^p)$ is given by (2). The first term ($\epsilon > 0$) describes the standard substitution effect. The second term ($(\epsilon - 1)\Phi(K^p)[1 - (d \ln(MC^p)/d \ln(P))]$) reflects the fairness channel and can be decomposed into two subterms. The first subterm ($(\epsilon - 1)\Phi(K^p) > 0$) appears because a higher price mechanically raises the perceived markup and thus lowers the perceived fairness of the transaction, which reduces demand. The second subterm ($-(\epsilon - 1)\Phi(K^p)(d \ln(MC^p)/d \ln(P)) < 0$) appears because a higher price may also signal a higher marginal cost and thus raises the perceived fairness of the transaction, which in turn increases demand.

3.3. No Fairness Concerns

Before studying the more realistic and interesting case in which customers care about fairness, we briefly examine the benchmark case in which they do not care about fairness.

Without fairness concerns ($\theta = 0$), the fairness measure F is always one, its elasticity Φ is zero, and the price elasticity of demand E is constant, equal to ϵ (see equation (5)). In that case, the profit-maximizing markup takes a standard value of $\epsilon/(\epsilon - 1)$. Since the markup does not

depend on the marginal cost, changes in marginal cost are fully passed through into the price. We denote by

$$\sigma \equiv \frac{d \ln(P)}{d \ln(MC)}$$

the marginal-cost passthrough, which measures the percentage change in price when the marginal cost increases by one percent. Since $P = K \cdot MC$, and here K is independent from MC , the passthrough is one. The following lemma summarizes the findings:

LEMMA 1. *When customers do not care about fairness ($\theta = 0$), the profit-maximizing markup is $K = \epsilon/(\epsilon - 1)$, and the marginal-cost passthrough is $\sigma = 1$.*

3.4. Rational or Proportional Inference

When the marginal cost is unobservable, customers must infer it from the price. With fairness concerns, the equilibrium price then depends upon how customers perform this inference. We consider several inference processes. We begin by analyzing the monopoly's pricing when customers rationally invert the price to uncover the hidden marginal cost.

The monopoly prices according to (4), which specifies that $P = K \cdot MC$ where $K = E/(E - 1)$ is the profit-maximizing markup and MC is the monopoly's marginal cost. Equation (5) shows that the elasticity of demand E depends upon the function $P \mapsto MC^P(P)$, which gives customers' perception of the monopoly's marginal cost. Hence, we can write the profit-maximizing price as the following function of the marginal cost: $MC \mapsto P(MC) = MC \cdot E(MC^P)/(E(MC^P) - 1)$. To uncover the true marginal cost, customers must invert this price function. When the firm follows a separating strategy—charging different prices for different levels of MC —this is easy and yields a function that maps the profit-maximizing price to the true marginal cost:

$$P \mapsto MC(P) = P \cdot \frac{E(MC^P) - 1}{E(MC^P)}.$$

Correctly inverting the firm's price reveals the firm's true marginal cost, so $MC^P(P) = MC(P)$. Therefore, to uncover the true marginal cost by observing the monopoly's price, rational customers need to solve the following functional equation:

$$(6) \quad MC(P) = P \cdot \frac{E(MC) - 1}{E(MC)}.$$

Solving this functional equation yields the function $P \mapsto MC(P)$ that gives the true marginal cost associated with any price.

To solve the functional equation (6), rational customers guess that $MC(P) = P/K^b$, where

K^b is a constant. Under this guess, $d \ln(MC^p)/d \ln(P) = d \ln(MC)/d \ln(P) = 1$ so (5) implies that $E = \epsilon$. The functional equation can be rewritten as $P/K^b = P \cdot (\epsilon - 1)/\epsilon$ for all P . By identification, we find that $K^b = \epsilon/(\epsilon - 1)$. To conclude,

$$MC(P) = \frac{\epsilon - 1}{\epsilon} P$$

is indeed a solution to (6). Rational customers thus form correct beliefs that

$$MC^p(P) = \frac{\epsilon - 1}{\epsilon} P,$$

correctly perceiving the firm's markup to be $K = \epsilon/(\epsilon - 1)$.

A rational customer recognizes that marginal cost is proportional to price and correctly estimates the factor of proportionality, K^b . Note, however, that the conclusion that $E = \epsilon$, and the firm's markup $K = \epsilon/(\epsilon - 1)$ by extension, does not depend upon customers correctly estimating the factor of proportionality. If customers were instead to use the wrong value of K^b —to infer proportionally, but not rationally—then the firm would still price the same as it would if customers did not care about fairness. Indeed, if customers infer proportionally, the perceived marginal cost is

$$MC^p(P) = \frac{P}{K^b}$$

for some $K^b \geq 1$. As $d \ln(MC^p)/d \ln(P) = d \ln(MC)/d \ln(P) = 1$, equation (5) implies that the price elasticity of demand is $E = \epsilon$ and thus the profit-maximizing markup is $\epsilon/(\epsilon - 1)$. Finally, since the markup does not depend on marginal costs with either rational or proportional inference, changes in marginal costs are fully passed through into prices, and the marginal-cost passthrough equals one. The following lemma summarizes these results:

LEMMA 2. *When customers care about fairness ($\theta > 0$), and rationally or proportionally infer marginal costs from price, the profit-maximizing markup in any separating equilibrium is $K = \epsilon/(\epsilon - 1)$, and the marginal-cost passthrough is $\sigma = 1$. The markup and passthrough are the same as in the absence of fairness concerns.*

Without fairness concerns, the price affects demand by determining customers' budget sets. With fairness concerns and hidden marginal costs, the price exerts two additional effects on demand. But with rational or proportional inference, these two effects cancel each other out, explaining why markup and passthrough are the same as without fairness concerns. First, when the purchase price is high relative to the marginal cost of production, customers deem the transaction to be less fair, which reduces customers' marginal utility from consuming the good

and therefore demand for the good. Second, a higher price signals a higher marginal cost, and a higher marginal cost raises the perceived fairness of the transaction and thus the marginal utility of consumption. With rational or proportional inference, the increase in the perceived marginal cost is as large as the observed price increase, so that only the effect of the price on demand through customers' budget sets remains.

3.5. Underinference

A customer who does not sufficiently introspect about the relationship between price and marginal cost will likely underappreciate the information conveyed by the price. We now analyze the market when customers stop short of rational inference by failing to think through how prices signal marginal costs; in particular, they make cost inferences that are subproportional.

We have seen how a rational customer can uncover the marginal cost by inverting the firm's pricing rule. To treat cases in which the customer underinfers about marginal cost from price, we assume that it uses a simple belief-updating rule:

$$(7) \quad MC^p(P) = (MC^b)^\gamma \left(\frac{P}{K^b} \right)^{1-\gamma}.$$

The parameter $MC^b > 0$ denotes the customer's prior belief about the firm's marginal cost. The factor $K^b \geq 1$ represent the customer's perceived factor of proportionality between price and marginal cost. The parameter $\gamma \in [0, 1]$ characterizes the sophistication of the customer's inferences. The case in which $\gamma = 0$ corresponds to the customer believing that the firm uses a proportional markup rule with factor K^b . As we saw in the last section, even if K^b does not coincide with the firm's equilibrium markup, the equilibrium passthrough still equals one. When $\gamma = 1$, the customer fails to update her belief about marginal cost at all from price. In this case, the customer naively maintains her prior belief MC^b . When $\gamma \in (0, 1)$, the customer commits two distinct types of error. First, by placing positive weight on MC^b , the customer fails to learn enough from price. Second, to the extent that the consumer does infer from price, he infers the wrong thing whenever the monopolist does something other than markup with fixed factor K^b .

With the belief-updating rule (7), customers perceive the monopoly's markup to be

$$(8) \quad K^p(P) = (K^b)^{1-\gamma} \left(\frac{P}{MC^b} \right)^\gamma.$$

For $\gamma < 1$, customers do appreciate that a higher price reflects a higher marginal cost, but

do not raise their estimate of the marginal cost sufficiently. Thus, the perceived markup is an increasing function of the observed price: they see higher prices as less fair. Formally, with underinference, an increase in P leads to an increase in $K^P(P)$ (as shown by (8)) and thus to a decrease in $F(K^P)$ (as shown by (1)). Furthermore, since the functions $K^P(P)$ and $F(K^P)$ are differentiable, customers enjoy a small price reduction as much as they dislike a small price increase: the demand curve faced by the monopoly has no kinks.

Although customers correctly perceive the markup as a real variable, underinference ties their estimates of the markup to the nominal variable MC^b . In this way, it induces a specific form of money illusion: the perceived markup now depends on the price level P ; in fact, a higher price causes customers to perceive a higher markup. On the other hand, there is no money illusion with rational or proportional inference because the customer understands that a higher price reflects a higher marginal cost and that the markup is constant.

The combination of fairness and underinference modifies the price elasticity of demand, E , in two important ways. Using (7), we rewrite (5):

$$(9) \quad E = \epsilon + (\epsilon - 1)\gamma\Phi(K^P(P)),$$

where the perceived markup $K^P(P)$ is given by (8). We have seen that without fairness concerns ($\Phi = 0$), or with fairness concerns and proportional inference ($\gamma = 0$), the price elasticity E is constant and equal to ϵ . But with fairness concerns ($\Phi > 0$) and underinference ($\gamma > 0$), things are different: the price elasticity E is always greater than ϵ ; and price elasticity E is increasing in the price P , because $\Phi(K^P)$ and $K^P(P)$ are increasing in K^P and P .

The properties of the price elasticity E under fairness concerns and underinference have direct implications for the markup charged by the monopoly, because the profit-maximizing markup is $K = E/(E - 1)$. The following proposition formalizes these findings:

PROPOSITION 1. *When customers care about fairness ($\theta > 0$), and underinfer marginal costs from prices ($\gamma > 0$), the profit-maximizing markup K is defined by*

$$(10) \quad K = 1 + \frac{1}{\epsilon - 1} \cdot \frac{1}{1 + \gamma\Phi(K^P(K \cdot MC))},$$

implying that $K < \epsilon/(\epsilon - 1)$, and the marginal-cost passthrough is given by

$$(11) \quad \sigma = 1 / \left[1 + \frac{\gamma^2\Phi(\theta - \Phi)}{(1 + \gamma\Phi)(\epsilon + (\epsilon - 1)\gamma\Phi)} \right],$$

implying that $\sigma < 1$. The markup is lower than without fairness concerns or with proportional

inference. And unlike in the cases without fairness concerns or with proportional inference, the passthrough is below one.

The formal proof of the proposition appears in Appendix A, but the intuition behind it is simple. When customers care about the fairness of prices but underinfer marginal cost from price, they become more price-sensitive. Indeed, an increase in the price increases the opportunity cost of consumption—as in the standard case without fairness—and also increases the perceived markup, which reduces the marginal utility of consumption and therefore demand. This heightened price-sensitivity raises E above ϵ and drives the markup below $\epsilon/(\epsilon - 1)$.

Furthermore, after an increase in marginal cost, the monopoly optimally lowers its markup, which produces a marginal-cost passthrough below one—prices are somewhat rigid. This occurs because consumers underappreciate the increase in marginal cost that accompanies a higher price. Since the perceived markup increases, the price elasticity of demand increases. In response, the monopoly reduces its markup, which mitigates the price increase. Thus, our model generates incomplete passthrough of marginal costs into prices, in line with the evidence presented in Section 2.1. Last, although customers believe that transactions are less fair when marginal costs increase, they are wrong: transactions actually become more fair.

In the long-run, consumers may acclimate to prices, and eventually learn marginal costs; consequently, they may come to judge the equilibrium markup as fair. In this case, the markup and passthrough take a simpler form. When $K^P = K^f$, then $\Phi = \theta/2$, which greatly simplifies expressions (10) and (11). To obtain closed-form expressions for markup and passthrough, we consider such equilibria.

COROLLARY 1. *When customers care about fairness ($\theta > 0$), underinfer marginal costs ($\gamma > 0$), and are acclimated (this requires the belief parameters MC^b and K^b to be such that in equilibrium $K^P = K^f$ and $F(K^P) = 1$), the markup and passthrough are*

$$(12) \quad K = 1 + \frac{1}{\epsilon - 1} \cdot \frac{1}{1 + \gamma\theta/2}$$

$$(13) \quad \sigma = 1 \left/ \left[1 + \frac{\gamma^2\theta^2/4}{(1 + \gamma\theta/2)(\epsilon + (\epsilon - 1)\gamma\theta/2)} \right] \right.$$

The markup decreases with the competitiveness of the market (ϵ), concerns for fairness (θ), and inference error (γ). The passthrough increases with the competitiveness of the market (ϵ), but decreases with concerns for fairness (θ) and inference error (γ).

The corollary shows that the passthrough is smaller in less-competitive markets, and approaches one as the market becomes perfectly competitive ($\epsilon \rightarrow \infty$). This implies that prices

are more rigid in less-competitive markets, and that prices become flexible in perfectly competitive markets. This property echoes the finding of Carlton (1986) that prices are more rigid in industries that are more concentrated. It is also consistent with the finding by Amiti, Itskhoki, and Konings (2014) that firms with high market power pass through changes in marginal costs driven by exchange-rate shocks much less than firms with low market power.

Another implication of the corollary is that more concern for fairness and a larger error in inference lead to a lower markup, exactly like a more competitive environment. Moreover, the markup goes to one as the concern for fairness grows arbitrarily large ($\theta \rightarrow \infty$).

Finally, more concern for fairness and a larger error in inference lead to a lower passthrough. This implies that prices are more rigid in fairness-oriented markets. This result accords well with the results reported by Kackmeister (2007), who finds that the fairness of transactions matters less today than it did in 1890 due to weaker current personal relationships between retailers and customers. Kackmeister also documents that retail prices were much more rigid in 1889–1891 than in 1997–1999.

4. New Keynesian Model

We now embed our pricing model into a simple New Keynesian model, thus replacing the standard assumption of Calvo pricing. When customers care about the fairness of prices and underinfer marginal costs from prices, the markup charged by monopolistic firms is not constant but depends on the rate of inflation. This property has several important implications, including the nonneutrality of money.

4.1. Assumptions

The model is dynamic and set in discrete time. The economy is composed of a continuum of households indexed by $j \in [0, 1]$ and a continuum of firms indexed by $i \in [0, 1]$. Households supply labor services, consume goods, and hold riskless nominal bonds. Firms use labor services to produce goods. Since the goods produced by firms are imperfect substitutes for one another, and the labor services supplied by households are also imperfect substitutes, each firm exercises some monopoly power on the goods market, and each household exercises some monopoly power on the labor market.

Fairness Concerns. We introduce concerns for fairness as in Section 3 but generalize the setup to allow for a continuum of firms and goods.

We assume that each firm's technology and, hence, marginal cost are unobservable to other firms and households. When a household purchases good i at price P_i , it infers that firm i 's nominal marginal cost of production is $MC_i^p(P_i)$. Having inferred the marginal cost, the household deduces that the markup charged by firm i is $K_i^p(P_i) = P_i/MC_i^p(P_i)$. This perceived markup determines the perceived fairness of the transaction with firm i , measured by

$$F_i(K_i^p) = \frac{2}{1 + \left(K_i^p/K_i^f\right)^{\theta_i}}.$$

The parameters K_i^f and θ_i may be specific to good i . In absolute value, the elasticity of F_i with respect to K_i^p is

$$\Phi_i(K_i^p) \equiv -\frac{d \ln(F_i)}{d \ln(K_i^p)} = \theta_i \frac{\left(K_i^p/K_i^f\right)^{\theta_i}}{1 + \left(K_i^p/K_i^f\right)^{\theta_i}}.$$

An amount Y_{ij} of good i bought by household j at a unit price P_i yields a fairness-adjusted consumption $Z_{ij} = F_i(K_i^p(P_i))Y_{ij}$. Household j 's fairness-adjusted consumption of the different goods aggregates into a consumption index

$$(14) \quad Z_j = \left(\int_0^1 Z_{ij}^{(\epsilon-1)/\epsilon} di \right)^{\epsilon/(\epsilon-1)},$$

where $\epsilon > 1$ is the elasticity of substitution between different goods, which describes the household's love of variety: as $\epsilon \rightarrow \infty$, goods become perfect substitutes.

Finally, the fairness-adjusted price index

$$(15) \quad Q = \left[\int_0^1 \left(\frac{P_i}{F_i(K_i^p(P_i))} \right)^{1-\epsilon} di \right]^{1/(1-\epsilon)}$$

represents the price of one unit of Z_j .

Inference About Marginal Costs. The dynamic model has the advantage over the static model of providing a natural candidate for the nominal anchor that households use to infer nominal marginal costs. Whereas the static model takes as its nominal anchor a parameter (MC^b), the dynamic model uses the current perception of nominal marginal cost. Households' perceptions

of firm i 's nominal marginal cost evolve according to

$$(16) \quad MC_i^p(t) = (MC_i^p(t-1))^\gamma \left(\frac{P_i(t)}{K_i^b} \right)^{1-\gamma}.$$

In the law of motion, $MC_i^p(t-1)$ is last period's perception of the marginal cost, $P_i(t)/K_i^b$ is the marginal cost under proportional inference, and $\gamma \in [0, 1]$ measures the sophistication of households' inference. With $\gamma = 0$, then $MC_i^p(t) = P_i(t)/K_i^b$ so households set their beliefs about marginal costs in proportion to observed prices. With $\gamma = 1$, then $MC_i^p(t)$ is constant so households fail to update their beliefs about marginal costs. With $\gamma \in (0, 1)$, households partially adjust their beliefs in the direction of the true nominal marginal cost.

Following the same logic as in the static model of Section 3, we can show that with rational inference the nominal marginal cost remains $MC_i(P_i(t)) = P_i(t) \cdot (\epsilon - 1)/\epsilon$; furthermore, rational households can back out this cost by following the same strategy as in the static model.¹¹ Hence, rational inference is a special case of (16) with $\gamma = 0$ and $K_i^b = \epsilon/(\epsilon - 1)$.

Households. Households work, own the firms, spend part of their income on consumption, and save the rest of their income in riskless nominal bonds. Households derive utility from consuming goods and disutility from working. The utility depends on the fairness-adjusted consumption index Z_j and the amount N_j of labor supplied. Household j 's utility at time 0 is

$$(17) \quad \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\ln(Z_j) - \frac{N_j(t)^{1+\eta}}{1+\eta} \right)$$

where \mathbb{E}_0 is the expectation conditional on period-0 information, $\beta > 0$ is the time discount factor, and $\eta > 0$ is the inverse of the Frisch elasticity of labor supply.

Households sell or buy one-period bonds. Household j holds $B_j(t)$ bonds in period t . Bonds purchased in period t have a price $X(t)$, mature in period $t + 1$, and pay one unit of money at maturity. Bonds are traded on a perfectly competitive market at a price determined by monetary policy.

Household j 's budget constraint in period t is

$$(18) \quad \int_0^1 P_i(t) Y_{ij}(t) di + X(t) B_j(t) = W_j(t) N_j(t) + B_j(t-1) + V_j(t),$$

¹¹Under rational inference, the firm's optimization problem reduces to a collection of static optimization problems: at each time t , the firm maximizes current profits. The firm's problem therefore coincides with that in the static model of Section 3, and rational households can follow the same strategy of solving a functional equation at each time t .

where $V_j(t)$ are dividends from ownership of firms. In addition, household j is subject to a solvency constraint preventing Ponzi schemes: $\lim_{T \rightarrow \infty} \mathbb{E}_t [B_j(T)] \geq 0$ for all t .

Household j maximizes the utility (17) by choosing sequences for the nominal wage of labor service j , the amount of labor service j supplied, the amounts of consumption of goods $i \in [0, 1]$, and the amount of bonds held, $\left\{ W_j(t), N_j(t), [Y_{ij}(t)]_{i=0}^1, B_j(t) \right\}_{t=0}^{\infty}$. The maximization is subject to the flow budget constraint (18), to the solvency condition, and to the constraint imposed by firms' demand for labor service j . The household takes as given the initial endowment of bonds $B_j(-1)$, and the sequences for prices and dividends, $\left\{ X(t), [P_i(t)]_{i=0}^1, V_j(t) \right\}_{t=0}^{\infty}$.

Firms. Firm i hires labor to produce output using the production function

$$(19) \quad Y_i(t) = A_i(t)N_i(t)^\alpha,$$

where $Y_i(t)$ is its output of good i , $A_i(t)$ is its technology level, $\alpha < 1$ is the extent of diminishing marginal returns to labor, and

$$(20) \quad N_i(t) = \left(\int_0^1 N_{ij}(t)^{(\nu-1)/\nu} dj \right)^{\nu/(\nu-1)}$$

is an employment index. In the employment index, $N_{ij}(t)$ is the quantity of labor service j hired by firm i , and $\nu > 1$ is the elasticity of substitution between different labor services. The level of technology $A_i(t)$ is exogenous, possibly stochastic, and is unobservable to households—making the firm's marginal cost unobservable.

Firm i chooses sequences for the price of good i , the output of good i , and the amounts of labor services employed, $\left\{ P_i(t), Y_i(t), [N_{ij}(t)]_{j=0}^1 \right\}_{t=0}^{\infty}$, in order to maximize the present-discounted value of profits

$$(21) \quad \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \Gamma(t) \left(P_i(t)Y_i(t) - \int_0^1 W_j(t)N_{ij}(t)dj \right) \right],$$

where

$$(22) \quad \Gamma(t) \equiv \beta^t \cdot \frac{Q(0)}{Q(t)} \cdot \frac{Z(0)}{Z(t)}$$

is the stochastic discount factor for nominal payoffs in period t . The maximization is subject to the production constraint (19), to the demand for good i , and to the law of motion of the perceived marginal cost, given by (16). All the firm's profits are rebated to households.

Monetary Policy. The nominal interest rate is determined by monetary policy, which follows a simple rule:

$$(23) \quad i(t) = i_0(t) + \mu\pi(t),$$

where $i_0(t) > 0$ is exogenous and possibly stochastic, and the parameter $\mu > 0$ gives the response of the interest rate to inflation.

4.2. Optimal Pricing

We now characterize the optimal pricing for firms. Once this is done, it is easy to completely describe the equilibrium. The derivations in the next subsections are mostly standard and relegated to Appendix C.

To maximize their utility, households make two decisions: first, they choose how to divide their wealth across goods and bonds; second, they determine the wage for their labor service. Integrating over all households, we find that the demand for good i is given by

$$(24) \quad Y_i^d(t, P_i(t), MC_i^p(t-1)) = Z(t) \cdot F \left([K^b]^{1-\gamma} \left[\frac{P_i(t)}{MC_i^p(t-1)} \right]^\gamma \right)^{\epsilon-1} \cdot \left(\frac{P_i(t)}{Q(t)} \right)^{-\epsilon}.$$

where $Z(t) \equiv \int_0^1 Z_j(t) dj$ describes the level of aggregate demand. The demand increases with aggregate demand, Z , and decreases with the price of good i , P_i . The equation can be written as $Z_i^d \equiv F \cdot Y_i^d = Z \cdot [(P_i/F)/Q]^{-\epsilon}$. As the price of one unit of Z_i is P_i/F and the price of one unit of Z is Q , the relative price of Z_i is $(P_i/F)/Q$. Hence, this alternative formulation says that the demand for Z_i equals aggregate demand Z times the relative price of Z_i to the power of $-\epsilon$. This is the standard expression for demand curves in this type of models.

We also find that it is optimal for household j to smooth fairness-adjusted consumption over time according to an Euler equation:

$$X(t) = \beta \mathbb{E}_t \left[\frac{Q(t)Z_j(t)}{Q(t+1)Z_j(t+1)} \right]$$

The analysis will focus on symmetric equilibria: all households receive the same dividends; all firms share a common technology; all households post the same wage; and all firms set the same price. In such an equilibrium, all parameters and variables are the same for all households and firms, so we will drop the subscripts i and j on parameters and variables. In particular, in a symmetric equilibrium, $Z_j(t) = F(t)Y(t)$ and $Q(t) = P(t)/F(t)$. In such an equilibrium, the

Euler equation simplifies to

$$(25) \quad X(t) = \beta \mathbb{E}_t \left[\frac{P(t)Y(t)}{P(t+1)Y(t+1)} \right].$$

This is the standard consumption Euler equation.

Next, the demand for labor service j from firms is

$$(26) \quad N_j^d(t, W_j(t)) = N(t) \cdot \left(\frac{W_j(t)}{W(t)} \right)^{-\nu},$$

where

$$(27) \quad W(t) \equiv \left(\int_0^1 W_j(t)^{1-\nu} dj \right)^{1/(1-\nu)}$$

is the nominal wage index and $N(t) \equiv \int_0^1 N_i(t) di$ is aggregate employment. The labor demand increases with aggregate employment but decreases with the relative wage of labor service j , W_j/W . Given labor demand (26), household j sets its wage $W_j(t)$ to maximize utility. The optimal wage satisfies

$$\frac{W_j(t)}{Q(t)} = \frac{\nu}{\nu - 1} N_j(t)^\eta Z_j(t).$$

As the wage of labor service j is W_j and the price of one unit of fairness-adjusted consumption is Q , the real wage of labor service j is W_j/Q . Hence, household j sets its real wage at a markup of $\nu/(\nu - 1) > 1$ over its marginal rate of substitution between leisure and fairness-adjusted consumption, $N_j^\eta Z_j$. In the symmetric case, $Z_j(t) = F(t)Y(t)$ and $Q(t) = P(t)/F(t)$, so the previous equation simplifies to

$$(28) \quad \frac{W(t)}{P(t)} = \frac{\nu}{\nu - 1} N(t)^\eta Y(t).$$

To maximize profits, firms also make two decisions: first, they choose how much of each labor service to hire; second, they determine the price of their good. Integrating over all firms, we find that the demand for labor service j is given by (26).

Next, we turn to firm i 's pricing. Let E_i be the price elasticity of the demand for good i , in absolute value:

$$(29) \quad E_i(t) = -\frac{\partial \ln(Y_i^d)}{\partial \ln(P_i)} = \epsilon + (\epsilon - 1)\gamma\Phi(K_i^P(t)),$$

where the elasticity Φ is given by (2). Although the demand is not the same as in the static model, the expression for E_i remains the same (see (9)). In the static model the profit-maximizing markup is given by $K_i = E_i/(E_i - 1)$. In the dynamic model, however, we will see that the profit-maximizing markup is not necessarily given by $E_i/(E_i - 1)$, since E_i does not capture the effect of P_i on future perceived marginal costs and thus future demands.

When firm i charges a price $P_i(t)$ and faces a nominal marginal cost $MC_i(t)$, then the markup charged by firm i is

$$K_i(t) = \frac{P_i(t)}{MC_i(t)}.$$

We define the quasi elasticity $D_i(t) > 1$ by

$$(30) \quad D_i(t) = \frac{K_i(t)}{K_i(t) - 1}.$$

In the static model, when the price is optimal, $D_i(t) = E_i(t)$. In the dynamic model, the gap between $D_i(t)$ and $E_i(t)$ indicates how much the slow adjustment of customers' beliefs matters. When firms change their price, they affect perceived marginal costs today and in the future (through (16)); thus, the price today affects demands in the future. This effect is not captured by $E_i(t)$ but is captured by $D_i(t)$.

We find that in a symmetric equilibrium, to maximize profits, firms should set prices such that

$$(31) \quad \beta \mathbb{E}_t \left[\frac{E(t+1) - (1-\gamma)\epsilon}{D(t+1)} \right] + (1-\gamma\beta) = \frac{E(t)}{D(t)}.$$

This forward-looking equation gives the quasi elasticity $D(t)$ when prices are optimal, and thus the profit-maximizing markup $K(t) = D(t)/(D(t) - 1)$.¹²

Finally, in a symmetric equilibrium, all real variables are determined by the goods-market markup. We establish these links here. We begin by computing marginal costs. The nominal marginal cost is the nominal wage divided by the marginal product of labor: $MC(t) = W(t)/(\alpha A(t)N(t)^{\alpha-1})$. Using (28) and (19), we obtain

$$\frac{MC(t)}{P(t)} = \frac{\nu}{(\nu-1)\alpha} N(t)^{1+\eta}.$$

The nominal marginal cost increases with employment because the real wage increases with employment and the production function has diminishing marginal returns to labor. The nominal

¹²Appendix C shows that the equation admits a slightly more complicated expression in an equilibrium that is not symmetric.

marginal cost also increases with the labor-market markup, $\nu/(\nu - 1)$. The goods-market markup is the price over the marginal cost: $K(t) = P(t)/MC(t)$. Thus, the previous equation implies that the goods-market market directly govern employment:

$$(32) \quad N(t)^{1+\eta} = \frac{(\nu - 1)\alpha}{\nu} \cdot \frac{1}{K(t)}.$$

Then, employment determines output and real wage through (28) and (19). Finally, nominal profits in period t are turnover minus wage bill: $V(t) = P(t)Y(t) - W(t)N(t)$. Moreover,

$$\frac{1}{K(t)} = \frac{MC(t)}{P(t)} = \frac{W(t)/P(t)}{\alpha A(t)N(t)^{\alpha-1}} = \frac{W(t)/P(t)}{\alpha Y(t)/N(t)}.$$

Hence real profits are governed by the goods-market markup:

$$(33) \quad \frac{V(t)}{P(t)} = Y(t) \cdot \left(1 - \frac{\alpha}{K(t)}\right).$$

4.3. Steady-State Equilibrium

We first describe the steady-state equilibrium in a symmetric case. It is convenient to manipulate all the variables in log form. We use the following notation: for a variable $C(t)$, we denote the log of $C(t)$ by $c(t) \equiv \ln(C(t))$, and the steady-state values of $C(t)$ and $c(t)$ by \bar{C} and \bar{c} . Following common practice, we define the inflation rate between t and $t + 1$ as $\pi(t + 1) = \ln(P(t + 1)/P(t))$ and the nominal interest between t and $t + 1$ as $i(t) = -\ln(X(t))$.¹³ Finally, we define the real interest rate as $r(t) = i(t) - \pi(t)$ and the time discount rate as $\rho = -\ln(\beta)$. In steady state, all real variables are constant and all nominal variables grow at a constant rate $\bar{\pi}$.

Since $Y(t) = Y(t + 1)$ in steady state, (25) implies

$$(34) \quad \bar{i} = \bar{\pi} + \rho.$$

Hence, in steady state, the nominal interest rate is the time discount rate plus the inflation rate. In other words, the real interest rate equals the time discount rate: $\bar{r} \equiv \bar{i} - \bar{\pi} = \rho$.

¹³If $P(t + 1)$ and $P(t)$ are close enough, then $\pi(t + 1) \approx (P(t + 1) - P(t))/P(t)$, which is another common way of defining inflation. If $X(t) \approx 1$, then $i(t) \approx (1 - X(t))/X(t)$, which is the yield of a bond purchased at time t , and is another common way of defining the interest rate. Both $\pi(t + 1)$ and $i(t)$ describe what happens between periods t and $t + 1$, but we index the inflation rate with $t + 1$ and the interest rate with t because the inflation rate is determined in period $t + 1$ while the interest rate is determined in period t .

Equation (34) combined with the monetary policy rule (23) implies that

$$(35) \quad \bar{\pi} = \frac{\rho - \bar{i}_0}{\mu - 1}.$$

Hence steady-state inflation rate is determined by the exogenous component of the monetary-policy rule, \bar{i}_0 . Inflation is higher when the exogenous component is lower.

Then, (16) shows that in steady state the perceived markup is determined by inflation:

$$(36) \quad \bar{k}^p = k^b + \frac{\gamma}{1 - \gamma} \bar{\pi}.$$

Households perceive higher markups when inflation is higher, and the response of the perceived markup to inflation is stronger with higher underinference. Of course in steady state, prices grow at the inflation rate, and the perceived marginal cost also grows at the inflation rate. Hence, each period, customers adjust their perception of nominal marginal costs by the right amount. What they get wrong is the level of nominal marginal costs; this error is induced by inflation.

Next, combining equations (29), (30), and (31), we obtain the steady-state relationship between the goods-market markup and the rate of inflation:

PROPOSITION 2. *Without fairness concerns ($\theta = 0$) or with proportional inference ($\gamma = 0$), the steady-state goods-market markup is $\bar{K} = \epsilon/(\epsilon - 1)$. But with fairness concerns ($\theta > 0$) and underinference ($\gamma > 0$), the steady-state goods-market markup is*

$$(37) \quad \bar{K} = 1 + \frac{1}{\epsilon - 1} \cdot \frac{1}{1 + \frac{(1-\beta)\gamma}{1-\beta\gamma} \Phi(K^p(\bar{\pi}))},$$

where the elasticity Φ is given by (2), the perceived markup $K^p(\bar{\pi})$ by (36), and steady-state inflation $\bar{\pi}$ by (35). This implies that $\bar{K} < \epsilon/(\epsilon - 1)$ and that \bar{K} is decreasing in $\bar{\pi}$.

Equation (37) is the counterpart to (10) in the static model. The two equations have the same structure. Since $\gamma \mapsto (1 - \beta)\gamma/(1 - \beta\gamma)$ is increasing from 0 to 1 when γ increases from 0 to 1, for any $\gamma \in [0, 1]$, the static model with inference parameter $\gamma^s = (1 - \beta)\gamma/(1 - \beta\gamma)$ achieves the same allocation as the steady state of the dynamic model with inference parameter γ .

Proposition 2 gives the goods-market markup in steady state. Steady-state employment, output, real wage, and real profits are directly governed by this markup, through the employment equation (32), the production constraint (19), the wage-setting equation (28), and the equation for real profits (33). Using these equations and Proposition 2, we characterize the effect of steady-state inflation on real variables and the shape of the long-run Phillips curve, which links steady-state inflation to steady-state employment.

COROLLARY 2. *Without fairness concerns ($\theta = 0$) or with proportional inference ($\gamma = 0$), the steady-state goods-market markup is independent of inflation. Thus money is superneutral: steady-state inflation has no effect on steady-state employment, output, fairness measure, real wages, and real profits. As a consequence, the long-run Phillips curve is vertical:*

$$\bar{n} = \frac{1}{1 + \eta} \left[\ln(\alpha) - \ln\left(\frac{\nu}{\nu - 1}\right) - \ln\left(\frac{\epsilon}{\epsilon - 1}\right) \right].$$

With fairness concerns ($\theta > 0$) and underinference ($\gamma > 0$), the steady-state goods-market markup is decreasing in steady-state inflation. Thus money is not superneutral: in steady state, higher inflation leads to higher employment, higher output, lower fairness measure, higher real wages, but lower real profits when $K < 1 + \alpha + \eta$. As a consequence, the long-run Phillips curve is upward sloping:

$$(38) \quad \bar{n} = \frac{1}{1 + \eta} \left[\ln(\alpha) - \ln\left(\frac{\nu}{\nu - 1}\right) - \ln\left(1 + \frac{1}{\epsilon - 1} \cdot \frac{1}{1 + \frac{(1-\beta)\gamma}{1-\beta\gamma} \Phi(K^p(\bar{\pi}))}\right) \right].$$

The proof is not complicated but a bit long, so we relegate it to Appendix A. The effect of inflation on real profits depends on parameter values. In the case $K < 1 + \alpha + \eta$, real profits fall when inflation is higher. This is the case that seems the most relevant in practice: since most estimates of K are less than 2, α is typically above 2/3, and microestimates of η are above 1 (see Section 4.5).

The superneutrality of money is the property that the steady-state inflation rate has no influence on the steady-state levels of real variables. In the model, an increase in steady-state inflation is engineered by reducing the exogenous component of the monetary-policy rule, \bar{i}_0 (see equation (35)). We find that without fairness concerns or with proportional inference, then money is superneutral. Of course, in that case, the long-run Phillips curve is vertical: steady-state employment is independent of steady-state inflation.

On the other hand, if households care about fairness and underinfer, then money is not superneutral. In that case, after an increase in steady-state inflation, households underappreciate the increase in nominal marginal costs, so they attribute the higher prices partly to higher nominal marginal costs and partly to higher markups, which they find unfair. Since the perceived fairness of the transactions on the goods market decreases, the elasticity of the demand for goods increases. In response, firms reduce their markups. We have showed that in equilibrium, employment is a decreasing function of the markup; therefore, a lower markup implies higher employment, which in turn implies higher output. After an increase in inflation, households mistakenly believe that transactions on the goods market are less fair, but transactions are in fact

more fair (since markups are lower), and firms actually suffer lower real profits.

With fairness concerns and underinference, the long-run Phillips curve is not vertical but upward sloping. This property of the model is consistent with evidence that higher average inflation leads to lower average unemployment (for example, King and Watson 1994).¹⁴ In our model, in steady state, higher inflation leads to higher employment because it reduces goods-market markups. There is also evidence that this mechanism operates. Benabou (1992) finds that in the US retail sector for 1948–1990, higher average inflation leads to lower average markup. Using aggregate US data for 1953–2000, Banerjee and Russell (2005) also find that higher average inflation leads to lower average goods-market markup.

Our mechanism complements the traditional mechanism for an upward-sloping long-run Phillips curve: that because of downward nominal wage rigidity, steady-state inflation erodes real wages and thus reduces unemployment (Tobin 1972; Akerlof, Dickens, and Perry 1996; Benigno and Ricci 2011). While our mechanism operates on the goods market instead of the labor market, the psychological origin of the two mechanisms could be similar, since one possible source of wage rigidity is fairness concerns of workers.

4.4. Equilibrium Dynamics

We now analyze equilibrium dynamics, working with the deviations of the log variables from their steady-state values for convenience. For a variable $C(t)$ other than the interest or inflation rate, we denote the log-deviation of $C(t)$ from its steady-state value by $\widehat{c}(t) \equiv \ln(C(t)) - \ln(\bar{C})$. For inflation and interest rates, we denote the deviation (not log-deviation) from steady state by $\widehat{\pi}(t) \equiv \pi(t) - \bar{\pi}$, $\widehat{i}_0(t) \equiv i_0(t) - \bar{i}_0$, and $\widehat{r}(t) = r(t) - \bar{r}$.

The first equilibrium condition is the law of motion of the perceived markup, which derives from the inference mechanism (16):

$$(39) \quad \widehat{k}^p(t) = \gamma \left[\widehat{\pi}(t) + \widehat{k}^p(t-1) \right].$$

This equation shows that the perceived markup today tends to be high if inflation is high or if the past perceived markup was high. Past beliefs matter because people use them as baseline for their current beliefs. Inflation matters because people do not fully appreciate the effect of inflation on nominal marginal costs.

The second condition is the *dynamic IS equation*, obtained by combining the Euler equa-

¹⁴See also King and Watson (1997) for a thorough discussion of the empirical specifications under which superneutrality can be rejected.

tion (25) with the monetary-policy rule (23):

$$(40) \quad \alpha \widehat{n}(t) + \mu \widehat{\pi}(t) = \alpha \mathbb{E}_t [\widehat{n}(t+1)] + \mathbb{E}_t [\widehat{\pi}(t+1)] - \widehat{i}_0(t) - \widehat{a}(t) + \mathbb{E}_t [\widehat{a}(t+1)].$$

Here the dynamic IS equation involves the log-deviation of employment, \widehat{n} , although usually it involves the output gap, which is the gap between the actual level of output and the level of output when the markup is at its long-run level (in the standard New Keynesian model, this is also the level of output when prices are flexible). However, the log-deviation of employment and output gap are related by $y(t) - y^n(t) = \alpha \widehat{n}(t)$, so using one or the other is equivalent.¹⁵

The final condition is the *short-run Phillips curve*, obtained from optimal pricing equation (31):

$$(41) \quad (1 - \beta\gamma)\gamma \widehat{k}^p(t-1) + (1 - \beta\gamma)\gamma \widehat{\pi}(t) - \Lambda_1 \widehat{n}(t) = \beta\gamma \mathbb{E}_t [\widehat{\pi}(t+1)] - \Lambda_2 \mathbb{E}_t [\widehat{n}(t+1)],$$

where

$$\Lambda_1 \equiv (1 + \eta) \frac{\epsilon + (\epsilon - 1)\gamma \overline{\Phi}}{(\theta - \overline{\Phi})\gamma \overline{\Phi}} \left[1 + \frac{(1 - \beta)\gamma \overline{\Phi}}{1 - \beta\gamma} \right]$$

$$\Lambda_2 \equiv (1 + \eta)\beta \frac{\epsilon + (\epsilon - 1)\overline{\Phi}}{(\theta - \overline{\Phi})\overline{\Phi}} \left[1 + \frac{(1 - \beta)\gamma \overline{\Phi}}{1 - \beta\gamma} \right].$$

Just as the long-run Phillips curve relates steady-state inflation to steady-state employment, the short-run Phillips curve relates current inflation and to current employment. In addition to current inflation and employment, this Phillips curve also incorporates the expectations of inflation and employment, a feature of New Keynesian models (see Galí 2008, p. 49). Moreover, the Phillips curve includes a backward-looking element, $\widehat{k}^p(t-1)$, which appears due to customers' backward-looking perceptions of marginal costs. In the standard New Keynesian model, the Phillips curve excludes backward-looking elements, although earlier Phillips curves included them.

Using (39), we can write $\widehat{k}^p(t-1)$ as a function of past inflation rates:

$$\widehat{k}^p(t) = \sum_{i=0}^{+\infty} \gamma^{i+1} \widehat{\pi}(t-i).$$

¹⁵Equation (19) implies that log output and log employment are related by $y(t) = a(t) + \alpha n(t)$. Moreover, when the markup is at its long-run level, employment also is at its long-run level, so that the log of the natural level of output is $y^n(t) = a(t) + \alpha \bar{n}$. Thus, the output gap is directly determined by the log-deviation of employment: $y(t) - y^n(t) = \alpha \widehat{n}(t)$. The output gap is negative whenever employment is below its long-run level.

Combining this result with the expression of the short-run Phillips curve offers an alternative formulation of the Phillips curve that highlights the presence of past inflation rates:

$$(1 - \beta\gamma)\gamma \sum_{i=0}^{+\infty} \gamma^i \widehat{\pi}(t - i) - \Lambda_1 \widehat{n}(t) = \beta\gamma \mathbb{E}_t [\widehat{\pi}(t + 1)] - \Lambda_2 \mathbb{E}_t [\widehat{n}(t + 1)].$$

Thus, the short-run Phillips curve in our model is a hybrid of backward- and forward-looking elements. This property accords well with available estimates, which tend to indicate that both lagged inflation and expected future inflation enter significantly in the hybrid Phillips curve (Mavroeidis, Plagborg-Moller, and Stock 2014, Table 2). Although the short-run Phillips curve in the standard New Keynesian model is purely forward-looking, backward-looking components can be appended to it. For example, Christiano, Eichenbaum, and Evans (2005) assume that the firms that are not able to reset their price at time t simply index their price to past inflation.

Combining (39), (40), and (41), and performing some algebra yields the system of difference equations characterizing the equilibrium:

$$(42) \quad \begin{bmatrix} \widehat{k}^p(t) \\ \mathbb{E}_t [\widehat{\pi}(t + 1)] \\ \mathbb{E}_t [\widehat{n}(t + 1)] \end{bmatrix} = \mathbf{A} \begin{bmatrix} \widehat{k}^p(t - 1) \\ \widehat{\pi}(t) \\ \widehat{n}(t) \end{bmatrix} + \mathbf{B} \cdot \epsilon(t)$$

where

$$\mathbf{A} \equiv \begin{bmatrix} \gamma & \gamma & 0 \\ \frac{(1-\beta\gamma)\alpha\gamma}{\Lambda_2+\alpha\beta\gamma} & \frac{\Lambda_2\mu+\alpha\gamma(1-\beta\gamma)}{\Lambda_2+\alpha\beta\gamma} & \frac{(\Lambda_2-\Lambda_1)\alpha}{\Lambda_2+\alpha\beta\gamma} \\ \frac{-(1-\beta\gamma)\gamma}{\Lambda_2+\alpha\beta\gamma} & \frac{[\beta(\mu+\gamma)-1]\gamma}{\Lambda_2+\alpha\beta\gamma} & \frac{\Lambda_1+\alpha\beta\gamma}{\Lambda_2+\alpha\beta\gamma} \end{bmatrix}, \quad \mathbf{B} \equiv \begin{bmatrix} 0 \\ \frac{\Lambda_2}{\Lambda_2+\alpha\beta\gamma} \\ \frac{\beta\gamma}{\Lambda_2+\alpha\beta\gamma} \end{bmatrix},$$

and $\epsilon(t) \equiv \widehat{i}_0(t) + \widehat{a}(t) + \mathbb{E}_t [\widehat{a}(t + 1)]$ is an exogenous shock realized at time t . The system (42) determines employment $\widehat{n}(t)$, inflation $\widehat{\pi}(t)$, and perceived markup $\widehat{k}^p(t)$. The other variables are directly obtained from these three variables. By log-linearizing (32), (19), (28), and (33), we find that the goods-market markup is given by $\widehat{k}(t) = -(1+\eta)\widehat{n}(t)$, output by $\widehat{y}(t) = \widehat{a}(t) + \alpha\widehat{n}(t)$, the real wage by $\widehat{w}(t) - \widehat{p}(t) = (\eta + \alpha)\widehat{n}(t) + \widehat{a}(t)$, and real profits by $\widehat{v}(t) - \widehat{p}(t) = \widehat{y}(t) + \left[\alpha / (\overline{K} - \alpha) \right] \widehat{k}(t)$. Last, the nominal and real interest rates are $\widehat{i}(t) = \widehat{i}_0(t) + \mu\widehat{\pi}(t)$ and $\widehat{r}(t) = \widehat{i}_0(t) + (\mu - 1)\widehat{\pi}(t)$.

4.5. Calibration

We calibrate our model: for standard parameters, we use the usual empirical evidence; for the new fairness and inference parameters, we use evidence on passthrough. Table 3 summarizes the calibrated values of the parameters.

We start by calibrating the three parameters central to our theory: the fairness parameter θ , the inference parameter γ , and the elasticity of substitution across goods ϵ . These three parameters jointly determine the average goods-market markup and markup dynamics in response to shocks—which determine passthrough dynamics. Hence, to calibrate θ , γ , and ϵ , we use empirical evidence on markups and passthrough dynamics.

Using firm-level data for the US economy between 1950 and 2014, De Loecker and Eeckhout (2017) provides compelling evidence on the goods-market markup in the United States. They find that the average markup hovers between 1.2 and 1.3 in the 1950–1980 period and rises from 1.2 to 1.7 in the 1980–2014 period. The average value of the markup since 2000 is about 1.5; we use this value as steady-state markup. Then, equation (37) with $\bar{K} = 1.5$ gives a first condition that links θ , γ , and ϵ .

The markups computed by De Loecker and Eeckhout are consistent with earlier evidence for the United States, as reviewed in Rotemberg and Woodford (1995). The marketing literature estimates the markup for a typical product to be below 2, and the industrial organization literature finds markups range between 1.2 and 1.7 (p. 261, p. 266). Several papers using US microdata provide similar estimates of the markup: 1.3 in the automobile industry (Berry, Levinsohn, and Pakes 1995, p. 882), 1.6 in the ready-to-eat cereal industry (Nevo 2001, Table 8, column 1), and 1.6 in the coffee industry (Nakamura and Zerom 2010, Table 6).¹⁶ Finally, Barsky et al. (2003, p. 166) discover that for most national-brand items retailed in supermarkets, markups range between 1.4 and 2.1.

To finish calibrating θ , γ , and ϵ , we turn to the empirical evidence on passthrough. We have seen in Section 2.1 that a median estimate of the short-run passthrough is 0.5. This estimate imposes a second condition on ϵ , θ and γ . To obtain a third and final condition, we use an estimate of the long-run passthrough. The inference parameter γ determines how quickly people learn about marginal costs: it therefore determines how quickly the passthrough increases toward complete passthrough. Burstein and Gopinath (2014, Table 7.4) discover short-run and long-run passthroughs for import prices in response to exchange-rate fluctuations in the United States and seven other countries. The short-run passthroughs are measured on impact while the long-run passthroughs are measured 2 years after the exchange-rate shock. On average, the short-run passthrough is 0.45 while the long-run passthrough is 0.69. Hence, after 2 years, the passthrough increases by $(0.69 - 0.45)/0.45 = 55\%$ compared to the instantaneous passthrough. On the other hand, Benzarti and Carloni (2016, Figure 5) find that even after more than two years, the passthrough of value-added tax into restaurant prices remains broadly equal to the instantaneous

¹⁶Berry, Levinsohn, and Pakes (1995) find that on average $(P - MC)/P = 0.239$, which translates into a markup of $K = P/MC = 1 - 1/0.239 = 1.3$. Nevo (2001) finds that a median estimate of $(P - MC)/P$ is 0.372, which translate into a median markup estimate of $K = P/MC = 1 - 1/0.372 = 1.6$.

Table 3. Parameter Values in Simulations

| Value | Description | Source or target |
|--|--|--|
| A. Common parameters | | |
| $\beta = 0.99$ | Quarterly discount factor | Annual rate of return = 4% |
| $\alpha = 1$ | Marginal returns to labor | Labor share = 2/3 |
| $\eta = 1.1$ | Inverse of Frisch elasticity of labor supply | Chetty et al. (2013, Table 2) |
| $\mu = 1.5$ | Response of interest rate to inflation | Gali (2008, p. 52) |
| $\zeta = 3/4$ | Persistence of interest-rate shock | Gali (2008, p. 52), Gali (2011, p. 26) |
| B. Parameters of the New Keynesian model with fairness | | |
| $\epsilon = 2.1$ | Elasticity of substitution across goods | Steady-state markup = 1.5 |
| $\theta = 25$ | Fairness concern | Short-run passthrough = 0.5 |
| $\gamma = 0.87$ | Amount of underinference | Long-run passthrough = 0.7 |
| C. Parameters of the standard New Keynesian model | | |
| $\epsilon = 3$ | Elasticity of substitution across goods | Steady-state markup = 1.5 |
| $\kappa = 2/3$ | Share of firms keeping price unchanged | Average price duration = 3 quarters |

passthrough. For our calibration, we consider an intermediate passthrough adjustment: we set the two-year passthrough to be 40% higher than the instantaneous passthrough. Applying this coefficient to our short-run passthrough of 0.5, we obtain an estimate of the long-run passthrough of $0.5 \times 1.4 = 0.7$.

We then take the perspective of one firm from our New Keynesian model, and simulate passthrough dynamics in response to an exogenous increase in the firm's marginal costs, keeping everything else constant. (Appendix C describes the firm's problem and pricing, and the simulation used to calibrate the parameters.) The fairness parameter θ primarily affects the level of passthrough while the inference parameter γ affects how the passthrough evolves over time. Based on the simulation, we set $\epsilon = 2.1$, $\theta = 25$, and $\gamma = 0.87$ to match short-run and long-run passthroughs of 0.5 and 0.7, together with a steady-state markup of 1.5.

In addition to the main parameters— ϵ , γ , and θ —we must calibrate other parameters related to inference and fairness: K^b and K^f . We assume that in steady state the perceived markup matches the actual markup: $\overline{K^p} = \overline{K} = 1.5$; this constraint determines K^b . We also assume that the steady-state markup is fair: $K^f = 1.5$. This calibration implies that in steady state, people perceive the correct markup, and they deem this perceived markup to be fair. With this calibration, $\overline{F} = 1$ and $\overline{\Phi} = \theta/2$.

Next we calibrate the labor-supply parameter η , which governs the response of employment to shocks. We set $\eta = 1.1$, which corresponds to a Frisch elasticity of labor supply of $1/1.1 = 0.9$. This value is the median microestimate of the Frisch elasticity for aggregate hours, which combines the labor-supply responses at the intensive and extensive margins (Chetty et al. 2013,

Table 2, row 3, column 3).

We then set the quarterly discount factor to $\beta = 0.99$, giving an annual rate of return on bonds of 4%. We also set the production-function parameter to $\alpha = 1$. This calibration guarantees that the labor share takes its conventional value of $2/3$. Indeed, in the model the steady-state labor share is $\alpha/\bar{K} = 2/3 \times \alpha$.

Finally, we calibrate the monetary-policy rule by setting the response of the nominal interest rate to inflation to $\mu = 1.5$, which is consistent with observed variations in the federal funds rate since the 1980s (Gali 2008, p. 52).

In addition to our New Keynesian model, we also calibrate a simple standard New Keynesian model (described in Appendix D) for benchmark purposes. For the parameters common to the two models, we use the same parameter values—except for ϵ . In the standard New Keynesian model, the steady-state markup is $\epsilon/(\epsilon - 1)$, so we set $\epsilon = 3$ to match a markup of 1.5.

The standard New Keynesian model uses Calvo pricing: nonneutrality arises because a fraction of firms, κ , cannot update their prices each period. We calibrate κ , which is the key parameter behind the nonneutrality of monetary policy, as in the New Keynesian literature: using microevidence on the frequency of price adjustments. If a share κ of firms keep their price fixed each period, the average duration of a price spell is $1/(1 - \kappa)$ (Gali 2008, p. 43). Nakamura and Steinsson (2013) review the literature measuring price rigidity in the United States: in the microdata underlying the Consumer Price Index, the mean duration of price spells is about 3 quarters (Table 1, column 4). Hence, we set $1/(1 - \kappa) = 3$, which implies $\kappa = 2/3$.

4.6. The Effects of Monetary Policy

In the long run, monetary policy determines steady-state inflation, and the effects of monetary policy are described by a long-run Phillips curve linking steady-state inflation to steady-state employment. In the short run, monetary policy determines the nominal interest rate, and the effects of monetary policy are described by impulse responses to an unexpected increase in interest rate.

Long Run. Figure 2 displays two versions of the long-run Phillips curve: one describes the relationship between steady-state inflation and steady-state goods-market markup, and the other one the relationship between steady-state inflation and steady-state employment. The curve is computed from Proposition 2 using the parameter values in Table 3. Furthermore, we set the belief function such that at an inflation of 2%—our reference since it is the implicit inflation target in the United States—steady-state markup is 1.5. Then employment is measured as the percentage deviation from the employment level when inflation is 2%.

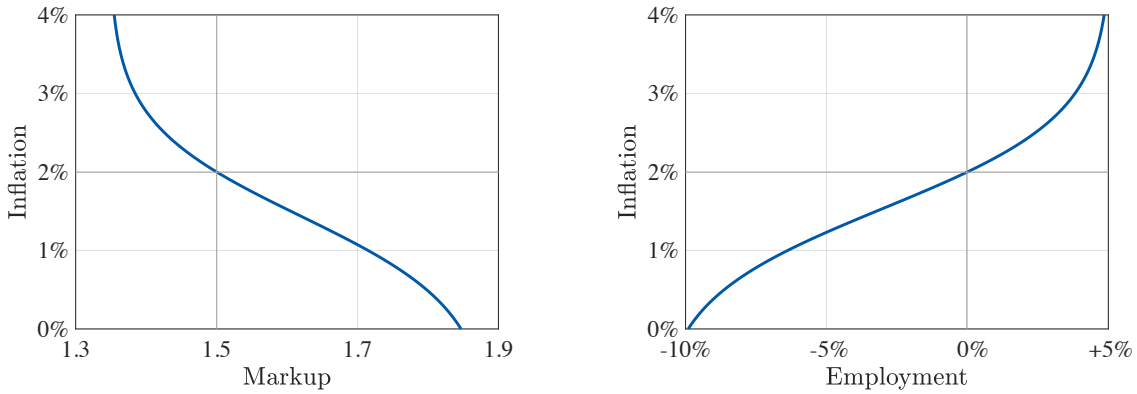


Figure 2. Long-Run Phillips Curve

Notes: The left-hand panel is one version of the long-run Phillips curve: it gives the relationship between steady-state inflation and steady-state goods-market markup. It is constructed using (37) under the calibration in Table 3. The right-hand panel is another version of the long-run Phillips curve: it gives the relationship between steady-state inflation and steady-state employment. It is constructed using (38) under the calibration in Table 3; on the x-axis, employment is measured as the percent deviation from the employment level when steady-state inflation is 2%.

The left-hand panel shows that higher inflation leads to lower markup. As a consequence, as showed in the right-hand panel, higher inflation leads to higher employment. Quantitatively, if inflation raises from 2% to 4%, the markup falls from 1.5 to 1.35, and employment increases by 5%. On the other hand, if inflation falls from 2% to 1%, the markup rises from 1.5 to 1.7, and employment falls by 6%. This effect of steady-state inflation on the goods-market markup could explain part of the variation in markup measured by De Loecker and Eeckhout (2017) in the United States between 1980 and 2014. They find that the average markup increased from 1.2 to 1.7 over that period. At the same time, average inflation fell from above 5% to around 1.5%. Through our mechanism, this drop in inflation could explain part of the increase in markup.

Short Run. Next, we study the response to an unexpected transitory shock to monetary policy, focusing on the equilibrium around the steady state with zero inflation.¹⁷ To obtain a steady state with zero inflation, we set the exogenous component of the monetary-policy rule appropriately: $\bar{i}_0 = \rho$. We assume that the exogenous component of the monetary-policy rule follows an AR(1) process:

$$\widehat{i}_0(t) = \zeta \cdot \widehat{i}_0(t-1) + \epsilon^i(t)$$

¹⁷In our model it would be simple to study the equilibrium around a steady state with positive or negative inflation. But in the standard New Keynesian model, it is more complicated to study equilibria around steady states with nonzero inflation (see Coibion and Gorodnichenko 2011). To simplify the simulations of the standard model, following the literature, we therefore assume that steady-state inflation is zero.

where $\zeta \in (0, 1)$ and $\{\epsilon^i(t)\}$ is a white noise process with mean zero. A positive realization of ϵ^i corresponds to a contractionary monetary-policy shock, leading to a rise in the nominal interest rate, given inflation. In the simulations we set $\zeta = 3/4$, which corresponds to a moderately persistent policy shock (Gali 2008, p. 52; Gali 2011, p. 26).

Before launching the simulation, we verify that under our calibration, the model admits a unique, determinate rational-expectations equilibrium. Since two variables are nonpredetermined at time t ($\widehat{n}(t)$ and $\widehat{\pi}(t)$) and one is predetermined ($\widehat{k}^p(t-1)$), the solution to the dynamical system (42) is unique if exactly one eigenvalue of the matrix \mathbf{A} is within the unit circle, and exactly two are outside the unit circle (see Woodford 2003, p. 672). Furthermore, in that case, the equilibrium is a saddle path. Under our calibration, the eigenvalues of \mathbf{A} are $1.02 + 0.02i$, $1.02 - 0.02i$, and 0.4 : since exactly one eigenvalue is within the unit circle, the rational-expectations equilibrium exists and is unique.

Figure 3 displays the dynamic response to a contractionary monetary-policy shock comprising an increase of 25 basis point of ϵ^i at time 0. Without any response of inflation, this shock would lead to an increase of the annualized nominal interest rate by one percentage point. In the figure, the responses of the real interest rate and inflation are expressed in annual terms (by multiplying by 4 the responses of the variables $\widehat{\pi}(t)$ and $\widehat{r}(t)$). The responses of the other variables are expressed as percentage deviations from their steady-state values.

The tightening of monetary policy generates a decrease in the inflation rate, and an increase in the real interest rate. Inflation is negative for about two quarters and close to zero after that. The deflation leads to a decrease in perceived goods-market markups, as customers underinfer the decrease in marginal costs from lower prices. Firms take advantage of lower perceived markups by raising actual markups. The actual goods-market markup rises by more than 1% above its steady-state value, and thus output and employment fall, by about 0.7% below their steady-state value (the responses of output and employment are the same since we calibrate the production function to be linear). Since labor demand falls, the real wage falls. Since goods-market markups rise, and despite the reduction in output, real profits increase.

Monetary-policy shocks influence output, employment and other real variables, meaning that monetary policy is nonneutral. A large amount of evidence documents the nonneutrality of money: an important early contribution is Friedman and Schwartz (1963), much of the evidence is summarized in Christiano, Eichenbaum, and Evans (1999), and recent work include Romer and Romer (2004) and Christiano, Eichenbaum, and Evans (2005). Of course, many models of monetary nonneutrality have been developed (see Blanchard 1990; Mankiw and Reis 2010), but, with the exception of Rotemberg (2005), none invoking the fairness of prices.

In our New Keynesian model, as in the standard model, the response of real variables

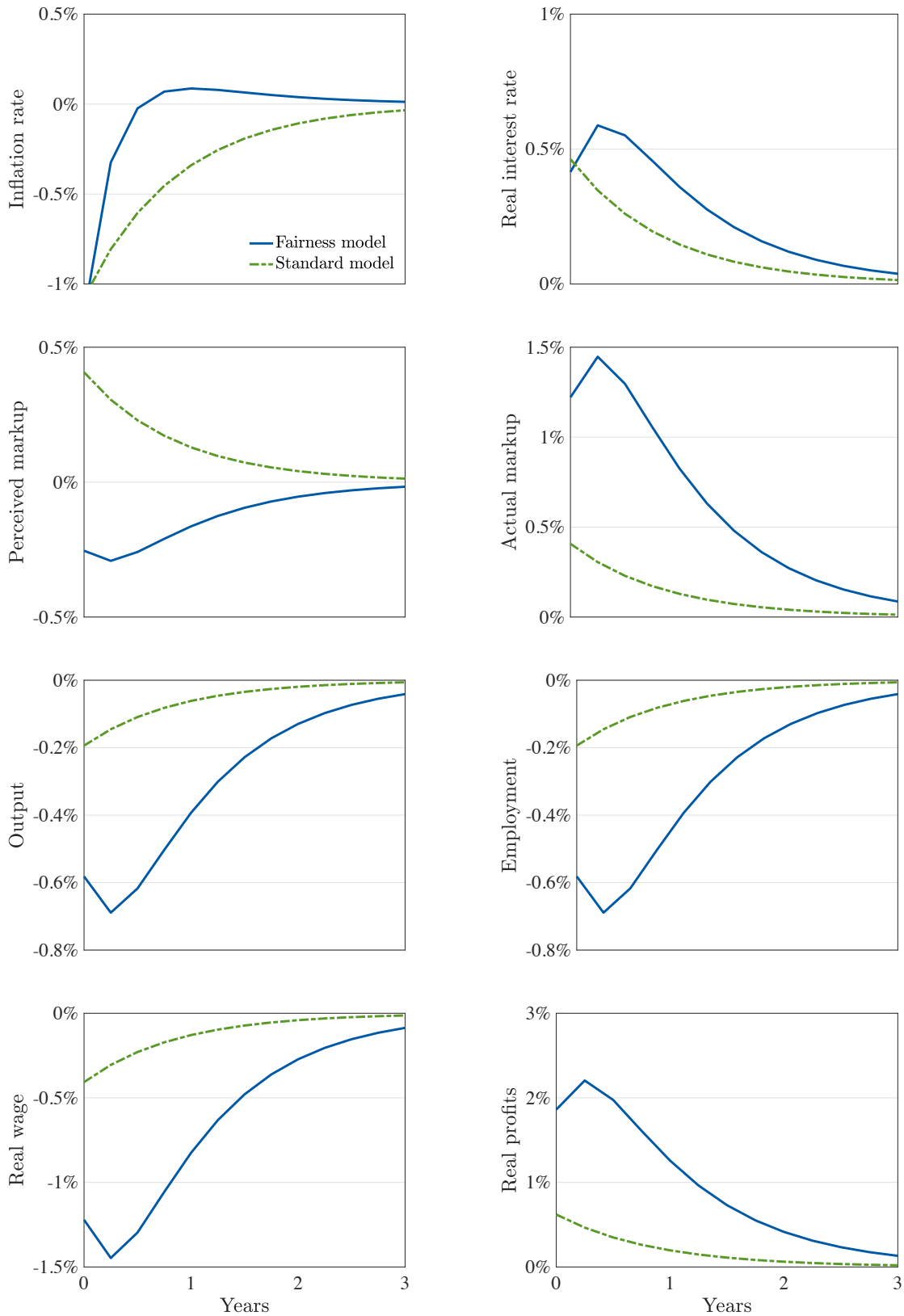


Figure 3. Response to Tighter Monetary Policy

Notes: This figure describes the response of the fairness and standard New Keynesian models to an increase in the exogenous component of the monetary-policy rule, i_0 , by 1 percentage point (annualized). The fairness model is described in Section 4.4. The standard model is described in Appendix D. The two models are calibrated in Table 3. The real interest rate and inflation rate are deviations from steady state (measured in percentage points) and annualized. The other variables are percentage deviations from steady state.

to monetary-policy shocks is driven by the response of the goods-market markup. Here, the markup rises after an increase in nominal interest rate, which then drives the response of output, employment, real wage, and real profits. If business cycles are mostly generated by aggregate-demand shocks, then our model would predict that goods-market markups are countercyclical. There is indeed some evidence of countercyclical markups (for example, Rotemberg and Woodford 1999). But measuring aggregate markups is challenging, so the empirical evidence is not definitive.

There are two qualitative differences between the impulse responses of the New Keynesian model with fairness to those of the standard New Keynesian model. The first concerns the response of the perceived markup. In the fairness model, households believe that markups on the goods market are lower and transactions are more fair when they observe the lower inflation generated by an increase in interest rate. By contrast, in the standard model, households correctly infer markups from available information, so they understand that markups on the goods market are higher. Hence, unlike the standard model, our model explains Shiller (1997)'s finding that people are angry after an increase in inflation triggered by expansionary monetary policy. This happens because people perceive transactions as less fair when inflation rises, which reduces their consumption utility, and triggers a feeling of displeasure (the exact definition of "angry"). Despite people's perceptions, transactions have however become more fair.

The second difference is that the responses of output and employment are hump-shaped in the New Keynesian model with fairness, but not in the standard New Keynesian model. This property of the fairness model coincides with empirical evidence: output is estimated to respond to a monetary-policy shock in a hump-shaped fashion, peaking after several quarters (Romer and Romer 2004; Christiano, Eichenbaum, and Evans 2005). Of course the standard New Keynesian model can be extended in various ways to obtain hump-shaped responses (for example, Christiano, Eichenbaum, and Evans 2005).

Quantitatively, there also are differences in the impulse responses of the two models. The response of inflation is much more transient in the fairness model: inflation reaches zero in about a year in the fairness model and in about three years in the standard model. Moreover, the response of the markup, employment, and output is about three times as large in the fairness model. So monetary policy shocks are more amplified in the fairness model calibrated to match microevidence on passthrough dynamics than in the standard model calibrated to match microevidence on price dynamics.

Finally, there is one important discrepancy between the impulse responses in our model and those estimated in US data. These discrepancy also applies to the simple standard New Keynesian model simulated here. Monetary-policy shocks are estimated to have a delayed and

gradual effect on inflation (Romer and Romer 2004; Christiano, Eichenbaum, and Evans 2005). In our model, the response of inflation is immediate and transient. It is not clear yet how this issue can be addressed.

5. Conclusion

In many contexts, prices are somewhat rigid—they only partially respond to changes in marginal costs. To explain price rigidity, this paper presents a model of monopolistic pricing in which customers care about the fairness of prices: customers derive more utility from a good priced at a low markup than at a high markup. This assumption is motivated by copious evidence that firms stabilize prices out of fairness for their customers, and that customers are very concerned about fairness and consider a fair price to be a fair markup over marginal cost.

We find that preferences for fairness alone cannot explain price rigidity. When marginal costs are hidden but customers infer them rationally from prices, prices are flexible: the passthrough of marginal costs into prices is one. Yet extensive laboratory evidence indicates that people tend not to draw sufficient inference from their observations. When people underinfer hidden marginal costs from prices, our model generates some price rigidity: the passthrough of marginal costs into prices is strictly less than one. The logic for the result is simple. When prices rise following an increase in marginal costs, customers underappreciate the increase in marginal costs and partially misattribute higher prices to higher markups. As they perceive transactions as less fair, the price elasticity of their demand for goods rises, and firms respond by reducing markups. Hence, the passthrough of marginal costs into prices is less than one.

Our model of price rigidity could be useful in the study of optimal monetary policy. Most New Keynesian models used to study optimal monetary policy rely on the assumption of infrequent pricing from Calvo (1983). The Calvo model of pricing does not provide a theory of price rigidity: it is only a modeling device to generate price rigidity. Nevertheless, Calvo pricing has been immensely popular because it offers a tractable way to introduce price rigidity in a macroeconomic model. There exist models of pricing that are more realistic than the Calvo model; but they have not been nearly as successful because they are much less tractable. By contrast, the complexity of our model is comparable to that of the Calvo model, and our microfoundations accord well with evidence collected by surveys of firms and customers. Building on reasonable microfoundations is especially important to study optimal monetary policy because the choice of microfoundations determines the effects of monetary policy on social welfare; these effects in turn determine the outcome of the policy analysis.

References

- Akerlof, George A. 1982. "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics* 97 (4): 543–569.
- Akerlof, George A., William T. Dickens, and George L. Perry. 1996. "The Macroeconomics of Low Inflation." *Brookings Papers on Economic Activity* 1996 (1): 1–76.
- Akerlof, George A., and Janet L. Yellen. 1990. "The Fair Wage-Effort Hypothesis and Unemployment." *Quarterly Journal of Economics* 105 (2): 255–283.
- Alvarez, Luis J., and Ignacio Hernando. 2005. "Price Setting Behaviour of Spanish Firms: Evidence from Survey Data." ECB Working Paper 538.
- Amirault, David, Carolyn Kwan, and Gordon Wilkinson. 2006. "Survey of Price-Setting Behaviour of Canadian Companies." Bank of Canada Working Paper 06-35.
- Amiti, Mary, Oleg Itskhoki, and Jozef Konings. 2014. "Importers, Exporters, and Exchange Rate Disconnect." *American Economic Review* 104 (7): 1942–1978.
- Apel, Mikael, Richard Friberg, and Kerstin Hallsten. 2005. "Microfoundations of Macroeconomic Price Adjustment: Survey Evidence from Swedish Firms." *Journal of Money, Credit and Banking* 37 (2): 313–338.
- Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review* 98 (5): 1998–2031.
- Aucremanne, Luc, and Martine Druant. 2005. "Price-Setting Behaviour in Belgium: What Can Be Learned from an Ad Hoc Survey?" ECB Working Paper 448.
- Ball, Laurence, and David Romer. 1990. "Real Rigidities and the Non-Neutrality of Money." *Review of Economic Studies* 57 (2): 183–203.
- Banerjee, Anindya, and Bill Russell. 2005. "Inflation and Measures of the Markup." *Journal of Macroeconomics* 27 (2): 289–306.
- Barsky, Robert B., Mark Bergen, Shantanu Dutta, and Daniel Levy. 2003. "What Can the Price Gap Between Branded and Private-Label Products Tell Us About Markups?" In *Scanner Sata and Price Indexes*, edited by Robert C. Feenstra and Matthew D. Shapiro, chap. 7: University of Chicago Press.
- Benabou, Roland. 1992. "Inflation and Markups: Theories and Evidence from the Retail Trade Sector." *European Economic Review* 36: 566–574.
- Benigno, Pierpaolo, and Luca Antonio Ricci. 2011. "The Inflation-Output Trade-Off with Downward Wage Rigidities." *American Economic Review* 101 (4): 1436–1466.
- Benjamin, Daniel J. 2015. "A Theory of Fairness in Labor Markets." *Japanese Economic Review* 66 (2): 182–225.
- Benzarti, Youssef, and Dorian Carloni. 2016. "Who Really Benefits from Consumption Tax Cuts? Evidence from a Large VAT Reform in France." <http://dx.doi.org/10.2139/ssrn.2629380>.
- Benzarti, Youssef, Dorian Carloni, Jarkko Harju, and Tuomas Kosonen. 2017. "What Goes Up May Not Come Down: Asymmetric Incidence of Value-Added Taxes." http://conference.nber.org/confer/2017/SI2017/PE/Benzarti_Carloni_Harju_Kosonen.pdf.
- Bergin, Paul R., and Robert C. Feenstra. 2001. "Pricing-To-Market, Staggered Contracts, and Real Exchange Rate Persistence." *Journal of International Economics* 54 (2): 333–359.
- Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium."

- Econometrica* 63 (4): 841–890.
- Bieger, Thomas, Isabelle Engeler, and Christian Laesser. 2010. “In What Condition is a Price Increase Perceived as Fair? An Empirical Investigation in the Cable Car Industry.” In *20th Annual CAUTHE Conference*, 1–12. Hobart, Tasmania: School of Management, University of Tasmania.
- Blanchard, Olivier J. 1990. “Why Does Money Affect Output? A Survey.” In *Handbook of Monetary Economics*, edited by Benjamin M. Friedman and Frank H. Hahn, vol. 2, chap. 15. Amsterdam: Elsevier.
- Blinder, Alan S., Elie R. D. Canetti, David E. Lebow, and Jeremy B. Rudd. 1998. *Asking about Prices*. New York: Russell Sage Foundation.
- Burstein, Ariel, and Gita Gopinath. 2014. “International Prices and Exchange Rates.” In *Handbook of International Economics*, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, vol. 4, chap. 7. Amsterdam: Elsevier.
- Bushong, Benjamin, Matthew Rabin, and Joshua Schwartzstein. 2015. “A Model of Relative Thinking.” <http://www.dartmouth.edu/~jschwartzstein/papers/NormMay15.pdf>.
- Calvo, Guillermo A. 1983. “Staggered Prices in a Utility-Maximizing Framework.” *Journal of Monetary Economics* 12 (3): 383–398.
- Carillo, Juan, and Thomas Palfrey. 2011. “No Trade.” *Games and Economic Behavior* 71 (1): 66–87.
- Carlsson, Mikael, and Oskar Nordstrom Skans. 2012. “Evaluating Microfoundations for Aggregate Price Rigidities: Evidence from Matched Firm-Level Data on Product Prices and Unit Labor Cost.” *American Economic Review* 102 (4): 1571–1595.
- Carlton, Dennis W. 1986. “The Rigidity of Prices.” *American Economic Review* 76 (4): 637–658.
- Charness, Gary, and Matthew Rabin. 2002. “Understanding Social Preferences with Simple Tests.” *Quarterly Journal of Economics* 117 (3): 817–869.
- Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2013. “Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities.” In *NBER Macroeconomics Annual 2012*, edited by Daron Acemoglu, Jonathan Parker, and Michael Woodford, chap. 1. Chicago: University of Chicago Press.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 1999. “Monetary Policy Shocks: What Have We Learned And To What End?” In *Handbook of Macroeconomics*, edited by John B. Taylor and Michael Woodford, vol. 1, chap. 2. Amsterdam: Elsevier.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 2005. “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy.” *Journal of Political Economy* 113 (1): 1–45.
- Coibion, Olivier, and Yuriy Gorodnichenko. 2011. “Monetary Policy, Trend Inflation, and the Great Moderation: An Alternative Interpretation.” *American Economic Review* 101 (1): 341–370.
- De Loecker, Jan, and Jan Eeckhout. 2017. “The Rise of Market Power and the Macroeconomic Implications.” NBER Working Paper 23687.
- De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. 2016. “Prices, Markups and Trade Reform.” *Econometrica* 84 (2): 445–510.
- Dornbusch, Rudiger. 1985. “Exchange Rates and Prices.” *American Economic Review* 77 (1): 93–107.
- Dotsey, Michael, and Robert G. King. 2005. “Implications of Date-Dependent Pricing for Dynamic Macroeconomic Models.” *Journal of Monetary Economics* 52 (1): 213–242.
- Dufwenberg, Martin, Paul Heidhues, Georg Kirchsteiger, Frank Riedel, and Joel Sobel. 2011. “Other-

- Regarding Preferences in General Equilibrium.” *Review of Economic Studies* 78 (2): 613–639.
- Eichenbaum, Martin, and Jonas D.M. Fisher. 2007. “Estimating the Frequency of Price Re-Optimization in Calvo-Style Models.” *Journal of Monetary Economics* 54 (7): 2032–2047.
- Esponda, Ignacio, and Emanuel Vespa. 2014. “Hypothetical Thinking and Information Extraction in the Laboratory.” *American Economic Journal: Microeconomics* 6 (4): 180–202.
- Fehr, Ernst, and S. Gächter. 2000. “Fairness and Retaliation: The Economics of Reciprocity.” *Journal of Economic Perspectives* 14 (3): 159–181.
- Fehr, Ernst, Lorenz Goette, and Christian Zehnder. 2009. “A Behavioral Account of the Labor Market: The Role of Fairness Concerns.” *Annual Review of Economics* 1 (1): 355–384.
- Fehr, Ernst, Alexander Klein, and Klaus M. Schmidt. 2007. “Fairness and Contract Design.” *Econometrica* 75 (1): 121–154.
- Fehr, Ernst, and Klaus M. Schmidt. 1999. “A Theory of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics* 114 (3): 817–868.
- Fraichard, Julien. 2006. “Le Commerce en 2005.” http://www.insee.fr/fr/ffc/docs_ffc/ref/comfra06bb.pdf.
- Frey, Bruno S., and Werner W. Pommerehne. 1993. “On the Fairness of Pricing—An Empirical Survey Among the General Population.” *Journal of Economic Behavior & Organization* 20 (3): 295–307.
- Friedman, Milton, and Anna J. Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton, NJ: Princeton University Press.
- Gabaix, Xavier. 2014. “A Sparsity-Based Model of Bounded Rationality.” *Quarterly Journal of Economics* 129 (4): 1661–1710.
- Gabaix, Xavier. 2016. “A Behavioral New Keynesian Model.” NBER Working Paper 22954.
- Gali, Jordi. 2008. *Monetary Policy, Inflation, and the Business Cycle*. Princeton, NJ: Princeton University Press.
- Gali, Jordi. 2011. *Unemployment Fluctuations and Stabilization Policies: A New Keynesian Perspective*. Cambridge, MA: MIT Press.
- Gennaioli, Nicola, and Andrei Shleifer. 2010. “What Comes to Mind.” *Quarterly Journal of Economics* 125 (4): 1399–1433.
- Gielissen, Robert, Chris E. Dutilh, and Johan J. Graafland. 2008. “Perceptions of Price Fairness: An Empirical Research.” *Business & Society* 47 (3): 370–389.
- Gopinath, Gita, and Roberto Rigobon. 2008. “Sticky Borders.” *Quarterly Journal of Economics* 123 (2): 531–575.
- Hall, R. L., and C. J. Hitch. 1939. “Price Theory and Business Behaviour.” *Oxford Economic Papers* 2: 12–45.
- Hall, Simon, Mark Walsh, and Anthony Yates. 2000. “Are UK Companies’ Prices Sticky?” *Oxford Economic Papers* 52 (3): 425–446.
- Hoerberichts, Marco, and Ad Stokman. 2006. “Pricing Behaviour of Dutch Companies: Main Results from a Survey.” ECB Working Paper 607.
- Holt, Charles A., and Roger Sherman. 1994. “The Loser’s Curse.” *American Economic Review* 84 (3): 642–652.
- Jones, Douglas N., and Patrick C. Mann. 2001. “The Fairness Criterion in Public Utility Regulation: Does Fairness Still Matter?” *Journal of Economic Issues* 35 (1): 153–172.
- Kackmeister, Alan. 2007. “Yesterday’s Bad Times Are Today’s Good Old Times: Retail Price Changes

- Are More Frequent Today Than in the 1890s.” *Journal of Money, Credit and Banking* 39 (8): 1987–2020.
- Kagel, John, and Dan Levin. 2002. *Common Value Auctions and the Winner’s Curse*. Princeton, NJ: Princeton University Press.
- Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler. 1986. “Fairness as a Constraint on Profit Seeking: Entitlements in the Market.” *American Economic Review* 76 (4): 728–41.
- Kalapurakal, Rosemary, Peter R. Dickson, and Joel E. Urbany. 1991. “Perceived Price Fairness and Dual Entitlement.” In *Advances in Consumer Research*, edited by Rebecca H. Holman and Michael R. Solomon, vol. 18, 788–793. Provo, UT: Association for Consumer Research.
- Kaplan, Steven Laurence. 1996. *The Bakers of Paris and the Bread Question: 1700–1755*. Durham, NC: Duke University Press.
- Kimball, Miles S. 1995. “The Quantitative Analytics of the Basic Neomonetarist Model.” *Journal of Money, Credit and Banking* 27 (4): 1241–1277.
- King, Robert G., and Mark W. Watson. 1994. “The Post-War US Phillips Curve: A Revisionist Econometric History.” *Carnegie-Rochester Conference Series on Public Policy* 41: 157–219.
- King, Robert G., and Mark W. Watson. 1997. “Testing Long-Run Neutrality.” *Economic Quarterly* 83 (3): 69–101.
- Klenow, Peter J., and Jonathan L. Willis. 2016. “Real Rigidities and Nominal Price Changes.” *Economica* 83 (331): 443–472.
- Konow, James. 2001. “Fair and Square: The Four Sides of Distributive Justice.” *Journal of Economic Behavior & Organization* 46 (2): 137–164.
- 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.
- Kwapil, Claudia, Joseph Baumgartner, and Johann Scharler. 2005. “The Price-Setting Behaviour of Austrian Firms: Some Survey Evidence.” ECB Working Paper 464.
- Loupias, Claire, and Roland Ricart. 2004. “Price Setting in France: New Evidence from Survey Data.” ECB Working Paper 448.
- Lunnemann, Patrick, and Thomas Y. Matha. 2006. “New Survey Evidence on the Pricing Behavior of Luxembourg Firms.” ECB Working Paper 617.
- Mankiw, N. Gregory, and Ricardo Reis. 2010. “Imperfect Information and Aggregate Supply.” In *Handbook of Monetary Economics*, edited by Benjamin M. Friedman and Michael Woodford, vol. 3A, chap. 5. Amsterdam: Elsevier.
- Martins, Fernando. 2005. “The Price Setting Behaviour of Portuguese Firms: Evidence From Survey Data.” ECB Working Paper 562.
- Mavroeidis, Sophocles, Mikkel Plagborg-Moller, and James H. Stock. 2014. “Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve.” *Journal of Economic Literature* 52 (1): 124–188.
- Maxwell, Sarah. 1995. “What Makes a Price Increase Seem “Fair”?” *Pricing Strategy & Practice* 3 (4): 21–27.
- Melitz, Marc J., and Gianmarco I. P. Ottaviano. 2008. “Market Size, Trade, and Productivity.” *Review of Economic Studies* 75 (1): 295–316.
- Miller, Judith A. 1999. *Mastering the Market. The State and the Grain Trade in Northern France,*

- 1700–1860. Cambridge: Cambridge University Press.
- Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer. 2008. “Coarse Thinking and Persuasion.” *Quarterly Journal of Economics* 123 (2): 577–619.
- Nakamura, Emi, and Jon Steinsson. 2013. “Price Rigidity: Microeconomic Evidence and Macroeconomic Implications.” *Annual Review of Economics* 5: 133–163.
- Nakamura, Emi, and Dawit Zerom. 2010. “Accounting for Incomplete Pass-Through.” *Review of Economic Studies* 77 (3): 1192–1230.
- Nevo, Aviv. 2001. “Measuring Market Power in the Ready-To-Eat Cereal Industry.” *Econometrica* 69 (2): 307–342.
- Okun, Arthur M. 1975. “Inflation: Its Mechanics and Welfare Costs.” *Brookings Papers on Economic Activity* 6 (2): 351–402.
- Okun, Arthur M. 1981. *Prices and Quantities: A Macroeconomic Analysis*. Washington, DC: Brookings Institution.
- Piron, Robert, and Luis Fernandez. 1995. “Are Fairness Constraints on Profit-Seeking Important?” *Journal of Economic Psychology* 16 (1): 73–96.
- Rabin, Matthew. 1993. “Incorporating Fairness into Game Theory and Economics.” *American Economic Review* 83 (5): 1281–1302.
- Romer, Christina D., and David Romer. 2004. “A New Measure of Monetary Policy Shocks: Derivation and Implications. .” *American Economic Review* 94 (4): 1055–1084.
- Rotemberg, Julio J. 2005. “Customer Anger at Price Increases, Changes in the Frequency of Price Adjustment and Monetary Policy.” *Journal of Monetary Economics* 52 (4): 829–852.
- Rotemberg, Julio J. 2011. “Fair Pricing.” *Journal of the European Economic Association* 9 (5): 952–981.
- Rotemberg, Julio J., and Michael Woodford. 1995. “Dynamic General Equilibrium Models with Imperfectly Competitive Product Markets.” In *Frontiers of Business Cycle Research*, edited by Thomas F. Cooley, chap. 9. Princeton, NJ: Princeton University Press.
- Rotemberg, Julio J., and Michael Woodford. 1999. “The Cyclical Behavior of Prices and Costs.” In *Handbook of Macroeconomics*, edited by John B. Taylor and Michael Woodford, vol. 1, chap. 16. Amsterdam: Elsevier.
- Samuelson, William F., and Max H. Bazerman. 1985. “The Winner’s Curse in Bilateral Negotiations.” In *Research in Experimental Economics*, edited by Vernon L. Smith, vol. 3, 105–137. Greenwich, CT: JAI Press.
- Schmidt, Klaus M. 2011. “Social Preferences and Competition.” *Journal of Money, Credit and Banking* 43 (5): 207–231.
- Shafir, Eldar, Peter Diamond, and Amos Tversky. 1997. “Money Illusion.” *Quarterly Journal of Economics* 112 (2): 341–374.
- Shiller, Robert J. 1997. “Why Do People Dislike Inflation?” In *Reducing Inflation: Motivation and Strategy*, edited by Christina D. Romer and David H. Romer, chap. 1. Chicago: University of Chicago Press.
- Shiller, Robert J., Maxim Boycko, and Vladimir Korobov. 1991. “Popular Attitudes Towards Free Markets: The Soviet Union and the United States Compared.” *American Economic Review* 81 (3): 385–400.
- Tobin, James. 1972. “Inflation and Unemployment.” *American Economic Review* 62 (1/2): 1–18.
- Tversky, Amos, and Daniel Kahneman. 1973. “Availability: A Heuristic for Judging Frequency and

- Probability.” *Cognitive Psychology* 5 (2): 207–232.
- Urbany, Joel E., Thomas J. Madden, and Peter R. Dickson. 1989. “All’s Not Fair in Pricing: An Initial Look at the Dual Entitlement Principle.” *Marketing Letters* 1 (1): 17–25.
- Warhaftig, Itamar. 1987. “Consumer Protection: Price and Wage Levels.” In *Crossroads: Halacha and the Modern World*, vol. 1, 49–69. Alon Shvut-Gush Etzion, Israel: Zomet Institute.
- Wezsacker, Georg. 2010. “Do We Follow Others when We Should? A Simple Test of Rational Expectations.” *American Economic Review* 100 (5): 2340–2360.
- Woodford, Michael. 2003. *Interest and Prices*. Princeton, NJ: Princeton University Press.
- Xia, Lan, Kent B. Monroe, and Jennifer L. Cox. 2004. “The Price Is Unfair! A Conceptual Framework of Price Fairness Perceptions.” *Journal of Marketing* 68 (4): 1–15.
- Zajac, Edward E. 1985. “Perceived Economic Justice: The Example of Public Utility Regulation.” In *Cost Allocation: Methods, Principles, Applications*, edited by H. Peyton Young, chap. 7. Amsterdam: North Holland.

Appendix A. Long Proofs

Proof of Proposition 1

We first derive the expression for the profit-maximizing markup K , given by (10). This expression directly follows from the result that $K = E/(E - 1)$ and from the expression (9) for E . Since $K^P(P)$ is strictly increasing in P , $\Phi(K^P)$ is strictly increasing in K^P , and $\gamma > 0$, the right-hand side of (10) is strictly decreasing in K ; it is also strictly positive for $K \geq 0$. Hence, (10) always has a unique solution, so that K is well-defined and unique.

Next we derive the expression for the marginal-cost passthrough σ . Equation (9) gives the price elasticity of demand as a function of the perceived markup: $E(K^P)$. Thus the profit-maximizing markup is a function of the perceived markup: $K(K^P) = E(K^P)/(E(K^P) - 1)$. Finally, (8) gives the perceived markup as a function of the price: $K^P(P)$. The price therefore satisfies $P = K(K^P(P)) \cdot MC$. Taking logs and differentiating, we obtain

$$\sigma = \frac{d \ln(P)}{d \ln(MC)} = 1 + \frac{d \ln(K)}{d \ln(K^P)} \cdot \frac{d \ln(K^P)}{d \ln(P)} \cdot \frac{d \ln(P)}{d \ln(MC)}.$$

Using $d \ln(K^P)/d \ln(P) = \gamma$ and $d \ln(P)/d \ln(MC) = \sigma$, we reshuffle the above equation and obtain

$$(A1) \quad \sigma = 1 \left/ \left(1 - \gamma \frac{d \ln(K)}{d \ln(K^P)} \right) \right.$$

Using (9) and (2), we obtain the following elasticity:

$$\frac{d \ln(E)}{d \ln(K^P)} = \frac{E - \epsilon}{E} \cdot \frac{d \ln(\Phi)}{d \ln(K^P)} = \frac{E - \epsilon}{E} (\theta - \Phi).$$

Since $K(K^P) = E(K^P)/(E(K^P) - 1)$, the elasticities of K and E are related:

$$\frac{d \ln(K)}{d \ln(K^P)} = \left(1 - \frac{E}{E - 1} \right) \frac{d \ln(E)}{d \ln(K^P)} = -\frac{1}{E - 1} \cdot \frac{d \ln(E)}{d \ln(K^P)}.$$

Combining the last two equations yields

$$-\frac{d \ln(K)}{d \ln(K^P)} = \frac{E - \epsilon}{(E - 1)E} (\theta - \Phi) = \frac{\gamma \Phi (\theta - \Phi)}{[1 + \gamma \Phi] [\epsilon + (\epsilon - 1) \gamma \Phi]}.$$

Using this last equation and (A1), we obtain (11).

Proof of Corollary 2

First Case: $\theta = 0$ or $\gamma = 0$. In steady state, the goods-market markup $\bar{k} = \ln(\bar{K})$ directly determines all variables. Indeed, (32) gives

$$(A2) \quad (1 + \eta)\bar{n} = \ln(\alpha) - \ln\left(\frac{\nu}{\nu - 1}\right) - \bar{k}.$$

So \bar{k} determines \bar{n} . Then, the production constraint (19) links output to employment:

$$(A3) \quad \bar{y} = \bar{a} + \alpha\bar{n}$$

So \bar{k} determines \bar{y} . Further, the wage-setting equation (28) links real wage to employment and output:

$$(A4) \quad \bar{w} - \bar{p} = \ln\left(\frac{\nu}{\nu - 1}\right) + \eta\bar{n} + \bar{y}.$$

So \bar{k} determines $\bar{w} - \bar{p}$. Finally, real profits are given by (33):

$$(A5) \quad \bar{v} - \bar{p} = \bar{y} + \ln\left(1 - \frac{\alpha}{\bar{K}}\right).$$

So \bar{k} determines $\bar{v} - \bar{p}$. Proposition 2 shows that with $\theta = 0$ or $\gamma = 0$, then \bar{k} is independent of steady-state inflation. This implies that in this first case, all variables are independent of steady-state inflation: money is superneutral. The expression for the long-run Phillips curve comes from (A2) and the value of \bar{k} given by Proposition 2.

Second Case: $\theta > 0$ and $\gamma > 0$. Since $K^P(\pi)$ is increasing in $\bar{\pi}$ when $\gamma > 0$, and $\Phi(K^P)$ is decreasing in K^P when $\theta > 0$, equation (37) implies that in this second case, steady-state goods-market markup is decreasing in steady-state inflation. Then, (A2) implies that steady-state employment is increasing in steady-state inflation. Next, (A3) implies that steady-state output is increasing in steady-state inflation. It follows from these results and (A4) that steady-state real wage is increasing in steady-state inflation. The next step is to compute the response of steady-state real profits to steady-state inflation. Equation (A3) implies that $d\bar{y}/d\bar{n} = \alpha$ and equation (A2) implies that $d\bar{n}/d\bar{k} = -1/(1 + \eta)$ so $d\bar{y}/d\bar{k} = -\alpha/(1 + \eta)$. Equation (A5) yields

$$\frac{d(\bar{v} - \bar{p})}{d\bar{k}} = \frac{d\bar{y}}{d\bar{k}} + \frac{\alpha}{\bar{K} - \alpha},$$

which implies

$$(A6) \quad \frac{d(\bar{v} - \bar{p})}{d\bar{k}} = \alpha \cdot \left(\frac{1}{\bar{K} - \alpha} - \frac{1}{1 + \eta} \right).$$

Hence, $d(\bar{v} - \bar{p})/d\bar{k} > 0$ if and only if $1/(\bar{K} - \alpha) > 1/(1 + \eta)$, or equivalently, $\bar{K} < 1 + \alpha + \eta$. Higher steady-state inflation leads to lower steady-state goods-market markup, and when $\bar{K} < 1 + \alpha + \eta$, to lower steady-state real profits. Finally, the expression for the long-run Phillips curve comes from (A2) and the value of \bar{k} given by (37).

Appendix B. Signaling

To study how firms strategically reveal cost information to customers, we add to our monopoly model the option for the firm to credibly reveal its cost to customers before choosing its price. The model predicts that firms would want to signal marginal costs when they are subject to a cost increase that is sufficiently large; they would never want to signal marginal costs after a cost decrease. We provide empirical evidence that firms seem to behave this way.

Observable Marginal Costs

Before analyzing the more realistic and interesting case in which marginal costs are unobservable and firms can reveal or conceal them, we begin by briefly studying pricing when marginal costs are observable. This case will be a useful basis for comparison.

When marginal costs are observable, the perceived marginal cost is the true marginal cost: $MC^p(P) = MC$. Hence, the perceived markup equals the true markup: $K^p(P) = P/MC = K$. Equation (5) therefore implies that the price elasticity of demand is $E = \epsilon + (\epsilon - 1)\Phi(K) > \epsilon$. When marginal costs are observable, the concern for fairness increases the price elasticity of the demand curve faced by the monopoly; thus, the profit-maximizing markup is lower than without fairness concerns. However, since the markup does not depend on marginal cost, changes in marginal cost are fully passed through into price, so the passthrough remains one. The following lemma summarizes these results:

LEMMA A1. *When customers care about fairness ($\theta > 0$) but observe marginal costs, the profit-maximizing markup K is defined by*

$$(A7) \quad K = 1 + \frac{1}{\epsilon - 1} \cdot \frac{1}{1 + \Phi(K)},$$

implying that $K < \epsilon/(\epsilon - 1)$, and the marginal-cost passthrough is $\sigma = 1$. The markup is lower than in the absence of fairness concerns, but the passthrough identical.

In the absence of fairness concerns, the price affects demand solely by through the substitution effect. With fairness concerns and observable marginal costs, the price also influences the perceived fairness of the transaction: when the purchase price is high relative to the marginal cost of production, customers deem the transaction to be less fair, which reduces customers' marginal utility from consuming the good. Hence, with fairness concerns and observable marginal costs, a high price also reduces demand by lowering the marginal utility of consumption. As a result, the monopoly's demand is more price elastic than without fairness concerns, and the profit-maximizing markup is accordingly lower.

The lemma predicts that when customers care about fairness but observe costs, the passthrough of marginal costs into prices is one, whereas the passthrough is strictly below one when costs are not observed. Renner and Tyran (2004) provide evidence from a laboratory experiment that in customer markets, price rigidity after a temporary cost shock is much more pronounced when costs are observable than when they are not. Kachelmeier, Limberg, and Schadewald (1991a,b) also find that in laboratory experiments that disclosing information on changes in marginal costs hastens price convergence relative to the convergence observed in markets with no disclosures.

Unobserved Marginal Costs and Signaling

We now turn to the case with unobservable marginal costs and the possibility to signal them. When the firm reveals its marginal cost, the firm's profits are $V^r = (P^r - MC)Y^r$, where the r superscript denotes the firm's decision to "reveal." Furthermore, when the firm reveals its marginal cost, $K^p = K^r$. Using $P^r = K^r \cdot MC$ and substituting the expression for demand (3) yields

$$(A8) \quad V^r = (K^r \cdot MC - MC) \cdot F(K^r)^{\epsilon-1} \cdot (K^r \cdot MC)^{-\epsilon} = MC^{1-\epsilon} \cdot (K^r - 1) \cdot (K^r)^{-\epsilon} \cdot F(K^r)^{\epsilon-1}.$$

When the firm credibly reveal its marginal cost, the profit-maximizing markup is the same as when the marginal cost is observable, so K^r is given by (A7).

When instead the firm conceals its marginal cost before choosing its price, we assume that the customer still uses the inference rule described by (7). When the firm conceals, a rational customer can invert the price to uncover the marginal cost, just as it does absent the possibility of cost disclosure. And when the firm conceals, a fully naive customer infers nothing from the firm's disclosure or pricing decision. Accordingly, the inference rule remains the mixture of

these two extreme cases. The firm's profits from concealing then are

$$(A9) \quad V^c = (K^c \cdot MC - MC) \cdot F(K^P)^{\epsilon-1} \cdot (K^c \cdot MC)^{-\epsilon} = MC^{1-\epsilon} \cdot (K^c - 1) \cdot (K^c)^{-\epsilon} \cdot F(K^P)^{\epsilon-1},$$

where the superscript c denotes “conceal”. The perceived markup is

$$K^P(K^c, MC) = (K^b)^{1-\gamma} \left(\frac{K^c \cdot MC}{MC^b} \right)^\gamma,$$

as in (8), and the profit-maximizing markup is given by (10), as when the marginal cost is not observable.

While the profits under revelation do not depend on customers' beliefs, the perceived markup, K^P , is an important determinant of the profits when the monopoly conceals. A higher K^P means a lower fairness measure $F(K^P)$ and a lower markup K^c , and therefore lower profits. Hence the monopoly may choose to conceal or reveal, depending on customers' beliefs:

LEMMA A2. *Assume that customers care about fairness ($\theta > 0$) and underinfer marginal costs ($\gamma > 0$). Consider the decision of the monopoly to conceal or reveal marginal cost depending on customers' beliefs, parameterized by K^b and MC^b . The monopoly's decision solely depends on $\lambda \equiv (K^b)^{1-\gamma} / (MC^b)^\gamma$. There exists a threshold λ_0 such the firm optimally conceals if $\lambda < \lambda_0$ and reveals if $\lambda > \lambda_0$. Equivalently, there exists a threshold K_0^P on the perceived markup such the firm optimally conceals if $K^P < K_0^P$ and reveals if $K^P > K_0^P$.*

Proof. The profits V^r if the monopoly reveals its marginal cost are independent of the belief parameters K^b and MC^b since customers do not need to make any inference when the firm reveals its cost.

On the other hand, the belief parameters K^b and MC^b determine the profits V^c if the monopoly conceals its marginal cost. In fact, in any equilibrium in which the firm conceals, the profits can be written as a function of the perceived markup K^P , which acts as a sufficient statistic summarizing the effect of K^b and MC^b on profits. Using (A9), we write equilibrium profits as a function of K^P :

$$V^c(K^P) = MC^{1-\epsilon} \cdot [K^c(K^P) - 1] \cdot K^c(K^P)^{-\epsilon} \cdot F(K^P)^{\epsilon-1},$$

where

$$K^c(K^P) = 1 + \frac{1}{(\epsilon - 1)(1 + \gamma\Phi(K^P))}.$$

Since $F(K^P)$ is decreasing in K^P and $\epsilon > 1$, then $F(K^P)^{\epsilon-1}$ is decreasing in K^P .

The function $(K - 1)K^{-\epsilon}$ is increasing in K on $[1, \epsilon/(\epsilon - 1)]$ and decreasing on $[\epsilon/(\epsilon - 1), \infty)$. Since $K^c \in (1, \epsilon/(\epsilon - 1))$, then $(K^c - 1)(K^c)^{-\epsilon}$ is increasing in K^c . Furthermore, since Φ is increasing in K^P , then $K^c(K^P)$ is decreasing in K^P . To conclude, $[K^c(K^P) - 1] \cdot K^c(K^P)^{-\epsilon}$ is decreasing in K^P .

Overall, $V^c(K^P)$ is decreasing in K^P . It has the following limits. When $K^P = 0$, $F(0) = 2$ and $\Phi(0) = 0$ so $K^c(0) = \epsilon/(\epsilon - 1)$. Since $F(K^r) < 2$ and $K^r < \epsilon/(\epsilon - 1)$, following the same arguments as above, $V^c(0) > V^r$. When $K^P \rightarrow \infty$, $F(\infty) = 0$ and $\Phi(\infty) = \theta$ so $K^c(\infty) = 1 + 1/[(\epsilon - 1)(1 + \theta)]$. Accordingly, $V^c(\infty) = 0$ so $V^c(\infty) < V^r$. We infer that there is a threshold K_0^P such that for any $K^P < K_0^P$, $V^c(K^P) > V^r$, and for any $K^P > K_0^P$, $V^c(K^P) < V^r$.

We now reformulate this result in terms of the underlying belief parameters. Using (8), we write the perceived markup as $K^P(\lambda) = \lambda \cdot P(\lambda)^\gamma$ where $\lambda \equiv (K^b)^{1-\gamma} / (MC^b)^\gamma$ and $P(\lambda)$ is implicitly defined by

$$P(\lambda) = K^c(\lambda \cdot P(\lambda)^\gamma) \cdot MC.$$

The elasticity of K^P with respect to λ is

$$\frac{d \ln(K^P)}{d \ln(\lambda)} = 1 + \gamma \frac{d \ln(P)}{d \ln(\lambda)}.$$

The elasticity of P with respect to λ satisfies

$$\begin{aligned} \frac{d \ln(P)}{d \ln(\lambda)} &= \frac{d \ln(K^c)}{d \ln(K^P)} \left(1 + \gamma \frac{d \ln(P)}{d \ln(\lambda)} \right) \\ &= \frac{\frac{d \ln(K^c)}{d \ln(K^P)}}{1 - \gamma \frac{d \ln(K^c)}{d \ln(K^P)}}. \end{aligned}$$

Combining these two results, we infer that

$$\frac{d \ln(K^P)}{d \ln(\lambda)} = \frac{1}{1 - \gamma \frac{d \ln(K^c)}{d \ln(K^P)}} = \sigma,$$

where we use (A1) to introduce the passthrough σ . Since $\sigma > 0$, $K^P(\lambda)$ is strictly increasing in λ . Furthermore, since K^c is bounded between 1 and $\epsilon/(\epsilon - 1)$, then $P(\lambda)$ is bounded between MC and $[\epsilon/(\epsilon - 1)] \cdot MC$ for any λ , which implies that $K^P(0) = 0$ and $\lim_{\lambda \rightarrow \infty} K^P(\lambda) = \infty$. Accordingly, the mapping $K^P(\lambda)$ is an increasing bijection from $[0, \infty)$ to $[0, \infty)$. This means that we can reformulate the results above in terms of λ instead of K^P . \square

The intuition for the lemma is simple. The monopoly optimally follows a threshold rule, concealing costs when the perceived markup is low and revealing costs when the perceived

markup would have been high. Because lower perceived markup lead to higher and less elastic demand—which means higher profits—a monopoly facing customers who tend to perceive low markups has much less incentive to reveal its cost and its markup than a monopoly facing customers who tend to perceive high markups.

The previous lemma shows that depending on parameter values, a monopoly may choose to conceal or reveal its cost. To obtain sharper predictions, we focus on a situation where customers appraise the markup they face as fair, which they may be particularly likely to do in the long run, once they have adapted to the environment.¹⁸ Being in an equilibrium in which customers are neither angry nor happy imposes constraints on the fairness and belief parameters. We take the parameters ϵ , θ , and γ as given. For customers to find the markup fair when the firm reveals its cost, it must be that $F(K^r) = 1$, which requires $K^r = K^f$ and therefore

$$(A10) \quad K^f = 1 + \frac{1}{\epsilon - 1} \cdot \frac{1}{1 + \theta/2}.$$

For customers to also perceive the markup as fair if the firm conceals, the parameters MC^b and K^b must be such that $K^p = K^f$. We have seen that there is a unique value of $\lambda = (K^b)^{1-\gamma} / (MC^b)^\gamma$ such that this happens. When customers are acclimated, they perceive the same markup whether the firm conceals or reveals its cost.

We now study whether the monopoly chooses to conceal or reveal when customers are in such acclimated situation. We find that the monopoly always prefers to conceal:

PROPOSITION A1. *Assume that customers care about fairness ($\theta > 0$) and are acclimated (K^f is given by (A10) and MC^b and K^b are such that in equilibrium $K^p = K^f$). If customers make some inference ($\gamma < 1$), then the monopoly always chooses to conceal its marginal cost: $V^c > V^r$. If customers do not make any inference ($\gamma = 1$), then the monopoly is indifferent between concealing and revealing its marginal cost: $V^c = V^r$.*

Proof. We first compute profits when the firm reveals its cost. Since $K^r = K^f$, $F(K^r) = 1$, and equation (A8) implies that profits are

$$V^r = MC^{1-\epsilon} \cdot (K^r - 1) \cdot (K^r)^{-\epsilon}.$$

¹⁸As noted by Kahneman, Knetsch, and Thaler (1986, pp. 730–731), “Psychological studies of adaption suggest that any stable state of affairs tends to become accepted eventually, at least in the sense that alternatives to it no longer come to mind. Terms of exchange that are initially seen as unfair may in time acquire the status of a reference transaction. . . . [People] adapt their views of fairness to the norms of actual behavior.”

Since $K^r = K^f$, $\Phi(K^r) = \theta/2$, and equation (A7) implies that the profit-maximizing markup is

$$K^r = 1 + \frac{1}{(\epsilon - 1)(1 + \theta/2)}.$$

Following the same logic, and using equations (A9) and (10), we can compute the profits and profit-maximizing markup when the firm conceals its cost:

$$V^c = MC^{1-\epsilon} \cdot (K^c - 1) \cdot (K^c)^{-\epsilon}$$

$$K^c = 1 + \frac{1}{(\epsilon - 1)(1 + \gamma\theta/2)}.$$

If $\gamma = 1$, then $K^c > K^r$ and $V^c = V^r$. But for any $\theta > 0$ and $\gamma < 1$, then $K^c > K^r$. Since the function $(K - 1)K^{-\epsilon}$ is strictly increasing in K for $K \in [1, \epsilon/(\epsilon - 1)]$, the result that $K^c > K^r$ implies that $V^c > V^r$: it is more profitable to conceal marginal costs. \square

The proposition says that if the monopoly had to choose between concealing and revealing its cost when customers are acclimated, which is likely to happen in the long run once customers have adapted to the markups and find them fair, then the monopoly would always choose to conceal its cost. Indeed, once customers are acclimated, they find transaction equally fair, whether the firm reveals or conceals, and the level of demand is the same in both situations. However, the demand is less elastic when the firm conceals: an increase in price signals some increase in cost, and triggers a smaller increase in perceived marginal cost as when the firm reveals. As the monopoly faces a more inelastic demand when it conceals its cost, it is able to extract higher profits.

At the limit where customers do not make any inference about marginal costs, however, the demand is as elastic whether the firm conceals or reveals, and the firm makes as much profits whether it conceals or reveals.

Proposition A1 shows that when customers are acclimated, it is more profitable for the monopoly to conceal its cost. Finally, we examine whether, starting from this situation, the monopoly may choose to reveal its cost in response to an increase or decrease in cost. We find that the monopoly will reveal its cost for a large-enough increase in marginal cost:

PROPOSITION A2. *Assume that customers care about fairness ($\theta > 0$) and are acclimated to some marginal cost \overline{MC} (K^f is given by (A10) and MC^b and K^b are such that in equilibrium $K^p = K^f$). At $MC = \overline{MC}$, the monopoly optimally conceals its cost. Then if customers underinfer marginal costs ($\gamma \in (0, 1)$), there exists a threshold $MC_0 > \overline{MC}$ such that the firm optimally conceals any marginal cost $MC < MC_0$ and reveals any marginal cost $MC > MC_0$.*

If customers do not make any inference ($\gamma = 1$), the threshold satisfies $MC_0 = \overline{MC}$, such that the firm reveals any cost increase but conceals any cost decrease. And if customers proportionally infer marginal costs ($\gamma = 0$), the firm never reveals its marginal cost.

Proof. Consider that the marginal cost is at some initial value \overline{MC} . Customers are acclimated to this marginal cost such that $K^r = K^f = K^p$. We have seen in Proposition A1 that in this situation, it is optimal for the firm to conceal its cost. Hence, $V^c(\overline{MC}) > V^r(\overline{MC})$. We now study what happens when the marginal cost MC departs from the initial value \overline{MC} .

We study the ratio of profits V^c/V^r . We have seen that at $MC = \overline{MC}$, $V^c/V^r > 1$. We determine how the ratio evolves when MC departs from \overline{MC} . Since V^c is given by equation (A9) and V^r by equation (A8), we have

$$\frac{V^c}{V^r} = \frac{\gamma(K^c)}{\gamma(K^r)} \cdot \frac{F(K^p)}{F(K^r)}$$

where $\gamma(K) = (K - 1)/K^\epsilon$, K^c satisfies (10), K^p is given by (8), and K^r is given by (A7). The auxiliary function $\gamma(K)$ is strictly increasing for $K \in [1, \epsilon/(\epsilon - 1)]$.

When MC increases, the following happens. First, K^r does not change so $F(K^r)$ and $\gamma(K^r)$ remain unchanged. Second, P also increases because the passthrough of marginal costs into prices (σ) is positive; then, as P increases, K^p increases. Third, the increase in K^p leads $F(K^p)$ to fall. Fourth, the increase in K^p leads $\Phi(K^p)$ to rise, K^c to fall, and $\gamma(K^c)$ to fall. To conclude, when MC increases, V^c/V^r decreases.

A first implication is that for all $MC < \overline{MC}$, then $V^c/V^r > 1$: it remains more profitable to conceal when marginal costs fall.

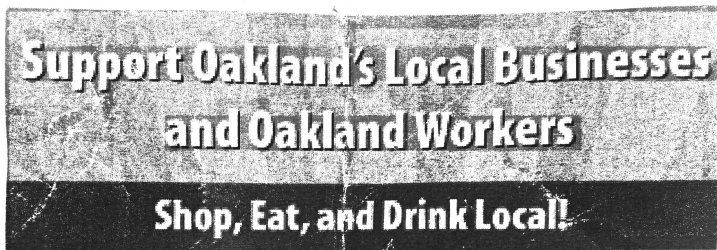
In addition, when $MC \rightarrow \infty$, then $P \rightarrow \infty$ (since $P \geq MC$), so $K^p \rightarrow \infty$, and $F(K^p) \rightarrow 0$. This implies that when $MC \rightarrow \infty$, then $V^c/V^r \rightarrow 0$. Since $V^c/V^r > 1$ for $MC = \overline{MC}$, V^c/V^r is strictly decreasing in MC , and $V^c/V^r \rightarrow 0$ when $MC \rightarrow \infty$, then there is a unique MC_0 such that $V^c/V^r > 1$ for any $MC < MC_0$ and $V^c/V^r < 1$ for any $MC > MC_0$. It is more profitable to reveal if and only if marginal costs are above MC_0 . \square

There is a simple logic for the result. We consider that the monopoly conceals its cost and that customers have adapted to the situation. Then we examine what happens if the marginal cost changes—either increases or decreases. A firm with a high marginal cost will tend to charge high prices. As the customer fails to adequately update beliefs about marginal cost from price, if the firm conceals its marginal cost, it will be wrongly perceived to use a high markup and regarded as unfair. On the other hand, a firm with a low marginal cost that conceals it will be wrongly perceived to use a low markup and regarded as fair. The former clearly has more incentive to reveal than the latter.

Starting from an equilibrium in which customers are acclimated and firms have the ability to signal their marginal costs, what happens to the marginal-cost passthrough? Let's focus on customers underinferring marginal costs. Before the cost shock, the monopoly conceals. If the marginal cost falls, the monopoly keeps concealing, so the passthrough is given by (13), which is less than one. This means that in response to a decrease in cost, prices are somewhat rigid. If the marginal cost rises but remains below MC_0 , the monopoly keeps concealing, so the passthrough is also given by (13). This means that in response to a small increase in cost, prices are also somewhat rigid. Finally, if the marginal cost rises above MC_0 , the monopoly switches from concealing to revealing. The original price is $\bar{P} = K^c \cdot \overline{MC}$ and the final price is $P = K^r \cdot MC$ so the marginal-cost passthrough is strictly less than one when $K^r < K^c$. As showed by (A10) and (12), since $\gamma < 1$, $K^r = K^f < K^c$, and the passthrough is indeed less than one even though the monopoly switches to revealing its costs. Hence prices are also somewhat rigid even in response to a large increase in cost. Depending on parameter values, the price rigidity may be more or less pronounced in response to a large cost increase, but of course as the marginal cost becomes infinitely large, the firm will eventually reveal its cost and the passthrough will converge to one.

In the special case in which customers make proportional inference, firms always conceal, and the passthrough is always one: hence prices are flexible. And in the special case in which customers do not make any inference, firms conceal cost decreases but reveal cost increases, so prices are downwardly rigid but upwardly flexible.

Our model thus predicts that in general a monopoly would not want to reveal its costs (Proposition A1). This could explain why we rarely see firms reveal their costs to customers. Our model also predicts that in response to a large-enough increase in production costs (but not a decrease in production costs), a monopoly will reveal its costs (Proposition A2). Indeed, there is evidence to this effect: while we have never observed a firm advertise a decrease in production costs, we frequently observe firms advertising cost increase. In fact, Okun (1981, p. 153) observed that "In many industries, when firms raise their prices, they routinely issue announcements to their customers, insisting that higher costs have compelled them to do so." Figure A1, panel A, presents examples of firms that reveal their costs in response to a substantial increase in labor costs. The two signs in the figure were posted in restaurants in Oakland and Berkeley in California following the more than 30% increase in minimum wage enacted there in March 2015. Many businesses responded by increasing prices; many also felt compelled to explain why. Figure A1, panel B, shows that some firms go to great lengths to document large increases in production costs. The figure comprises two displays posted side-by-side in a bakery in Ithaca, NY. The first reproduces several graphs from the *New York Times*, which plot the price

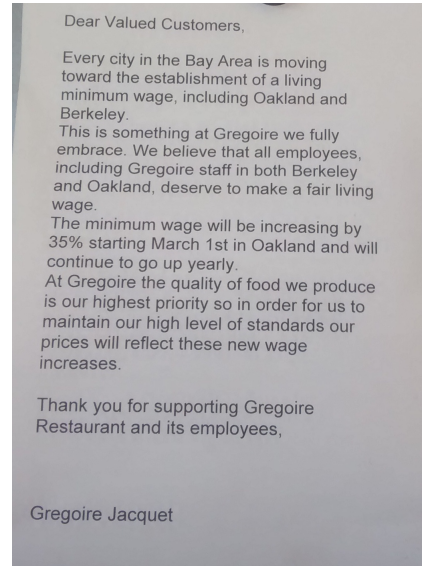


On March 2, 2015 Oakland's minimum wage increased from \$9 to \$12.25. Many businesses including Juhu Beach Club increased prices in response to increasing costs. Restaurants like ours, whose operations are labor-intensive, have raised prices more than most other businesses.

Juhu Beach Club is co-owned by two women, partners in life and business who are committed to creating a great workplace for all of our employees. Paying our staff fairly and fairly relative to others is an area we thoughtfully manage in our operation.

Thank you for supporting Juhu and our amazing team!

-Chef Preeti Mistry & Ann Nadeau



A. Large increases in minimum wage



February 28, 2008

TO OUR VALUED CUSTOMERS

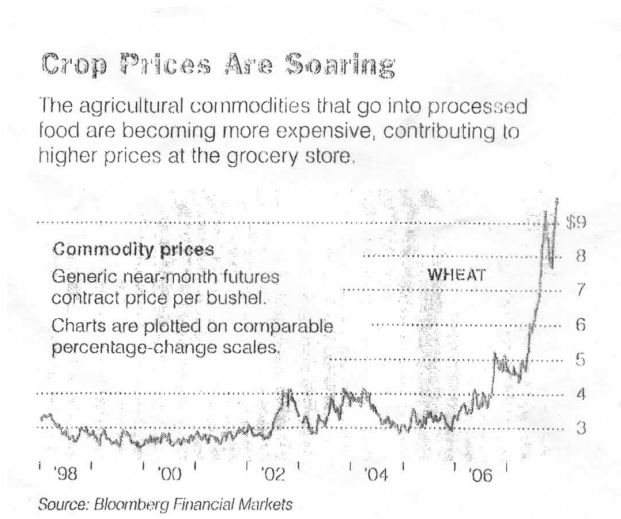
Wheat is continuing to hit record prices, vastly increasing our costs for flour. To cope with this, we are forced to impose a surcharge on bread and bagels, effective immediately. This will include sandwiches. Each week, we will recalculate the surcharge, according to the price of wheat. We hope that this will be temporary, but industry experts do not know when—or if—prices will stabilize.

- Our flour cost has more than tripled in the past month.
- On Monday (2/25/08) the price of March spring wheat on the Minneapolis Grain Exchange hit \$24 a bushel, double its cost two months ago and the highest price ever for wheat.
- The high-quality wheat we use to make artisan breads and bagels is getting harder to find.
- U.S. stocks of wheat are now at their lowest level in 60 years.

We can direct customers to substantial references for information about the wheat situation, online and in print.

When prices return to normal, we will drop the surcharge. Please bear with us as we try to address this very serious situation.

Sincerely,
The Brouse & Mehafeey Family



B. Large increase in wheat prices

Figure A1. Examples of Firms Revealing Large Cost Increases

Sources: Panel A: pictures taken at restaurants in Oakland, CA, and Berkeley, CA, in 2015 by Pascal Michailat. Panel B: picture taken at a bakery in Ithaca, NY, in 2008 by Daniel Benjamin.

of wheat, of soybeans, and of corn over time. The second explains that the increase in the wheat price translated into an increase in the price of flour, a key ingredient for bagels. The bakery promises to “drop the surcharge” when wheat prices return to normal.

A last implication of the model with signaling is that the response of prices to cost decreases and costs increases, especially large increases, may be asymmetric. As soon as the cost increase is large enough to make the firm reveal its cost, the price response is different from the response to a cost decrease of the same amplitude. In particular, as the cost increase becomes large, the passthrough becomes closer to one—that is, prices are close to flexible. In that situation, prices respond more strongly to cost increases than to cost decreases. This mechanism could explain some of the passthrough asymmetry documented by Benzarti et al. (2017) (see Section 2.1).

Appendix C. New Keynesian Model: Derivations

We derive various results related to the New Keynesian model of Section 4.

Optimal Pricing

Monopolistic firms set prices to maximize profits. We describe their optimal pricing strategy here. We start by deriving the demand faced by firms. To do that, we analyze the behavior of households.

Household j chooses

$$\left\{ W_j(t), N_j(t), [Y_{ij}(t)]_{i=0}^1, B_j(t) \right\}_{t=0}^{\infty}$$

to maximize (17) subject to the budget constraint (18), the labor-demand constraint $N_j(t) = N_j^d(t, W_j(t))$, and a solvency condition. Labor demand $N_j^d(t, W_j(t))$ gives the quantity of labor that firms would hire from household j in period t at a nominal wage $W_j(t)$. The household takes

$$\left\{ X(t), [F_i(t)]_{i=0}^1, [P_i(t)]_{i=0}^1, V_j(t) \right\}_{t=0}^{\infty}$$

as given. To solve household j 's problem, we set up the Lagrangian:

$$\begin{aligned} \mathcal{L}_j = & \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\ln(Z_j(t)) - \frac{N_j(t)^{1+\eta}}{1+\eta} \right. \\ & + \mathcal{A}_j(t) \left\{ W_j(t)N_j(t) + B_j(t-1) + V_j(t) - X(t)B_j(t) - \int_0^1 P_i(t)Y_{ij}(t)di \right\} \\ & \left. + \mathcal{B}_j(t) \left\{ N_j^d(t, W_j(t)) - N_j(t) \right\} \right] \end{aligned}$$

where $\mathcal{A}_j(t)$ is the Lagrange multiplier on the budget constraint in period t and $\mathcal{B}_j(t)$ is the Lagrange multiplier on the labor-demand constraint in period t .

We first compute the first-order conditions with respect to $Y_{ij}(t)$. We know that

$$\begin{aligned}\frac{\partial Z_{ij}}{\partial Y_{ij}} &= F_i \\ \frac{\partial Z_j}{\partial Z_{ij}} &= \left(\frac{Z_{ij}}{Z_j}\right)^{-1/\epsilon} di.\end{aligned}$$

Hence, the first-order conditions with respect to $Y_{ij}(t)$ are

$$(A11) \quad \left(\frac{Z_{ij}(t)}{Z_j(t)}\right)^{-1/\epsilon} \frac{F_i(t)}{Z_j(t)} = \mathcal{A}_j(t)P_i(t).$$

Manipulating and integrating the conditions (A11) over $i \in [0, 1]$, then using the definitions of Z_j and Q given by (14) and (15), we obtain

$$(A12) \quad \mathcal{A}_j(t)Q(t) = \frac{1}{Z_j(t)}.$$

Combining (A11) and (A12), we obtain the optimal consumption of good i for household j :

$$Y_{ij}(t) = \left(\frac{P_i(t)/F_i(t)}{Q(t)}\right)^{-\epsilon} \frac{Z_j(t)}{F_i(t)}.$$

Integrating the consumption of good i over all households yields the output of good i :

$$Y_i(t) = Z(t) \cdot F \left(\frac{P_i(t)}{MC_i^p(t)}\right)^{\epsilon-1} \cdot \left(\frac{P_i(t)}{Q(t)}\right)^{-\epsilon}.$$

Last, substituting $MC_i^p(t)$ by expression (16), we obtain the demand for good i :

$$Y_i^d(t, P_i(t), MC_i^p(t-1)) = Z(t) \cdot F \left([K^b]^{1-\gamma} \left[\frac{P_i(t)}{MC_i^p(t-1)} \right]^\gamma \right)^{\epsilon-1} \cdot \left(\frac{P_i(t)}{Q(t)}\right)^{-\epsilon}.$$

The derivatives of the function Y_i^d are

$$\begin{aligned} -\frac{\partial \ln(Y_i^d)}{\partial \ln(P_i)} &= \epsilon + (\epsilon - 1)\gamma\Phi(K_i^p) \equiv E_i(t) \\ \frac{\partial \ln(Y_i^d)}{\partial \ln(MC_i^p)} &= (\epsilon - 1)\gamma\Phi(K_i^p) = E_i(t) - \epsilon. \end{aligned}$$

where Φ is minus the elasticity of F , and is characterized by (2). The variable $E_i(t) > \epsilon$ is the absolute value of the price elasticity of the demand for good i .

The first-order condition with respect to $B_j(t)$ is

$$X(t)\mathcal{A}_j(t) = \beta\mathbb{E}_t [\mathcal{A}_j(t+1)].$$

Using equation (A12), we obtain

$$X(t) = \beta\mathbb{E}_t \left[\frac{Q(t)Z_j(t)}{Q(t+1)Z_j(t+1)} \right]$$

Since the wage set by household j depends on firms' demand for its labor, we turn to firms' problems before returning to the household's problem. Firm i chooses

$$\left\{ P_i(t), Y_i(t), [N_{ij}(t)]_{j=0}^1 \right\}_{t=0}^{\infty}$$

to maximize (21) subject to the production constraint (19), the demand constraint (24), and to the law of motion of beliefs (16). The firm takes

$$\left\{ A_i(t), [W_j(t)]_{j=0}^1, Q(t), Z(t) \right\}_{t=0}^{\infty}$$

as given. To solve firm i 's problem, we set up the Lagrangian:

$$\begin{aligned} \mathcal{L}_i &= \mathbb{E}_0 \sum_{t=0}^{\infty} \Gamma(t) \left[P_i(t)Y_i(t) - \int_0^1 W_j(t)N_{ij}(t)dj \right. \\ &\quad + C_i(t) \{ Y_i^d(t, P_i(t), MC_i^p(t-1)) - Y_i(t) \} + \mathcal{D}_i(t) \{ A_i(t)N_i(t)^\alpha - Y_i(t) \} \\ &\quad \left. + \mathcal{E}_i(t) \left\{ [MC_i^p(t-1)]^\gamma \left(\frac{P_i(t)}{K^b} \right)^{1-\gamma} - MC_i^p(t) \right\} \right] \end{aligned}$$

where $C_i(t)$ is the Lagrange multiplier on the demand constraint in period t , $\mathcal{D}_i(t)$ is the Lagrange multiplier on the production constraint in period t , and $\mathcal{E}_i(t)$ is the Lagrange multiplier on the

law of motion of the perceived marginal cost in period t .

Using the fact that

$$\frac{\partial N_i(t)}{\partial N_{ij}(t)} = \left(\frac{N_{ij}(t)}{N_i(t)} \right)^{-1/\nu} dj,$$

we find that the first-order conditions with respect to $N_{ij}(t)$ for all j are

$$(A13) \quad W_j(t) = \alpha \mathcal{D}_i(t) A_i(t) N_i(t)^{\alpha-1} \left(\frac{N_{ij}(t)}{N_i(t)} \right)^{-1/\nu}.$$

Manipulating and integrating the conditions (A13) over $j \in [0, 1]$, then using the definitions of N_i and W given by (20) and (27), we obtain

$$(A14) \quad \mathcal{D}_i(t) = \frac{W(t)}{\alpha A_i(t) N_i(t)^{\alpha-1}}.$$

Combining (A13) and (A14), we obtain the quantity of labor that firm i hires from household j :

$$N_{ij}(t) = \left(\frac{W_j(t)}{W(t)} \right)^{-\nu} N_i(t).$$

Integrating the quantities $N_{ij}(t)$ over all firms i yields the labor demand faced by household j :

$$N_j^d(t, W_j(t)) = \left(\frac{W_j(t)}{W(t)} \right)^{-\nu} N(t).$$

Having determined the demand for labor service j , we finish solving the problem of household j . The first-order conditions with respect to $N_j(t)$ and $W_j(t)$ are

$$\begin{aligned} N_j(t)^\eta &= \mathcal{A}_j(t) W_j(t) - \mathcal{B}_j(t) \\ \mathcal{A}_j(t) N_j(t) &= -\mathcal{B}_j(t) \frac{dN_j^d}{dW_j}. \end{aligned}$$

Combining these conditions, and using the fact that the elasticity of $N_j^d(t, W_j)$ with respect to W_j is $-\nu$, we find that

$$\begin{aligned} \mathcal{B}_j(t) &= \frac{N_j(t)^\eta}{\nu - 1} \\ W_j(t) &= \frac{\nu}{\nu - 1} \cdot \frac{N_j(t)^\eta}{\mathcal{A}_j(t)}. \end{aligned}$$

Using (A12), we find that household j sets its wage according to

$$\frac{W_j(t)}{Q(t)} = \frac{\nu}{\nu - 1} N_j(t)^\eta Z_j(t).$$

Next, we finish solving the problem of firm i . The first-order condition with respect to $Y_i(t)$ yields $P_i(t) = C_i(t) + \mathcal{D}_i(t)$. Using (A14), we obtain

$$C_i(t) = P_i(t) \left(1 - \frac{W(t)/P_i(t)}{A_i(t)\alpha N_i(t)^{\alpha-1}} \right).$$

Firm i 's nominal marginal cost is

$$(A15) \quad MC_i(t) = \frac{W(t)}{A_i(t)\alpha N_i(t)^{\alpha-1}}.$$

Hence, the first-order condition implies

$$C_i(t) = P_i(t) \left(1 - \frac{MC_i(t)}{P_i(t)} \right).$$

With the quasi elasticity $D_i(t) = K_i(t)/(K_i(t) - 1)$, we rewrite the first-order condition as

$$(A16) \quad C_i(t) = \frac{P_i(t)}{D_i(t)}.$$

The first-order condition with respect to $P_i(t)$ is

$$0 = Y_i(t) + C_i(t) \frac{\partial Y_i^d}{\partial P_i(t)} + (1 - \gamma) \mathcal{E}_i(t) \frac{MC_i^p(t)}{P_i(t)},$$

which implies

$$0 = 1 - \frac{C_i(t)}{P_i(t)} E_i(t) + (1 - \gamma) \frac{\mathcal{E}_i(t)}{Y_i(t) K_i^p(t)}.$$

Combining this equation with (A16) yields

$$(A17) \quad \frac{E_i(t)}{D_i(t)} - 1 = (1 - \gamma) \frac{\mathcal{E}_i(t)}{Y_i(t) K_i^p(t)}.$$

Finally, the first-order condition with respect to $MC_i^p(t)$ is

$$0 = \mathbb{E}_t \left[\frac{\Gamma(t+1)}{\Gamma(t)} C_i(t+1) \frac{\partial Y_i^d}{\partial MC_i^p} \right] + \gamma \mathbb{E}_t \left[\frac{\Gamma(t+1)}{\Gamma(t)} \mathcal{E}_i(t+1) \frac{MC_i^p(t+1)}{MC_i^p(t)} \right] - \mathcal{E}_i(t).$$

Multiplying this equation by $MC_i^p(t)/P_i(t)$, we get

$$0 = \mathbb{E}_t \left[\frac{\Gamma(t+1)}{\Gamma(t)P_i(t)} C_i(t+1)Y_i(t+1)(E_i(t+1) - \epsilon) + \gamma \frac{\Gamma(t+1)}{\Gamma(t)P_i(t)} \mathcal{E}_i(t+1)MC_i^p(t+1) \right] - \mathcal{E}_i(t) \frac{MC_i^p(t)}{P_i(t)}.$$

We now focus on a symmetric equilibrium, where $P_i(t) = P(t)$, $Z(t) = F(t)Y(t)$, and $Q(t) = P(t)/F(t)$. Using the definition of $\Gamma(t)$, given by (22), we find that in such equilibrium,

$$\frac{\Gamma(t+1)}{\Gamma(t)P_i(t)} = \beta \frac{Q(t)}{Q(t+1)P(t)} \cdot \frac{Z(t)}{Z(t+1)} = \frac{\beta}{P(t+1)} \cdot \frac{Y(t)}{Y(t+1)}.$$

Hence, the equation becomes

$$0 = \beta \mathbb{E}_t \left[C(t+1) \frac{Y(t)}{P(t+1)} (E(t+1) - \epsilon) + \gamma \mathcal{E}(t+1) \frac{Y(t)}{Y(t+1)} \cdot \frac{MC^p(t+1)}{P(t+1)} \right] - \mathcal{E}(t) \frac{MC^p(t)}{P(t)}.$$

Using (A16) and $K^p(t) = P(t)/MC^p(t)$, and dividing by $Y(t)$, we now obtain

$$0 = \beta \mathbb{E}_t \left[\frac{E(t+1) - \epsilon}{D(t+1)} + \gamma \frac{\mathcal{E}(t+1)}{Y(t+1)K^p(t+1)} \right] - \frac{\mathcal{E}(t)}{Y(t)K^p(t)}.$$

Finally, multiplying by $1 - \gamma$ and using (A17), we get

$$0 = \beta \mathbb{E}_t \left[(1 - \gamma) \frac{E(t+1) - \epsilon}{D(t+1)} + \gamma \frac{E(t+1)}{D(t+1)} - \gamma \right] - \frac{E(t)}{D(t)} + 1.$$

Rearranging the terms, we finally obtain

$$\beta \mathbb{E}_t \left[\frac{E(t+1) - (1 - \gamma)\epsilon}{D(t+1)} \right] = \frac{E(t)}{D(t)} - (1 - \gamma\beta).$$

This forward-looking equation gives the quasi elasticity $D(t)$ and thus the optimal markup $K(t)$.

Equilibrium Dynamics

We log-linearize the conditions describing a symmetric equilibrium. First, we rework (16) to obtain a law of motion for the perceived markup $K^p(t) = P(t)/MC^p(t)$. We find that

$$K^p(t) = \left(K^b\right)^{1-\gamma} \cdot (K^p(t-1))^\gamma \cdot \left(\frac{P(t)}{P(t-1)}\right)^\gamma.$$

Taking the log of this equation, and using $\pi(t) = p(t) - p(t-1)$, we find

$$(A18) \quad k^P(t) = (1 - \gamma)k^b + \gamma \cdot [\pi(t) + k^P(t-1)].$$

Subtracting the steady-state values of both sides yields the law of motion of the perceived markup:

$$(A19) \quad \widehat{k}^P(t) = \gamma \left[\widehat{\pi}(t) + \widehat{k}^P(t-1) \right].$$

Second, we take the log of the Euler equation (25):

$$y(t) = \mathbb{E}_t [y(t+1)] - (i(t) - \mathbb{E}_t [\pi(t+1)] - \rho).$$

Combining this equation with the monetary-policy rule (23) yields

$$y(t) + \mu\pi(t) = \mathbb{E}_t [y(t+1)] + \mathbb{E}_t [\pi(t+1)] + \rho - i_0(t).$$

Subtracting the steady-state values of both sides, we rewrite the equation as

$$(A20) \quad \widehat{y}(t) + \mu\widehat{\pi}(t) = \mathbb{E}_t [\widehat{y}(t+1)] + \mathbb{E}_t [\widehat{\pi}(t+1)] - \widehat{i}_0(t).$$

Then we log-linearize equation (19):

$$(A21) \quad \widehat{y}(t) = \widehat{a}(t) + \alpha\widehat{n}(t).$$

Combining this equation with (A20) yields the dynamic IS equation:

$$(A22) \quad \alpha\widehat{n}(t) + \mu\widehat{\pi}(t) = \alpha\mathbb{E}_t [\widehat{n}(t+1)] + \mathbb{E}_t [\widehat{\pi}(t+1)] - \widehat{i}_0(t) - \widehat{a}(t) + \mathbb{E}_t [\widehat{a}(t+1)].$$

Finally, we log-linearize (29):

$$\widehat{e}(t) = \frac{\overline{E} - \epsilon}{\overline{E}} \cdot \frac{d \ln(\Phi)}{d \ln(K^P)} \widehat{k}^P(t).$$

Furthermore, the elasticity of Φ with respect to K^P is

$$(A23) \quad \frac{d \ln(\Phi)}{d \ln(K^P)} = \theta - \Phi.$$

Thus, we have $\widehat{e}(t) = \Omega_0 \widehat{k}^p(t)$ where

$$\Omega_0 = \frac{\bar{E} - \epsilon}{\bar{E}} (\theta - \bar{\Phi}) = \frac{(\epsilon - 1)(\theta - \bar{\Phi})\gamma\bar{\Phi}}{\epsilon + (\epsilon - 1)\gamma\bar{\Phi}}.$$

Next, $D = K/(K - 1)$, so in log-linear form, $\widehat{d}(t) = -\Omega_1 \widehat{k}(t)$, where

$$\Omega_1 = \bar{D} - 1 = \frac{1}{\bar{K} - 1} = (\epsilon - 1) \left[1 + \frac{(1 - \beta)\gamma\bar{\Phi}}{1 - \beta\gamma} \right].$$

(We have used (37) to get the value of \bar{K} .) Finally, we log-linearize (31):

$$\widehat{e}(t) - \widehat{d}(t) = \Omega_3 \mathbb{E}_t [\widehat{e}(t + 1)] - \Omega_2 \mathbb{E}_t [\widehat{d}(t + 1)],$$

where

$$\begin{aligned} \Omega_3 &= \left[\frac{\bar{E} - (1 - \gamma\beta)\bar{D}}{\bar{E}} \right] \left[\frac{\bar{E}}{\bar{E} - (1 - \gamma)\epsilon} \right] = \beta \\ \Omega_2 &= \frac{\bar{E} - (1 - \gamma\beta)\bar{D}}{\bar{E}} = \beta\gamma \cdot \frac{\epsilon + (\epsilon - 1)\bar{\Phi}}{\epsilon + (\epsilon - 1)\gamma\bar{\Phi}}. \end{aligned}$$

To simplify Ω_3 and Ω_2 , we have used the following results:

$$\begin{aligned} \bar{D} &= 1 + (\epsilon - 1) \left[1 + \frac{(1 - \beta)\gamma\bar{\Phi}}{1 - \beta\gamma} \right] \\ (1 - \gamma\beta)\bar{D} &= (1 - \gamma\beta) + (\epsilon - 1) \left[(1 - \gamma\beta) + (1 - \beta)\gamma\bar{\Phi} \right] \\ &= (1 - \gamma\beta)\epsilon + (\epsilon - 1)(1 - \beta)\gamma\bar{\Phi} \\ \bar{E} - (1 - \gamma\beta)\bar{D} &= \beta\gamma \left[\epsilon + (\epsilon - 1)\bar{\Phi} \right] \\ \bar{E} - (1 - \gamma)\epsilon &= \gamma \left[\epsilon + (\epsilon - 1)\bar{\Phi} \right]. \end{aligned}$$

Next we log-linearize equation (32):

$$(A24) \quad \widehat{k}(t) = -(1 + \eta)\widehat{n}(t).$$

This implies that $\widehat{d}(t) = (1 + \eta)\Omega_1 \widehat{n}(t)$. Combining these results, we obtain

$$\Omega_0 \widehat{k}^p(t) - (1 + \eta)\Omega_1 \widehat{n}(t) = \beta\Omega_0 \mathbb{E}_t [\widehat{k}^p(t + 1)] - (1 + \eta)\Omega_1 \Omega_2 \mathbb{E}_t [\widehat{n}(t + 1)].$$

We define

$$\Lambda_1 \equiv (1 + \eta) \frac{\Omega_1}{\Omega_0} = (1 + \eta) \cdot \frac{1}{\gamma} \cdot \frac{\epsilon + (\epsilon - 1)\gamma\bar{\Phi}}{(\theta - \bar{\Phi})\bar{\Phi}} \left[1 + \frac{(1 - \beta)\gamma}{1 - \beta\gamma} \bar{\Phi} \right]$$

$$\Lambda_2 \equiv (1 + \eta) \frac{\Omega_1\Omega_2}{\Omega_0} = (1 + \eta) \cdot \beta \cdot \frac{\epsilon + (\epsilon - 1)\bar{\Phi}}{(\theta - \bar{\Phi})\bar{\Phi}} \left[1 + \frac{(1 - \beta)\gamma}{1 - \beta\gamma} \bar{\Phi} \right].$$

Using (A19) and these definitions, we obtain the short-run Phillips curve from the last equation:

$$(A25) \quad (1 - \beta\gamma)\gamma\widehat{k^p}(t - 1) + (1 - \beta\gamma)\gamma\widehat{\pi}(t) - \Lambda_1\widehat{n}(t) = \beta\gamma\mathbb{E}_t[\widehat{\pi}(t + 1)] - \Lambda_2\mathbb{E}_t[\widehat{n}(t + 1)].$$

Equations (A19), (A22), and (A25) jointly determine employment $\widehat{n}(t)$, inflation $\widehat{\pi}(t)$, and perceived markup $\widehat{k^p}(t)$. The other variables are directly obtained from these three variables.

To conclude, we combine the log-linear equilibrium conditions to obtain a dynamical system describing equilibrium dynamics. Combining (A19), (A22), and (A25), we obtain a system of difference equations:

$$\begin{bmatrix} \gamma & \gamma & 0 \\ 0 & \mu & \alpha \\ (1 - \beta\gamma)\gamma & (1 - \beta\gamma)\gamma & -\Lambda_1 \end{bmatrix} \begin{bmatrix} \widehat{k^p}(t - 1) \\ \widehat{\pi}(t) \\ \widehat{n}(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha \\ 0 & \beta\gamma & -\Lambda_2 \end{bmatrix} \begin{bmatrix} \widehat{k^p}(t) \\ \mathbb{E}_t[\widehat{\pi}(t + 1)] \\ \mathbb{E}_t[\widehat{n}(t + 1)] \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \epsilon(t),$$

where

$$\epsilon(t) \equiv \widehat{i}_0(t) + \widehat{a}(t) + \mathbb{E}_t[\widehat{a}(t + 1)]$$

is an exogenous shock realized at time t . The inverse of the matrix on the right-hand side is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha \\ 0 & \beta\gamma & -\Lambda_2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\Lambda_2}{\Lambda_2 + \alpha\beta\gamma} & \frac{\alpha}{\Lambda_2 + \alpha\beta\gamma} \\ 0 & \frac{\beta\gamma}{\Lambda_2 + \alpha\beta\gamma} & \frac{-1}{\Lambda_2 + \alpha\beta\gamma} \end{bmatrix}.$$

Premultiplying the system of difference equations by the inverse matrix, we rewrite the system as follows:

$$\begin{bmatrix} \widehat{k^p}(t) \\ \mathbb{E}_t[\widehat{\pi}(t + 1)] \\ \mathbb{E}_t[\widehat{n}(t + 1)] \end{bmatrix} = \mathbf{A} \begin{bmatrix} \widehat{k^p}(t - 1) \\ \widehat{\pi}(t) \\ \widehat{n}(t) \end{bmatrix} + \mathbf{B} \cdot \epsilon(t)$$

$$\mathbf{A} \equiv \begin{bmatrix} \gamma & \gamma & 0 \\ \frac{(1-\beta\gamma)\alpha\gamma}{\Lambda_2+\alpha\beta\gamma} & \frac{\Lambda_2\mu+\alpha\gamma(1-\beta\gamma)}{\Lambda_2+\alpha\beta\gamma} & \frac{(\Lambda_2-\Lambda_1)\alpha}{\Lambda_2+\alpha\beta\gamma} \\ \frac{-(1-\beta\gamma)\gamma}{\Lambda_2+\alpha\beta\gamma} & \frac{[\beta(\mu+\gamma)-1]\gamma}{\Lambda_2+\alpha\beta\gamma} & \frac{\Lambda_1+\alpha\beta\gamma}{\Lambda_2+\alpha\beta\gamma} \end{bmatrix} \quad \text{and} \quad \mathbf{B} \equiv \begin{bmatrix} 0 \\ \frac{\Lambda_2}{\Lambda_2+\alpha\beta\gamma} \\ \frac{\beta\gamma}{\Lambda_2+\alpha\beta\gamma} \end{bmatrix}.$$

Calibration

Using the derivations above, we study the behavior of a single firm i that faces an exogenous marginal cost $MC_i(t)$ and prices optimally given its monopolistic competitors. This is a simplified version of the firm problem in the New Keynesian model, abstracting from hiring decisions. The objective of the analysis is to link three key parameters of the New Keynesian model— ϵ , θ , and γ —to the dynamic behavior of the passthrough. Then we will use the empirical evidence on passthrough dynamics discussed in Section 4.5 to calibrate the three parameters. The calibrated parameters are used in the simulations of the New Keynesian model presented in Section 4.6. To simplify here, we assume that there is no underlying inflation, so that $MC_i(t)$ and $P_i(t)$ are constant in steady state.

Firm i chooses $\{P_i(t), Y_i(t)\}_{t=0}^{\infty}$ to maximize

$$\mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \cdot (P_i(t) - MC_i(t)) \cdot Y_i(t) \right],$$

subject to the demand constraint

$$(A26) \quad Y_i^d(P_i(t), MC_i^p(t-1)) = AD \cdot F \left([K^b]^{1-\gamma} \left[\frac{P_i(t)}{MC_i^p(t-1)} \right]^\gamma \right)^{\epsilon-1} \cdot P_i(t)^{-\epsilon}.$$

and to the law of motion of beliefs (16). The firm takes nominal marginal costs and aggregate demand $\{MC_i(t)\}_{t=0}^{\infty}$ as given. (We assume that the discount factor used by the firm at time t simply is β^t .) To solve firm i 's problem, we set up the Lagrangian:

$$\begin{aligned} \mathcal{L}_i = & \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[(P_i(t) - MC_i(t)) Y_i(t) \right. \\ & + C_i(t) \{ Y_i^d(P_i(t), MC_i^p(t-1)) - Y_i(t) \} \\ & \left. + \mathcal{E}_i(t) \left\{ [MC_i^p(t-1)]^\gamma \left(\frac{P_i(t)}{K^b} \right)^{1-\gamma} - MC_i^p(t) \right\} \right] \end{aligned}$$

where $C_i(t)$ is the Lagrange multiplier on the demand constraint in period t , and $\mathcal{E}_i(t)$ is the Lagrange multiplier on the law of motion of the perceived marginal cost in period t .

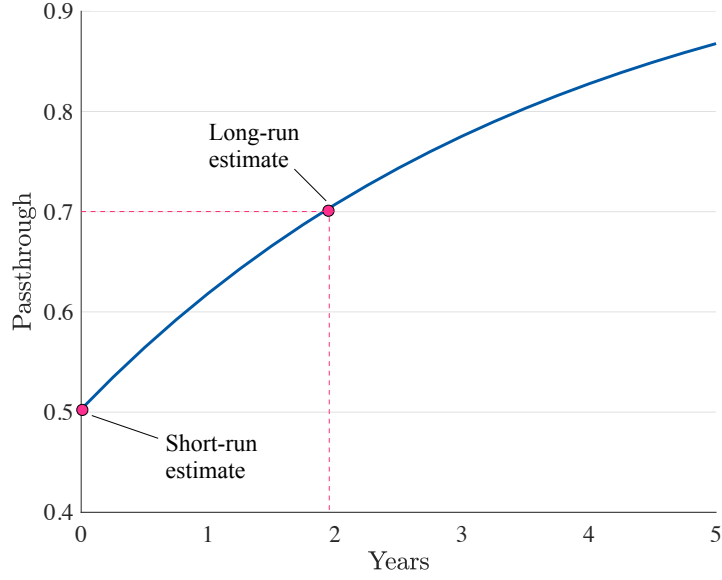


Figure A2. Simulated Passthrough Dynamics and Estimated Passthroughs

Notes: The simulated passthrough dynamics are obtained from the pricing model in Appendix C, under the calibration in Table 3. The passthrough estimates are obtained in Section 4.5.

The first-order condition with respect to $Y_i(t)$ yields

$$C_i(t) = P_i(t) \left(1 - \frac{MC_i(t)}{P_i(t)} \right).$$

With the quasi elasticity $D_i(t) = K_i(t)/(K_i(t) - 1)$, we rewrite the first-order condition as

$$(A27) \quad C_i(t) = \frac{P_i(t)}{D_i(t)}.$$

The first-order condition with respect to $P_i(t)$ is

$$0 = Y_i(t) + C_i(t) \frac{\partial Y_i^d}{\partial P_i(t)} + (1 - \gamma) \mathcal{E}_i(t) \frac{MC_i^p(t)}{P_i(t)},$$

which implies

$$0 = 1 - \frac{C_i(t)}{P_i(t)} E_i(t) + (1 - \gamma) \frac{\mathcal{E}_i(t)}{Y_i(t)} \cdot \frac{MC_i^p(t)}{P_i(t)},$$

where $E_i(t) \equiv \partial \ln(Y_i^d) / \partial \ln(P_i)$ is the price elasticity of demand. Combining this equation

with (A27) yields

$$(A28) \quad \frac{E_i(t)}{D_i(t)} - 1 = (1 - \gamma) \frac{\mathcal{E}_i(t)}{Y_i(t)} \cdot \frac{MC_i^p(t)}{P_i(t)}.$$

Finally, the first-order condition with respect to $MC_i^p(t)$ is

$$0 = \beta \mathbb{E}_t \left[C_i(t+1) \frac{\partial Y_i^d}{\partial MC_i^p} + \gamma \mathcal{E}_i(t+1) \frac{MC_i^p(t+1)}{MC_i^p(t)} \right] - \mathcal{E}_i(t).$$

Multiplying this equation by $(1 - \gamma)MC_i^p(t)$, and using $\partial \ln(Y_i^d)/\partial \ln(MC_i) = E_i - \epsilon$, we get

$$(1 - \gamma)\mathcal{E}_i(t)MC_i^p(t) = \beta \mathbb{E}_t \left[(1 - \gamma)C_i(t+1)Y_i(t+1)(E_i(t+1) - \epsilon) + \gamma(1 - \gamma)\mathcal{E}_i(t+1)MC_i^p(t+1) \right].$$

Hence, using (A27) and (A28), the equation becomes

$$\frac{Y_i(t)P_i(t)}{D_i(t)} (E_i(t) - D_i(t)) = \beta \mathbb{E}_t \left[\frac{Y_i(t+1)P_i(t+1)}{D_i(t+1)} \{ (1 - \gamma)(E_i(t) - \epsilon) + \gamma(E_i(t+1) - D_i(t+1)) \} \right].$$

We denote by

$$(A29) \quad V_i(t) \equiv Y_i(t) \cdot (P_i(t) - MC_i(t)) = \frac{Y_i(t)P_i(t)}{D_i(t)}$$

the profits of firm i in period t . We have

$$V_i(t) = Y_i(t) \cdot P_i(t) \cdot \left(1 - \frac{MC_i(t)}{P_i(t)} \right) = Y_i(t) \cdot P_i(t) \cdot \left(1 - \frac{1}{K_i(t)} \right) = \frac{Y_i(t) \cdot P_i(t)}{D_i(t)}.$$

Thus, the first-order condition simplifies to

$$(A30) \quad V_i(t) \cdot (E_i(t) - D_i(t)) = \beta \mathbb{E}_t [V_i(t+1) \cdot \{E_i(t) - (1 - \gamma)\epsilon - \gamma D_i(t+1)\}].$$

The eight equilibrium conditions describing firm i 's optimal pricing are equation (16), and equation (A26), equation (A30), equation (A29), $D_i(t) = K_i(t)/(K_i(t) - 1)$, $E_i(t) = \epsilon + (\epsilon - 1)\gamma\Phi(K_i^p(t))$, $K_i^p(t) = P_i(t)/MC_i^p(t)$, and $K_i(t) = P_i(t)/MC_i(t)$. The eight equilibrium variables $P_i(t)$, $MC_i^p(t)$, $Y_i(t)$, $V_i(t)$, $K_i(t)$, $K_i^p(t)$, $E_i(t)$, and $D_i(t)$. The firm takes as given the stochastic process for marginal cost, $MC_i(t)$.

To simulate passthrough dynamics, we solve this nonlinear dynamical system of eight equations (using Dynare). We assume that the firm is in steady state for some marginal cost \overline{MC}_i

and impose at time 0 an unexpected and permanent increase in MC_i by 1 percent.¹⁹ We compute the impulse responses of the equilibrium variables to this shock. Passthrough dynamics are directly obtained by computing the percentage change of firm i 's price over time: $\sigma_i(t) = (P_i(t) - \bar{P}_i) / \bar{P}_i \times 100$.

We use the simulations to calibrate our New Keynesian model. The three parameters that we need to calibrate are elasticity of substitution ϵ , the fairness concern θ , and the sophistication of inference γ . As discussed in Section 4.5, the three empirical moments that are matching to calibrate the parameters are a steady-state markup of 1.5, a passthrough of 40% on impact, and a passthrough of 76% after 2 years.

First, for any θ and γ , we set

$$\epsilon = 1 + \frac{1}{1.5 - 1} \cdot \frac{1}{1 + \frac{(1-\beta)\gamma}{1-\beta\gamma} \cdot \frac{\theta}{2}}.$$

This calibration ensures a steady-state markup of 1.5 (see equation (37), and note that we calibrate $\bar{K}^p = K^f$ so $\Phi(\bar{K}^p) = \theta/2$). Then, we repeat the simulation for various values of θ and γ until we obtain a passthrough of 40% on impact and 76% after 2 years. We match these two targets for $\theta = 75$ and $\gamma = 0.7$. The corresponding value of ϵ is $\epsilon = 2.1$. The simulated passthrough dynamics under this calibration are displayed in Figure A2.

Appendix D. The Standard New Keynesian Model

We describe the standard New Keynesian model that we use as a benchmark in the simulations presented in Figure 3. This model is borrowed from Galí (2008, Chapter 3).

The log-linear equations in the standard model are the same as in our fairness model, except two of them. First, the short-run Phillips curve (41) is replaced by

$$\widehat{\pi}(t) = \beta \mathbb{E}_t [\widehat{\pi}(t+1)] + \chi \alpha \widehat{n}(t)$$

where

$$\chi \equiv \frac{1 + \eta}{\alpha} \cdot \frac{(1 - \kappa)(1 - \beta\kappa)}{\kappa} \cdot \frac{\alpha}{\alpha + (1 - \alpha)\epsilon},$$

and κ is the fraction of firms keeping their prices unchanged each period.²⁰ This equation is obtained from equation (21) in Chapter 3 of Galí (2008), using the assumption that consumption utility is log. One of the implications of log utility is that the output gap in the standard model

¹⁹The steady-state values of the variables are the same as in our New Keynesian model.

²⁰We altered Galí's notation because some of his parameters were already used for other purposes in the paper.

is equal to $\alpha\widehat{n}(t)$, as showed in Section 4.4.

Second, since households rationally infer markups from prices, the perceived markup is not given by the law of motion (39) but follows the actual markup: $\widehat{k}^p(t) = \widehat{k}(t)$.

References

- Benzarti, Youssef, Dorian Carloni, Jarkko Harju, and Tuomas Kosonen. 2017. “What Goes Up May Not Come Down: Asymmetric Incidence of Value-Added Taxes.” http://conference.nber.org/confer/2017/SI2017/PE/Benzarti_Carloni_Harju_Kosonen.pdf.
- Gali, Jordi. 2008. *Monetary Policy, Inflation, and the Business Cycle*. Princeton, NJ: Princeton University Press.
- Kachelmeier, Steven J., Stephen T. Limberg, and Michael S. Schade Wald. 1991a. “Fairness in Markets: A Laboratory Investigation.” *Journal of Economic Psychology* 12 (3): 447–464.
- Kachelmeier, Steven J., Stephen T. Limberg, and Michael S. Schade Wald. 1991b. “A Laboratory Market Examination of the Consumer Price Response to Information about Producers’ Costs and Profits.” *Accounting Review* 66 (4): 694–717.
- Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler. 1986. “Fairness as a Constraint on Profit Seeking: Entitlements in the Market.” *American Economic Review* 76 (4): 728–41.
- Okun, Arthur M. 1981. *Prices and Quantities: A Macroeconomic Analysis*. Washington, DC: Brookings Institution.
- Renner, Elke, and Jean-Robert Tyran. 2004. “Price Rigidity in Customer Markets.” *Journal of Economic Behavior & Organization* 55 (4): 575–593.