Discussion Paper Series – CRC TR 224

Discussion Paper No. 743
Project C 02

The Misallocation Costs of Inflation: A Sufficient Statistics Approach

Klaus Adam[1]
Andrey Alexandrov[2]
Henning Weber[3]

May 2026

[1] University of Mannheim & CEPR
[2] Tor Vergata University of Rome
[3] Deutsche Bundesbank

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

The Misallocation Costs of Inflation: A Sufficient Statistics Approach*

Klaus Adam (University of Mannheim & CEPR)
Andrey Alexandrov (Tor Vergata University of Rome)
Henning Weber (Deutsche Bundesbank)

April 1, 2026

Abstract

The misallocation costs associated with different aggregate inflation rates can be estimated from micro price data via a set of sufficient statistics. We show that this works for a broad class of price-setting models and in the presence of unobserved product-level heterogeneity in pricing frictions and flexible prices. Applying the sufficient statistics approach to the micro price data underlying the U.K. consumer price index, we find large misallocation costs: aggregate productivity falls by about 1% if aggregate inflation is 8 percentage points above or below its optimal rate of 1.8%. Our findings provide important lessons for the calibration of sticky-price models: standard calibration targets can be uninformative about the sufficient statistics characterizing misallocation costs. To correctly capture these costs, models should be directly calibrated to the sufficient statistics that we uncover.

JEL Class. No.: E31, E58

*We thank Yvette Agape for valuable research assistance. Klaus Adam acknowledges support from the Deutsche Forschungsgemeinschaft through CRC-TR 224 (Project C02). The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem.

1 Introduction

The monetary models used in academia and central banks posit that excessively high rates of inflation distort relative prices, leading to a misallocation of demand and productive factors and a fall in aggregate productivity. These misallocation costs are central to the normative implications of monetary theory and motivate, for instance, the standard policy prescription that inflation should be kept low and stable.[1]

Despite their centrality in monetary theory, the quantitative importance of the misallocation costs associated with different inflation rates remains highly disputed, both in theory and in the data.

A theoretical assessment of misallocation costs through the lens of structural models is controversial because price-setting models differ in their assumptions about how prices react to inflation. As a result, misallocation costs depend strongly on the assumed form of price stickiness and on model parameterizations, with time-dependent pricing models typically generating much higher losses than state-dependent models (Nakamura, Steinsson, Sun and Villar (2018)).

An empirical assessment, in turn, is hindered by the fact that misallocation is determined by the gaps between actual prices and socially optimal prices. Importantly, these gaps are not directly observed because they depend on unobserved product characteristics. The literature typically controls for these unobservables via a number of fixed effects, assuming homogeneity in the dynamics of socially optimal prices and pricing frictions across products thereafter (Alvarez et al. (2019), Sheremirov (2020) and Sara-Zaror (2025)). However, it remains unclear whether these fixed effects fully control for unobserved heterogeneity.

In this paper, we develop a new approach for estimating misallocation costs associated with different aggregate inflation rates. This approach does not require taking a position on the particular form of price adjustment frictions and their parameterization. It also does not rely on homogeneity assumptions across products, leaving instead the cross-sectional distribution of socially optimal prices and pricing frictions almost entirely unrestricted.

We can nevertheless estimate the misallocation costs associated with different inflation rates via a set of sufficient statistics that can be recovered directly from micro price data. The set of pricing models covered by our approach includes the standard time- and state-dependent models used in the literature, their convex combinations, but also less standard models. In addition, our approach allows for rich unobserved heterogeneity at the level of individual products, including product-specific pricing frictions and product-specific

[1] See, for instance, Woodford (2003), Galí (2015), Adam and Weber (2019), and Acharya, Challe and Dogra (2023).

dynamics for productivity, mark-ups, and quality.

In a first step, we show that productivity losses associated with misallocation are determined by two key moments involving so-called price gaps, i.e., the gap between a product's actual (sticky) price and the product's hypothetical flexible price. Specifically, misallocation costs depend on the *variance of price gaps* and on the *covariance of price gaps with product-level mark-up shocks*.[2] Since the variance of price gaps and their covariance with mark-up shocks are endogenous to aggregate inflation, it is the response of these two moments to changes in inflation that determines how aggregate inflation affects misallocation.

In a second step, we show how the relationship between these two key moments and aggregate inflation can be recovered from panel micro price data by estimating a set of sufficient statistics. The statistics are sufficient in the sense that they summarize how productivity losses from misallocation depend on aggregate inflation for a broad class of underlying pricing frictions. To estimate the statistics, we do not have to rely on time variation in aggregate inflation. Instead, we levy cross-sectional heterogeneity of individual product price dynamics, which makes our estimation approach suitable for (but not limited to) environments with low and stable aggregate inflation rates.

We then apply our sufficient statistics approach to the micro price data underlying the construction of the U.K. consumer price index. We find that the annual inflation rate that minimizes misallocation is equal to 1.8%. It is positive because the flexible relative price decreases over the product life for the majority of products, causing some inflation to be optimal in the presence of sticky prices (Adam and Weber (2023)). Permanently higher or lower inflation rates generate virtually symmetric losses in aggregate productivity. These losses are negligible for aggregate inflation rates that are within a range of ±1 p.p. around the optimum (at most 0.02% of productivity), but increase steeply outside this range. For an inflation rate that is 8 p.p. higher or lower than the optimal rate, aggregate productivity losses reach about 1% per period. These results turn out to be highly stable across a number of alternative specifications for estimating the sufficient statistics and are quantitatively similar across different components of the consumer price index. This holds, for instance, for goods and services, even though goods prices adjust much more frequently than services prices.

In general, our results suggest that suboptimal inflation rates generate substantial misallocation costs. However, workhorse sticky-price models, calibrated in a standard way

[2] In contrast to product-level productivity and quality shocks, which are efficient, mark-up shocks are inefficient and distort the optimal allocation even under flexible prices. Price stickiness can either exacerbate or mitigate those distortions, depending on the sign of the covariance between mark-ups and price gaps, which explains why the covariance with mark-up shocks is relevant for misallocation costs.

to our data, may generate much lower costs. To illustrate this point, we calibrate a Calvo-plus model, which features time- and state-dependent price adjustments, to our micro price data, targeting standard moments of the price adjustment distribution. Surprisingly, the calibrated model generates productivity losses that are an order of magnitude lower than our empirical estimates. Increasing inflation 8 p.p. above its optimal level leads in the calibrated model to a productivity loss of only 0.06% instead of the 1% we estimate for the U.K. economy. Importantly, this difference is not due to our estimation procedure: applying the sufficient statistics approach to simulated data from the calibrated model correctly recovers this much lower loss.

This provides an important lesson for the calibration of sticky-price models: standard approaches to bring these models to the data may fail to capture the true misallocation costs of inflation. This is so because the models are typically calibrated to average data moments, e.g., the average price adjustment frequency or the average size of price adjustments. However, the misallocation costs of inflation are instead determined by the *sensitivity* of the variance and covariance of price gaps with respect to aggregate inflation, as captured by the sufficient statistics. We show that the typically targeted average moments are not informative about these sufficient statistics. This offers an important insight: if one seeks to construct models that capture the empirically observed misallocation costs of inflation, then one should calibrate directly to the sufficient statistics that we derive, rather than to the typically targeted data moments.

Our approach to evaluate misallocation costs of inflation differs from the previous literature, which relied predominantly on calibrated structural price-setting models to determine these costs, e.g., Coibion, Gorodnichenko and Wieland (2012). In an influential paper, Nakamura, Steinsson, Sun and Villar (2018) calibrate a menu cost model to match the median price adjustment frequency, the median absolute size of price adjustments, and how the absolute size of price adjustments varies with aggregate inflation, using the U.S. micro price data. With this approach, they conclude that the productivity and welfare losses of inflation are essentially zero. However, the targeted data moments are not part of the sufficient statistics that determine the misallocation costs associated with different inflation rates, which might explain the difference with our findings.

Our results that higher aggregate inflation rates generate misallocation costs aligns well with recent findings in Baqaee, Farhi and Sangani (2024) and Meier and Reinelt (2024) who infer price-induced misallocation from how product-specific mark-ups respond to inflationary demand shocks. Relatedly, several papers estimated the effects of inflation on *price dispersion*. Alvarez, Beraja, Gonzalez-Rozada and Neumeyer (2019) estimate a non-linear relationship between the cross-sectional dispersion of prices and inflation using data from Argentina. They find that cross-sectional price dispersion responds only weakly to inflation for inflation rates below 10%, but rises strongly for higher rates. Sheremirov (2020) uses U.S. supermarket scanner data and documents how local cross-sectional price dispersion correlates with local inflation over time. Sara-Zaror (2025) extends this empirical approach and documents that cross-sectional price dispersion strongly rises with the absolute deviation of inflation from zero, with the relationship becoming flatter for larger absolute inflation rates. Instead of estimating a relationship between the cross-sectional dispersion of prices and inflation over time, our sufficient statistics approach calls for estimating across-time dispersion of prices at the level of individual products and relating it to product-specific average price growth rates in the cross-section of products.

Although our method does not require recovering the price gap distribution, doing so might be an alternative approach to estimating misallocation costs. However, this requires rarely available marginal cost data or imposing strong assumptions on the flexible-price dynamics. Eichenbaum, Jaimovich and Rebelo (2011) exploit information on marginal costs and prices to identify mark-up distortions for U.S. supermarket products. Gardaglione, Gertler, Lenzu and Tielens (2025) follow a similar approach to recover price gaps for the Belgian manufacturing sector. Alvarez, Lippi and Oskolkov (2022) infer the price gap distribution from observed price changes assuming that pricing frictions are identical across products, and flexible prices follow a random walk with identical innovation variance and drift.[3] None of these papers estimates productivity losses associated with different inflation rates. In addition, our approach requires only data on posted prices and works under stationary flexible-price shocks. The latter is key because the random walk assumption is strongly rejected for the U.K. micro price data we use.

The remainder of the paper is structured as follows. Section 2 outlines demand and production structure of our framework. Section 3 introduces the general set of pricing frictions and the rich product-level heterogeneity. Section 4 derives analytical expressions for the productivity loss, the sufficient statistics capturing the effects of inflation, and explains how the sufficient statistics can be recovered from micro price data. Section 5 presents the U.K. micro price data and section 6 reports our main empirical results. Section 7 draws lessons for the calibration of sticky-price models that seek to capture the misallocation costs associated with suboptimal inflation rates. Section 8 provides additional empirical results and robustness checks for the main findings. The conclusion provides an outlook on future work.

[3] Baley and Blanco (2021) and Oskolkov and Lippi (2026) apply similar methods to identify the capital gap distribution from investment dynamics.

2 The Economic Framework

This section presents a general economic framework with a finite number of expenditure items and a continuum of products within each item. The demand structure is standard and captures how national statistical agencies construct the aggregate price index: it features Cobb-Douglas aggregation across expenditure items and CES aggregation within items, as in La'O and Tahbaz-Salehi (2022) and Adam and Weber (2023). The setup additionally allows for substantial heterogeneity in the dynamics of productivity, quality and flexible-price mark-ups across products and expenditure items. In the following, we outline the model structure relegating the details to appendix A.1.

2.1 Demand Structure

Time is discrete ($t = 0, 1, \dots$) and aggregate consumption C_t is a Cobb-Douglas aggregate of different expenditure categories C_{kt}, where $k \in \{1, ..., K\}$ indexes - in the language of statistical agencies - "expenditure items" or "items" in short. Each item k captures a certain product category, e.g., flatscreen TVs or vegetarian main courses. Letting $\psi_k \geq 0$ denote the expenditure weight of item k, with $\sum_k \psi_k = 1$, we have

$$C_t = \prod_{k=1}^{K} (C_{kt})^{\psi_k} .$$

Each item k is composed of many products $j \in [0, 1]$, i.e., specific TV models or specific vegetarian main courses, with item-level consumption being a Dixit-Stiglitz aggregate of the individual products in the item,

$$C_{kt} = \left(\int_0^1 (Q_{jkt} C_{jkt})^{\frac{\theta-1}{\theta}} \mathrm{dj} \right)^{\frac{\theta}{\theta-1}} ,$$

where C_{jkt} denotes the quantity consumed of product j (in physical units), Q_{jkt} its quality level, and $\theta > 1$ is the constant elasticity of substitution between products.

In line with micro price data, a share $\delta_k \in (0, 1)$ of products in item k exits the economy each period and gets replaced by new products to which we assign for simplicity the product indices of exiting products. Exit risk is i.i.d. over time and across products within an item, and might reflect changing consumer preferences or an exogenous loss of productivity. The quality of a product is given by

$$\ln Q_{jkt} = \ln Q_{jk}^0 + \ln \varepsilon_{jkt}^Q,$$

where $\ln Q_{jk}^0$ is a product-specific constant and $\ln \varepsilon_{jkt}^Q$ is an idiosyncratic stochastic quality component.[4]

[4]The constant is drawn at the time of product entry from a potentially time-varying item-specific

Letting P_{jkt} denote the nominal price of a physical unit of product j, the quality-adjusted price level P_{kt} for item k is given by

$$P_{kt} = \left(\int_0^1 \left(P_{jkt}/Q_{jkt} \right)^{1-\theta} \mathrm{dj} \right)^{\frac{1}{1-\theta}},$$

and the item-level inflation rate is denoted by

$$\pi_{kt} \equiv \ln(P_{kt}/P_{kt-1}). \tag{1}$$

The aggregate price level is given by

$$P_t = \prod_{k=1}^{K} \left(\frac{P_{kt}}{\psi_k} \right)^{\psi_k} \tag{2}$$

and the aggregate inflation rate is denoted by

$$\pi_t \equiv \ln(P_t/P_{t-1}). \tag{3}$$

Cost-minimization implies that the demand for product j satisfies

$$C_{jkt} = \left(\frac{P_{jkt}/Q_{jkt}}{P_{kt}} \right)^{-\theta} \frac{C_{kt}}{Q_{jkt}}, \tag{4}$$

and demand at the item level is

$$C_{kt} = \psi_k \left(\frac{P_{kt}}{P_t} \right)^{-1} C_t. \tag{5}$$

Analogously to item-level expenditure weight ψ_k, we define expenditure weight of product j in item k as:

$$\psi_{jkt} \equiv \frac{P_{jkt}C_{jkt}}{P_{kt}C_{kt}}. \tag{6}$$

In equilibrium, markets clear and product demand C_{jkt} equals product supply Y_{jkt} for all (j,k,t). Equivalent conditions hold at the item level ($Y_{kt} = C_{kt}$) and the aggregate level ($Y_t = C_t$).

2.2 Production Structure

Each product j in item k is produced with constant returns to scale

$$Y_{jkt} = A_{kt} G_{jkt}^{-1} L_{jkt}, \tag{7}$$

distribution, while the stochastic process driving $\ln \varepsilon^Q_{jkt}$ is drawn from a time-invariant set of item-specific stationary stochastic processes with mean zero. The value of $\ln \varepsilon^Q_{jkt}$ at the time of product entry is an independent draw from the product-specific stationary distribution.

where A_{kt} denotes item-specific productivity, G_{jkt}^{-1} product-specific productivity, and L_{jkt} is the labor input. Item-level productivity evolves according to

$$\ln A_{kt} = \ln \gamma_k + \ln A_{kt-1},$$

with γ_k capturing heterogeneity in productivity growth across items, e.g., the different productivity growth rates in the services versus goods components of the consumption basket. Product-specific (inverse) productivity evolves according to

$$\ln G_{jkt} = \ln G_{jk}^0 + s_{jkt} \cdot \ln \overline{G}_{jk} + \ln \varepsilon_{jkt}^G,$$

where $\ln G_{jk}^0$ is a product-specific constant, $\ln \overline{G}_{jk}$ is a product-specific time trend, $s_{jkt} \geq 0$ is product age, and $\ln \varepsilon_{jkt}^G$ is idiosyncratic stochastic productivity.[5] In line with the sticky-price literature, e.g., Midrigan (2011), Karadi and Reiff (2019), Cavallo, Lippi, and Miyahara (2024), Blanco, Boar, Jones, and Midrigan (2024), we assume that idiosyncratic quality and productivity shocks are inversely related:

$$\ln \varepsilon_{jkt}^G = \ln \varepsilon_{jkt}^Q.$$

Finally, we define labor input share of product j within item k as

$$\phi_{jkt} \equiv \frac{L_{jkt}}{L_{kt}} \tag{8}$$

where $L_{kt} = \int_0^1 L_{jkt} \mathrm{dj}$ is the item-level labor input.

2.3 Mark-ups and Flexible Prices

The real per-period profit (sales minus production costs) of a firm producing product j in item k is given by

$$\left((1+\tau_{jkt}) \frac{P_{jkt}}{P_t} - \frac{W_t}{A_{kt} P_t} G_{jkt} \right) Y_{jkt}, \tag{9}$$

where W_t denotes the nominal wage and τ_{jkt} is an output subsidy (or tax if negative). Suppose that firm j can set its price freely in every period to maximize the real per-period profit. From goods market clearing ($C_{jkt} = Y_{jkt}$) and the demand function (4) it follows that the optimal flexible product price P_{jkt}^* is equal to a mark-up over nominal marginal cost

$$P_{jkt}^* = \mu_{jkt}^f \frac{W_t}{A_{kt}} G_{jkt}, \tag{10}$$

where

$$\mu_{jkt}^f \equiv \frac{1}{1+\tau_{jkt}} \frac{\theta}{\theta - 1}$$

[5]The constant $\ln G_{jk}^0$ is drawn at the time of product entry from a potentially time-varying item-specific distribution and the time trend $\ln \overline{G}_{jk}$ from a time-invariant item-specific distribution of time trends.

is the flexible-price mark-up. We allow for product-specific flexible mark-up dynamics of the form

$$\ln \mu_{jkt}^{f} = \ln \mu_{jk}^{f,0} + s_{jkt} \cdot \ln \overline{\mu}_{jk}^{f} + \ln \varepsilon_{jkt}^{\mu},$$

where $\ln \mu_{jk}^{f,0}$ is a product-specific constant, $\ln \overline{\mu}_{jk}^{f}$ is a product-specific time trend, s_{jkt} denotes product age, and $\ln \varepsilon_{jkt}^{\mu}$ is an idiosyncratic stochastic mark-up component.[6]

2.4 Relative Price Trends on the Balanced Growth Path

We consider household preferences over consumption C_t and labor L_t that are consistent with a balanced growth path and in line with King, Plosser, and Rebelo (1988):

$$U(C_t, L_t) = \frac{[C_t \cdot V(L_t)]^{1-\sigma} - 1}{1-\sigma},$$

with $\sigma > 0$ and $V(L)$ ensuring strict concavity and Inada conditions. The representative household's time discount factor is given by $\beta \in (0,1)$.

On a balanced growth path, aggregate inflation is constant ($\pi_t = \pi$), the allocation of labor across items is constant over time ($L_{kt} = L_k$), item-level output Y_{kt} grows at the rate γ_k, and aggregate output Y_t grows at the rate γ of exogenous aggregate productivity,

$$\gamma \equiv \prod_{k=1}^{K} (\gamma_k)^{\psi_k}. \tag{11}$$

To ensure consistency of our setup with balanced growth, we impose a (necessary and sufficient) restriction on the dynamics of the joint within-item distribution of constants and trends present in the productivity, quality, and flexible-price mark-up processes, see condition 1 in appendix A.1.9.

On the resulting balanced growth path, the *quality-adjusted* flexible *relative* price

$$p_{jkt}^{*} \equiv \frac{P_{jkt}^{*}/Q_{jkt}}{P_{kt}}$$

evolves according to

$$\ln p_{jkt}^{*} = \ln p_{jk}^{*,0} - s_{jkt} \cdot \pi_{jk}^{*} + \ln \varepsilon_{jkt}^{\mu}, \tag{12}$$

where $\ln p_{jk}^{*,0}$ is an intercept that depends on the constants of the underlying fundamental processes and on the detrended real wage (which is constant on the balanced growth path):

$$\ln p_{jk}^{*,0} \equiv \ln \mu_{jk}^{f,0} + \ln G_{jk}^{0} - \ln Q_{jk}^{0} + \ln \frac{W_t}{P_{kt} A_{kt}}. \tag{13}$$

[6]The constant is drawn at the time of product entry from a potentially time-varying item-specific distribution, the time trend from a time-invariant item-specific distribution, and the stochastic process driving $\ln \varepsilon_{jkt}^{\mu}$ is drawn from a time-invariant set of item-specific stationary stochastic processes with mean zero. The value of $\ln \varepsilon_{jkt}^{\mu}$ at the time of product entry is an independent draw from the product-specific stationary distribution.

The time trend π^*_{jk} in equation (12) captures the average rate at which the flexible relative price changes over time; it is given by

$$\pi^*_{jk} \equiv -(\ln \overline{\mu}^f_{jk} + \ln \overline{G}_{jk}) \tag{14}$$

and depends on product-specific mark-up and productivity trends. In economic terms, π^*_{jk} captures the *desired rate of inflation* of the product in a setting with sticky prices, i.e., the rate of inflation that would cause relative prices to move (on average) at the rate of flexible prices, even in the absence of nominal price adjustments. Our setup allows desired inflation rates to differ across products within and across items. In fact, heterogeneity in desired inflation rates π^*_{jk} within an item will be key for our empirical approach later on.

2.5 Endogenous Aggregate Productivity

This section shows that aggregate productivity is endogenous and depends on relative prices. Specifically, item-level output is given by

$$Y_{kt} = \frac{A_{kt}}{\Delta_{kt}} L_k, \tag{15}$$

where Δ_{kt} denotes the endogenous component of item-level productivity, which captures the effects of misallocation associated with price stickiness and mark-ups. Since Δ_{kt} enters the denominator in equation (15), it can be interpreted as endogenous productivity *loss*: higher values of Δ_{kt} imply lower output per unit of labor. Naturally, endogenous productivity depends on product prices and is given by

$$\Delta_{kt} \equiv \int_0^1 \frac{G_{jkt}}{Q_{jkt}} (p_{jkt})^{-\theta} \, \mathrm{dj}, \tag{16}$$

with p_{jkt} denoting the sticky (quality-adjusted) relative price[7]

$$p_{jkt} \equiv \frac{P_{jkt}/Q_{jkt}}{P_{kt}}.$$

In the flexible-price *equilibrium*, the item-level endogenous productivity is given by

$$\Delta^f_{kt} \equiv \int_0^1 \frac{G_{jkt}}{Q_{jkt}} \left(p^*_{jkt} \frac{W^f_t / P^f_{kt}}{W_t / P_{kt}} \right)^{-\theta} \mathrm{dj}, \tag{17}$$

where W^f_t denotes the nominal wage and P^f_{kt} the item price level in the flexible-price equilibrium.[8] Aggregate output is given by

$$Y_t = \frac{A_t}{\Delta_t} L,$$

[7]We discuss price-setting frictions in the next section.

[8]Since p^*_{jkt} denotes the flexible price in the sticky-price equilibrium, it needs to be multiplied by the factor $(W^f_t/P^f_{kt})/(W_t/P_{kt})$ to transform it into the flexible price in the flexible-price equilibrium.

where $L = \sum_k L_k$,

$$A_t = \prod_{k=1}^{K} (\psi_k A_{kt})^{\psi_k} ,$$

and the endogenous aggregate productivity is given by

$$\Delta_t \equiv \prod_{k=1}^{K} \left(\frac{\psi_k}{\phi_k}\Delta_{kt}\right)^{\psi_k} , \tag{18}$$

where $\phi_k \equiv L_k/L$ is the item-level labor input share.

On the balanced growth path, endogenous productivity components are constant over time in each item, i.e., $\Delta_{kt} = \Delta_k$ and $\Delta_{kt}^f = \Delta_k^f$. As a result, the same holds true for the endogenous component of aggregate productivity, i.e., $\Delta_t = \Delta$ and $\Delta_t^f = \Delta^f$, where Δ_t^f denotes the endogenous component of aggregate productivity in the flexible-price equilibrium. From equation (5) it follows that the item-level inflation rates associated with a given aggregate inflation rate π are given by

$$\pi_k = \pi + \ln \frac{\gamma}{\gamma_k}. \tag{19}$$

With this demand and production structure, we state the assumptions on the underlying pricing frictions and within-item product heterogeneity in the next section.

3 Pricing Frictions and Product Heterogeneity

Our setup covers a broad class of pricing frictions, including time- and state-dependent frictions, commonly used in the literature. In addition, it allows for rich *product-level* heterogeneity *within* and *across* items, both in terms of pricing frictions and in terms of the stochastic processes driving mark-ups, productivity and quality.

3.1 The General Set of Admitted Pricing Frictions

Firms optimally choose the distribution of (history-dependent) price adjustment periods T_{jkh} $(h = 1, 2, ...)$ and the associated nominal prices P_{jkt} in these periods $(t = T_{jk1}, T_{jk2}, ...)$. Outside adjustment periods, the nominal price of the product remains constant so that inflation erodes the *relative* price. The choice of adjustment periods may be subject to costs or frictions, e.g., (random) menu costs, time-dependent restrictions, or both.

Specifically, the objective of firms is to maximize, subject to the product demand function

$$Y_{jkt} = \left(\frac{P_{jkt}/Q_{jkt}}{P_{kt}}\right)^{-\theta} \frac{Y_{kt}}{Q_{jkt}}, \tag{20}$$

the expected discounted sum of profits (sales minus production costs):

$$E_t \sum_{i=0}^{\infty} \beta^i \frac{U_{Ct+i}}{U_{Ct}} (1-\delta_k)^i \left((1+\tau_{jkt+i})\frac{P_{jkt+i}}{P_{t+i}} - \frac{W_{t+i}}{A_{kt+i}P_{t+i}} G_{jkt+i}\right) Y_{jkt+i}, \tag{21}$$

minus possible costs associated with price adjustments and subject to possible restrictions on the timing of adjustments. In equation (21), $\beta^i U_{Ct+i}/U_{Ct}$ denotes the discount factor of the representative household and δ_k is the product exit rate. We summarily refer to the price adjustment costs and the timing restrictions as 'pricing frictions', without modeling these frictions explicitly.[9] Instead, we consider a general set of pricing frictions satisfying the following assumption:

Assumption 1

1.1 *Pricing frictions are consistent with a balanced growth path on which aggregate and item-level variables have the same growth rates as under flexible prices.*

1.2 *Given the optimal distribution of adjustment periods T_{jkh} $(h = 1, 2, ...)$, the optimal reset price maximizes the sum of the expected discounted per-period profits* (9) *over the (stochastic) lifetime of the price.*

1.3 *The optimal distribution of adjustment periods T_{jkh} $(h = 1, 2, ...)$ is independent of product age and depends on aggregate inflation π only via the gap between actual and desired inflation $\pi_k - \pi^*_{jk} = \pi - \pi^*_{jk} + \ln(\gamma/\gamma_k)$.*

Assumption 1.1 is standard and rules out settings in which pricing frictions are not scale-invariant and therefore affect the growth rate of the economy.[10] Assumption 1.2 is also standard and applies to a wide class of models, including Calvo (1983), Taylor (1999), and fixed and random menu cost models (Golosov and Lucas (2007), Nakamura and Steinsson (2010), Alvarez, Lippi and Oskolkov (2022)). However, it rules out settings in which price adjustment costs depend on past prices, as for instance in Rotemberg (1982).[11] Assumption 1.3 is trivially satisfied with exogenous time-dependent frictions (Calvo (1983), Taylor (1999)), but also with menu costs that are, for instance, a (random) share of flexible-price sales, as we show in appendix A.7. This assumption rules out settings in which menu costs depend on product age or are expressed as a fixed share of the real wage as in Golosov and Lucas (2007).[12]

[9] Pricing frictions are the modeling element required to uniquely determine the distribution of a product's price adjustment times conditional on the history of shocks, see appendix A.1.7 for a definition.

[10] An example violating this assumption would be menu costs that are fixed in consumption units. In a growing economy, such costs are shrinking over time relative to firms' period profits and lead to different aggregate growth rates compared to the flexible-price economy.

[11] Optimal price setting then also needs to take into account how the current price choices affect current and future price adjustment costs.

[12] The latter is so because aggregate inflation generally affects the equilibrium real wage, and therefore would also affect the adjustment time distribution via its impact on effective menu costs. Such dependence

Although we assume that price adjustment periods T_{jkh} ($h = 1, 2, ...$) are chosen optimally, this is not strictly required for our approach: adjustment periods could instead be chosen according to some behavioral heuristics, as long as assumption 1 holds.

3.2 Product-Level Heterogeneity

Our setup allows pricing frictions to be product-specific. In particular, each product draws its frictions at the time of entry from a time-invariant item-specific set of frictions. In addition, it allows the fundamental stochastic processes for productivity (G_{jkt}), quality (Q_{jkt}) and flexible-price mark-ups (μ^f_{jkt}) to be product-specific. Furthermore, our setup imposes no restrictions on the joint distribution of pricing frictions and stochastic processes *across* items. However, a minimal set of assumptions on the *within-item* distributions of these objects is required.

To state these assumptions, we let ψ^d_{jkt} and ϕ^d_{jkt} denote the product-level expenditure and labor input shares in the deterministic flexible-price equilibrium, analogously to equations (6) and (8).[13] We then assume:

Assumption 2 *Within each expenditure item k the following holds in the cross-section of products j:*

2.1 The shares ψ^d_{jkt} and ϕ^d_{jkt} are mean-independent of the desired inflation rates π^*_{jk}, the processes driving productivity and mark-up shocks, and pricing frictions.

2.2 The desired inflation rates π^*_{jk} are independent of the processes driving productivity and mark-up shocks, and of pricing frictions.

Assumption 2.1 ensures that products with different fundamentals contribute equally (in expected terms) to item-level outcomes at the point of approximation because their average expenditure and labor input shares are independent of these fundamentals.[14] This assumption makes aggregation in the model consistent with our empirical approach that weighs products equally within an item. Assumption 2.2 rules out that desired inflation rates vary systematically with other product fundamentals within an item. Failure of this assumption would make it impossible to empirically recover how deviations of inflation from its desired value causally affect outcomes, because other fundamentals remain unobserved and thus cannot be controlled for in our regressions.[15]

---

of adjustment times on aggregate inflation makes it impossible to infer from the data observed in a given inflation regime how the economy would behave under a counterfactual inflation regime in which menu costs effectively differ.

[13]The deterministic flexible-price economy is the point around which we will approximate the economy.

[14]In contrast, in the stochastic sticky-price economy, average expenditure and labor input shares depend on the desired inflation rates, pricing frictions, and shocks.

[15]Note that we do not rule out that pricing frictions co-vary with the shock processes.

For the simple case in which *all* products *within an item* are subject to the same pricing friction and the same shock processes (allowing for idiosyncratic realizations), assumption 2 only requires that ψ^d_{jkt} and ϕ^d_{jkt} are mean-independent of desired inflation π^*_{jk} to preserve consistency with equal product weighting.

To facilitate the derivation of a closed-form expression for aggregate productivity losses due to misallocation, we adopt the following assumption:

Assumption 3 *The firm discount factor approaches one, i.e.,*

$$\beta\gamma^{1-\sigma}(1-\delta_k)\overline{G}_{jk}^{1-\theta}(\overline{\mu}^f_{jk})^{-\theta} \to 1 \qquad \textit{for all } j,k \tag{22}$$

The firm discount factor combines the household discount factor $\beta\gamma^{1-\sigma}$, discounting due to product exit risk $(1-\delta_k)$, and discounting due to the product-specific trends $\overline{G}_{jk}^{1-\theta}(\overline{\mu}^f_{jk})^{-\theta}$. Importantly, assumption 3 does not impose any restriction on the flexible-price trends of products, as defined in equations (12) and (14).

More generally, appendix A.6 shows that our setup imposes only minimal restrictions on the within-item flexible-price dynamics. It requires (i) time-invariance of the distributions of trends and shock processes, and limits time-variation in the distribution of intercepts to ensure consistency with balanced growth; (ii) some correlation between trends and intercepts to preserve equal product weighting in expectation in the deterministic flexible-price economy (assumption 2.1); and (iii) independence of trends from shock processes (assumption 2.2).

We now turn to the characterization of our main object of interest – the endogenous productivity loss due to price stickiness.

4 The Misallocation Costs of Inflation

This section shows how inflation π affects the endogenous component of aggregate productivity Δ on the balanced growth path.

We denote by $\mathcal{L}(\pi)$ the aggregate productivity loss in the sticky-price economy, relative to the flexible-price economy, when inflation is equal to π:

$$\mathcal{L}(\pi) = \ln\Delta(\pi) - \ln\Delta^f. \tag{23}$$

While inflation does not affect endogenous productivity under flexible prices Δ^f, it does so when prices are sticky. This happens via the following channels:

1. Different aggregate inflation rates imply different relative price trends during periods in which nominal prices do not adjust.

2. The anticipation of the above effect generally affects reset price choices during price adjustment periods.

3. Different inflation rates might also change the timing of price adjustments.

The importance of these channels depends on the specifics of the underlying pricing frictions. Our approach captures all three channels, as long as pricing frictions satisfy assumption 1.

We proceed in three steps. In section 4.1, we derive the relationship between the aggregate productivity loss $\mathcal{L}(\pi)$ and moments of the *price gap* distribution, where the price gap is defined as the difference between the actual and the flexible price, i.e.,

$$\ln z_{jkt} \equiv \ln p_{jkt} - \ln p^*_{jkt}.$$

In section 4.2, we show how the relevant moments of the price gap distribution depend on inflation and derive a set of sufficient statistics characterizing this dependence. In section 4.3, we show how the sufficient statistics can be recovered from micro price data. Respective results are formally derived in appendices A.2 and A.3.

4.1 Productivity Loss and Price Gaps

We start by considering the productivity loss $\mathcal{L}_k(\pi_k)$ in expenditure item k, defined analogously to aggregate loss (23) as

$$\mathcal{L}_k(\pi_k) \equiv \ln \Delta_k(\pi_k) - \ln \Delta^f_k. \tag{24}$$

The loss in item-level productivity depends on the item-level inflation rate π_k, which moves one-to-one with aggregate inflation π, see equation (19). In the following, we approximate the loss in equation (24) around the deterministic flexible-price economy. The next proposition relates the item-level productivity loss to moments of the price gap distribution:

Proposition 1 *Suppose assumptions 1-3 hold. The productivity loss in item k is*

$$\mathcal{L}_k(\pi_k) = \frac{\theta}{2}\Big(VAR_k(\ln z_{jkt}|\pi_k) + 2COV_k(\ln z_{jkt}, \ln \varepsilon^\mu_{jkt}|\pi_k)\Big) + O(3), \tag{25}$$

where $VAR_k(\cdot|\pi_k)$ and $COV_k(\cdot|\pi_k)$ denote the cross-sectional variance and covariance when inflation is equal to π_k and O(3) is a third-order approximation residual.

The dependence of the item-level loss on the variance of price gaps in proposition 1 is standard and captures misallocation costs due to sticky-price deviations from flexible prices. It is also present in Woodford (2003) and Galí (2008). The covariance between price gaps $\ln z_{jkt}$ and mark-up disturbances $\ln \varepsilon^\mu_{jkt}$ in equation (25) is new: it implies that the productivity loss is lower if the price gap covaries negatively with the mark-up shock. This is so because mark-up shocks are inefficient. For example, if the price gap is positive when the mark-up shock is negative, then price stickiness counteracts this inefficiency, leading to a lower loss.[16]

Interestingly, equation (25) reveals that there is a conflict between the objective pursued by firms and the objective of a social planner. While firms only seek to minimize the gap between actual and profit-maximizing flexible prices, the social planner also cares about the covariance term and prefers (ceteris paribus) a negative covariance between price gaps and mark-up shocks.

The productivity loss in proposition 1 scales with the demand elasticity θ because a higher demand elasticity implies that price gaps of a given size lead to larger quantity misallocation.

To aggregate item-level losses to an economy-wide loss, we impose the following assumption on the deterministic flexible-price economy, which is the point around which we approximate the aggregate loss (23):

Assumption 4 *In the deterministic flexible-price economy, item-level labor input shares are equal to item-level expenditure shares ($\phi_k^d = \psi_k$).*

Assumption 4 ensures that item-level expenditure shares capture the relative importance of item-level losses for the aggregate productivity loss.[17] As can be seen from equation (18), the economic importance of an expenditure item at the point of approximation is then solely determined by its expenditure weight. We then obtain:

Proposition 2 *Suppose assumptions 1-4 hold. The aggregate productivity loss associated with inflation π is*

$$\mathcal{L}(\pi) = \sum_{k=1}^{K} \psi_k \mathcal{L}_k(\pi_k) + O(3), \quad \text{with } \pi_k = \pi + \ln(\gamma/\gamma_k), \tag{26}$$

where O(3) is a third-order approximation residual.

Propositions 1 and 2 derive theoretically interesting connections between aggregate productivity and the joint distribution of price gaps and mark-up shocks. However, this

[16]Expression (25) does not have covariance terms involving shocks to productivity because these shocks are efficient and their interaction with price gaps does not matter for productivity losses. However, these shocks still drive the flexible price and affect the distribution of price gaps.

[17]This is equivalent to assuming that items have the same item-level mark-ups in the deterministic flexible-price economy, see appendix A.1.4 for details.

distribution is generally not identified by micro price data.[18] In fact, the subsequent sections show that the *level* of the productivity loss associated with a given inflation rate can generally not be recovered from micro price data. Nevertheless, we show that it is possible to recover how aggregate productivity *changes* with aggregate inflation.

4.2 Sufficient Statistics for the Productivity Loss

We now derive sufficient statistics characterizing how the aggregate productivity loss $\mathcal{L}(\pi)$ varies with inflation.

We start by considering the item-level loss $\mathcal{L}_k(\pi_k)$. To compute how the cross-sectional moments in equation (25) depend on inflation π_k, we consider a subset of products $J \subseteq [0,1]$ with identical desired inflation π^*_{jk}, fundamental shock processes, and pricing frictions.[19] We then perform a Taylor expansion to approximate how these moments for this subset of products depend on the *inflation gap* $\pi_k - \pi^*_{Jk}$, where π^*_{Jk} denotes the common desired inflation rate in subset J. This delivers

$$\begin{aligned} &VAR_{Jk}(\ln z_{jkt}|\pi_k) + 2COV_{Jk}(\ln z_{jkt}, \ln \varepsilon^{\mu}_{jkt}|\pi_k) \\ &= a_{Jk} + b_{Jk}\left(\pi_k - \pi^*_{Jk}\right) + c_{Jk}\left(\pi_k - \pi^*_{Jk}\right)^2 + O(3), \end{aligned} \tag{27}$$

where (a_{Jk}, b_{Jk}, c_{Jk}) denote the Taylor expansion coefficients and $O(3)$ is a third-order residual. To approximate the loss for *all* products in the item, one needs to appropriately aggregate over all possible subsets J. This aggregation step is non-trivial, as one needs to consider the variance both within and across subsets, and also take into account the heterogeneity in Taylor expansion coefficients. Nevertheless, the following proposition shows that one can equivalently aggregate across all products using a common set of expansion coefficients:

Proposition 3 *Suppose assumptions 1-3 hold, then*

$$\begin{aligned} &VAR_k(\ln z_{jkt}|\pi_k) + 2COV_k(\ln z_{jkt}, \ln \varepsilon^{\mu}_{jkt}|\pi_k) \\ &= a_k + b_k \int_0^1 \left(\pi_k - \pi^*_{jk}\right) dj + c_k \int_0^1 \left(\pi_k - \pi^*_{jk}\right)^2 dj + O(3), \end{aligned} \tag{28}$$

where (a_k, b_k, c_k) denote average Taylor expansion coefficients and $O(3)$ is a third-order approximation residual.

[18] Alvarez, Lippi and Oskolkov (2022) show that one can infer the distribution of price gaps from the distribution of price changes in the special case where the flexible price follows a random walk, and products face identical pricing frictions and random walk innovation variances. The random-walk assumption is, however, strongly rejected in our data, see section B.2.

[19] This ensures that all products $j \in J$ have the same price gap distribution.

The result that the same expansion coefficients (a_k, b_k, c_k) can be used to aggregate *across all products* within an item is important when seeking to estimate these coefficients empirically, as we do later. This key result emerges because the subset-specific Taylor coefficients, which depend on pricing frictions and shock processes, are independent of desired inflation π^*_{jk} due to assumption 2.2.

Propositions 1 and 3 then jointly imply the following:

Lemma 1 *Suppose assumptions 1-3 hold. Up to the second order, the sufficient statistics for the item-level loss $\mathcal{L}_k(\pi_k)$ consist of the coefficients (a_k, b_k, c_k), the distribution of desired inflation rates $\{\pi^*_{jk}\}_{j\in[0,1]}$, and the demand elasticity parameter θ:*

$$\mathcal{L}_k(\pi_k) = \frac{\theta}{2}\left(a_k + b_k \int_0^1 \left(\pi_k - \pi^*_{jk}\right) dj + c_k \int_0^1 \left(\pi_k - \pi^*_{jk}\right)^2 dj\right) + O(3), \tag{29}$$

where $O(3)$ is a third-order approximation residual.

The general set of price-setting frictions we consider leaves the sign and magnitude of the coefficients (a_k, b_k, c_k) unrestricted. Nevertheless, we can interpret the role of these coefficients in determining the productivity loss. In particular, coefficient a_k captures productivity losses due to price stickiness if item-level inflation is equal to the desired inflation for all products within an item.[20] A negative value of a_k would imply that price stickiness increases endogenous productivity by counteracting inefficiencies stemming from mark-up shocks. The coefficient b_k, in turn, determines first-order productivity losses associated with deviating from products' desired inflation. A negative b_k, for example, might arise if the covariance term in equation (25) falls more strongly with item-level inflation π_k than the variance of price gaps increases. Finally, the coefficient c_k captures how the item-level productivity loss changes around its extremum.[21] A positive value of c_k implies that productivity loss is bounded from below and the ability to reduce misallocation by means of inflation targeting is limited.

For special cases, the coefficients in lemma 1 can be determined analytically:

Proposition 4 *Suppose products in item k face the same Calvo price adjustment friction and that productivity and mark-up shocks are i.i.d. normal with mean zero and variance σ^2_G and σ^2_μ, respectively. Consider the continuous-time limit and let $\lambda > 0$ denote the Calvo*

[20] This would of course require that there is no heterogeneity in desired inflation rates $(\pi^*_{jk} = \pi^*_k)$.

[21] This becomes apparent once equation (29) is written in vertex form, as is done in proposition 7 later on for the aggregate loss.

price adjustment intensity. The coefficients in equation (29) are then given by:

$$a_k = \sigma_G^2 - \sigma_\mu^2 \tag{30}$$

$$b_k = 0 \tag{31}$$

$$c_k = 1/\lambda^2. \tag{32}$$

The economic interpretation of the coefficients (30)-(32) is straightforward. Price stickiness prevents firms from flexibly adjusting their prices in response to shocks. Since it is efficient to respond to productivity shocks, their variance increases productivity losses under sticky prices and pushes the intercept a_k towards positive values. In contrast, responding to mark-up shocks is inefficient, and therefore their variance decreases productivity under flexible prices and pushes the intercept a_k towards negative territory.[22] Interestingly, the coefficient b_k is zero with Calvo frictions, indicating that positive and negative deviations from desired inflation produce symmetric productivity effects. Finally, the coefficient c_k is positive, indicating that productivity losses increase with the squared deviation of inflation from the desired rate. The coefficient c_k, however, decreases with the intensity of the price adjustment λ because less severe pricing frictions (higher λ) allow firms to track more closely the time trend in the nominal flexible price.[23]

Although the example in proposition 4 suggests that c_k is positive, this is not necessarily always the case. Models in which price adjustment frequency responds sufficiently strongly to the inflation gap $(\pi_k - \pi_{jk}^*)$ can have $c_k < 0$. Similarly, it is not necessarily true that b_k is always equal to zero. In a menu cost setup, the covariance between mark-up shocks and price gaps can respond to changes in the inflation gap to first order, unlike in a Calvo setting.

Our main theoretical result explains what is needed to recover the aggregate productivity loss associated with different inflation rates:

Proposition 5 *Suppose assumptions 1-4 hold, we observe an actual equilibrium with aggregate inflation* π^{act} *and item-level inflation rates* $\{\pi_k^{act}\}_{k=1}^K$ *and wish to determine the aggregate loss* $\mathcal{L}(\pi)$ *at an aggregate inflation rate* $\pi = \pi^{act} + d$*, with* $d \gtreqless 0$*. To second order, the sufficient statistics allowing one to do so are (i) the expenditure weights* $\{\psi_k\}_{k=1}^K$ *and (ii) the sufficient statistics for the item-level losses stated in lemma 1. The associated*

[22] The strength of these effects would generally depend on the severity of pricing frictions, but not so in the continuous-time limit, which we consider here for analytic tractability.

[23] This time trend is nonzero whenever $\pi_k - \pi_{jk}^* \neq 0$.

aggregate loss is

$$\mathcal{L}(\pi^{act}+d) =$$
$$\frac{\theta}{2}\sum_{k=1}^{K}\psi_k\left(a_k + b_k\int_0^1\left(\pi_k^{act}-\pi_{jk}^{*}+d\right)dj + c_k\int_0^1\left(\pi_k^{act}-\pi_{jk}^{*}+d\right)^2 dj\right)+O(3), \tag{33}$$

where $O(3)$ *is a third-order approximation residual.*

The proposition follows directly from proposition 2 and lemma 1 and uses the fact that the item-level inflation rates π_k move one-to-one with aggregate inflation π, see equation (19).

The proposition shows that once one recovers inflation gaps $\pi_k^{act}-\pi_{jk}^{*}$ in an actual equilibrium observed in the data, one can compute the losses in other equilibria with different aggregate inflation rates (provided one knows also the remaining sufficient statistics). Our assumptions ensure that we can infer the behavior of a specific product in an alternative equilibrium from the behavior of other products in the actual equilibrium. Specifically, the distribution of price gaps for a product i in an equilibrium with alternative inflation rate $\pi^{act}+d$ is identical to the distribution of price gaps of some product j in the actual equilibrium if $\pi_{jk}^{*}=\pi_{ik}^{*}-d$.[24] This dependence of product behavior on π_{jk}^{*} in the actual equilibrium is in turn captured by the coefficients (b_k, c_k).

4.3 Recovering the Sufficient Statistics from Micro Price Data

We now show how some of the sufficient statistics in proposition 5 can be recovered from micro price data.

The expenditure weights $\{\psi_k\}_{k=1}^{K}$ are often readily available from statistical offices, as is the case with the UK's Office for National Statistics. The elasticity of product demand θ, however, cannot be recovered from micro price data alone.[25] Since the parameter θ scales productivity losses linearly, readers can easily adjust the losses we report later for alternative preferred values.

The following proposition shows how the parameters (b_k, c_k) and the distribution of desired inflation rates $\{\pi_{jk}^{*}\}_{j\in[0,1]}$ can be recovered from micro price data of item k:

Proposition 6 *Suppose assumptions 1-3 hold and we observe an actual equilibrium with aggregate inflation* π^{act} *and item-level inflation rates* $\{\pi_k^{act}\}_{k=1}^{K}$. *In population, regressing*

[24]For illustrative purposes this assumes that products are identical in terms of pricing frictions and shock processes, which is unproblematic, because these characteristics are independent of desired inflation, due to assumption 2.2.

[25]This would require observing also quantities sold at different prices.

the quality-adjusted relative price $\ln p_{jkt}$ *on a constant and a time trend identifies desired inflation* π^*_{jk}*:*

$$\ln p_{jkt} = \alpha_{jk} - t \cdot \pi^*_{jk} + u_{jkt}, \tag{34}$$

The following population regression across products j *in item* k *then identifies the coefficients* b_k *and* c_k*:*

$$VAR_{jk}\left(u_{jkt}\right) = \widetilde{a}_k + b_k(\pi^{act}_k - \pi^*_{jk}) + c_k(\pi^{act}_k - \pi^*_{jk})^2 + v_{jk}, \tag{35}$$

where

$$\widetilde{a}_k \equiv a_k + VAR_k(\ln \varepsilon^{\mu}_{jkt}).$$

Equation (34) reveals that the desired inflation rate π^*_{jk} is identified by the time trend of a product's relative price, which allows one to estimate the distribution of the desired inflation rates $\{\pi^*_{jk}\}_{j\in[0,1]}$. The intuition underlying this finding is straightforward: the time trend of the flexible relative price is given by $-\pi^*_{jk}$, see equation (12). Since the sticky price tracks the flexible price over time, it also inherits the flexible-price trend.

Perhaps more surprising is the result in equation (35): the residual variance of the detrended relative price, $VAR_{jk}\left(u_{jkt}\right)$, moves with the inflation gap in the cross-section of products in the very same way as the item-level productivity loss in lemma 1. This identifies the coefficients (b_k, c_k). Clearly, this requires that the desired inflation rates π^*_{jk} differ between products. We show in appendix B.1 that this is satisfied in our data.

Importantly, the coefficients a_k are generally not identified by regression (35). The estimated intercept reflects the combined effects of mark-up shocks ($VAR_k(\ln \varepsilon^{\mu}_{jkt})$) and the loss-relevant intercept a_k. Thus, the coefficient a_k is only identified in the special case in which mark-up shocks are absent. Therefore the *level* of productivity loss relative to a setting with flexible prices can generally not be recovered from the data. However, one can identify how productivity losses *change* with inflation via the coefficients (b_k, c_k), which will be the focus of the remainder of the paper.

In practice, we will run the first-stage regression (34) for every product in our sample and use the resulting estimates $\widehat{VAR}_{jk}\left(u_{jkt}\right)$ and $\widehat{\pi}^*_{jk}$ from these regressions to run the second-stage regression (35) to obtain coefficient estimates $(\widehat{b}_k, \widehat{c}_k)$. Section 6 will report the misallocation costs implied by these estimates, with details of the econometric approach being discussed in appendix C.

5 The UK Micro Price Data

We use the micro price data underlying the official U.K. consumer price index (CPI) as provided by the Office for National Statistics (ONS). We consider roughly 20 years of micro

Total number of price quotes used:	13.3 million			
Number of products:	234,047			
Number of expenditure items:	818			
	mean	median	min	max
Number of products per item	286	236	50	1098
Number of price quotes per item	16,219	13,346	1702	59,756
Number of price quotes per product	60	49	30	251
Number of price changes per product	7.4	6	0	183
Length of price spells (months)	8.0	6	1	119

Table 1: Basic product and price statistics

price data (February 1996 to December 2016) from a period in which aggregate inflation was rather stable, in line with our theoretical setup. The data are monthly and classified into narrowly defined expenditure items (e.g., flat panel TV 33inch, men's shoes trainers, vegetarian main course, etc.). Given the sample selection described below, we consider 818 different expenditure items and more than 13 million price observations over the sample period.

Starting from the raw micro price data, we delete products with duplicate price observations in a given month and also delete all price observations flagged by ONS as "invalid".[26] We then apply the ONS product definition and consider a product within an item as a sequence of price observations for a particular physical object or service sold in a particular location.[27] We split price time series into different products whenever a 'non-comparable' product substitution is recorded or whenever the product price is missing for more than five consecutive months, to avoid lumping together products that are potentially different.

We consider only expenditure items for which the item price index, computed from our micro price data, replicates the official item price index provided by ONS sufficiently well. Furthermore, we exclude cigarette items because price dynamics are largely the result of tax changes. We only consider products with a minimum length of 30 price observations. As a baseline, we use the Nakamura and Steinsson (2008) sales filter A, which removes V-shaped reductions in the sales price.[28] Within each expenditure item, we account for outliers by eliminating the 5% of products with the highest estimated values for $VAR_{jk}(u_{jkt})$ and

[26]Duplicate price quotes can arise because ONS does not disclose all available locational information underlying the data, so that in rare cases we cannot uniquely identify the product price.

[27]Following ONS practice and in line with the approach in Nakamura and Steinsson (2008), we continue price time series over so-called 'comparable' product substitution events.

[28]Section 8.2 shows that results are robust to using alternative sales filters.

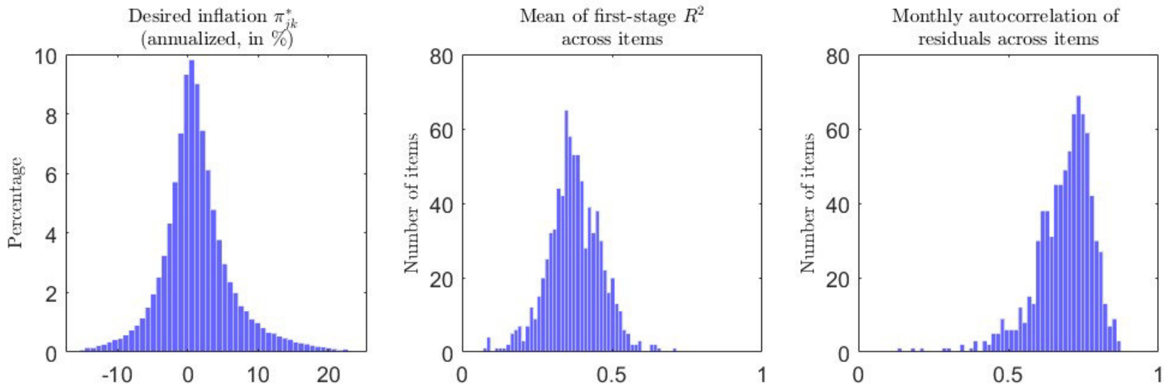

Figure 1: Outcomes from regression (34).

$(\pi^{act}_k - \pi^*_{jk})^2$ and consider only expenditure items containing at least 50 products.[29] This leaves us with a baseline sample containing 818 expenditure items that we use throughout our empirical analysis.

Table 1 provides basic descriptive statistics. On average, we have several hundred products and several thousand price observations per expenditure item. The average number of price observations per product is 60 and the average number of price changes per product is around 7. The table also reports information on the lengths of price spells.[30] On average, nominal prices stay constant for 8 months, with the median length being 6 months.

We then run regression (34) for every product in our sample using quality-adjusted prices. The left panel in figure 1 depicts the distribution of the estimated desired inflation rates π^*_{jk} across all products: 61% of products have a positive desired inflation rate, with some rates being relatively large. This reflects the fact that the relative price of products often tends to decline over the product life, as previously documented for U.S. durable goods in Bils (2009). In our data, certain expenditure categories, e.g., clothing and electronics, display on average relatively high rates of relative price decline, see Adam and Weber (2023) for further details.

The center panel in figure 1 reports the distribution of average R^2 values across items. For most items the time trend explains between 30-50% of the variation in the relative price over time. The right panel in figure 1 depicts the across-item distribution of the

[29] We also eliminate expenditure items for which the estimated residual variances are zero for all products. The latter occurs when prices never adjust within an item, which is the case for less than a handful of items capturing administered prices.

[30] These numbers are computed using uncensored price spells only, i.e., by ignoring the first and last price spell of every product, then computing the mean spell length for every product and finally reporting statistics for the pooled distribution of all products.

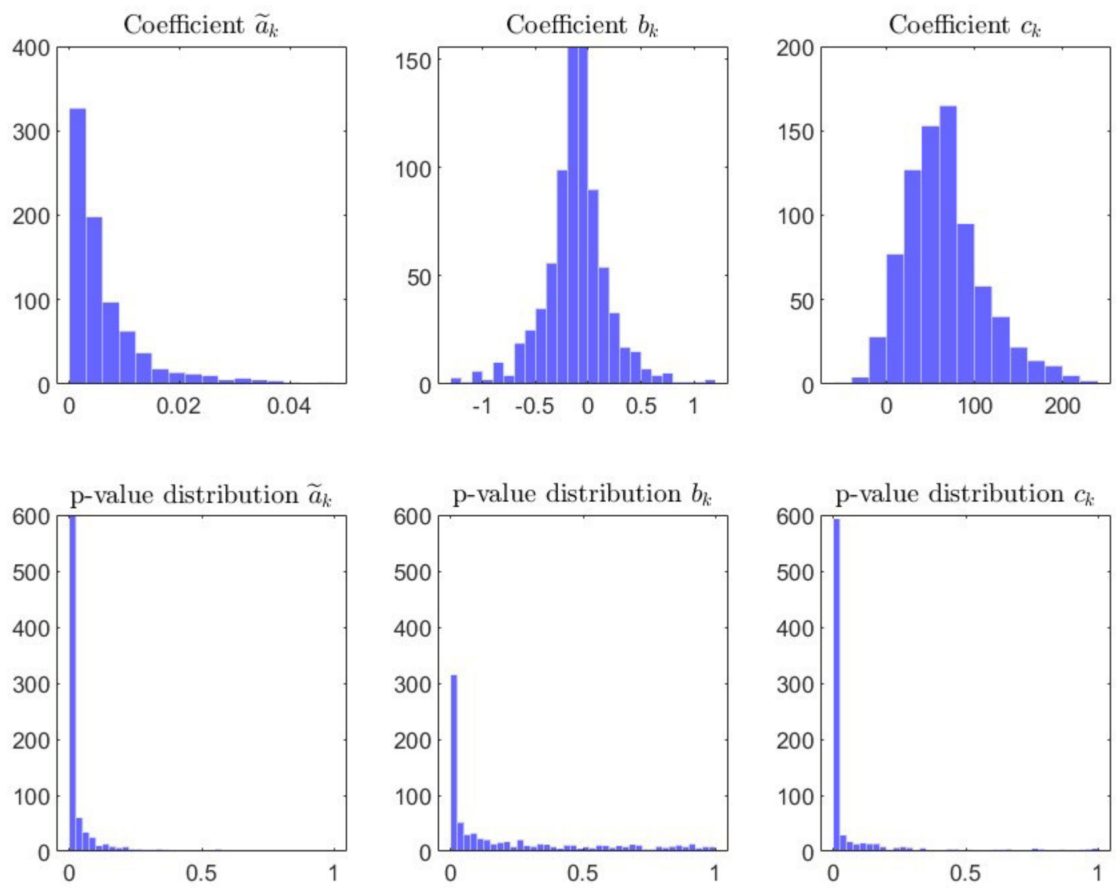

Figure 2: Outcomes from regression (35)

average *monthly* autocorrelation of residuals and shows that these are well bounded below one.

Our empirical strategy requires panel data on quality-adjusted prices, with sufficiently long product price time series and sufficient cross-sectional heterogeneity in desired inflation (π^*_{jk}) to estimate the relevant parameters in proposition 6. Therefore, we formally test in appendix B the null hypothesis of identical desired inflation rates within items. This hypothesis is rejected for virtually all items at the 1% significance level. However, we are not the first to show heterogeneity in price trends within narrowly defined product categories. This has previously been documented using scanner data in the literature studying inflation heterogeneity for households across the income distribution (Kaplan and Schulhofer-Wohl (2017) and Jaravel (2018)).[31]

We also reject in appendix B the hypothesis that the flexible price follows a random

[31] Jaravel (2018) explains this heterogeneity by differences in mark-up dynamics for products catering to high- and low-income households, driven by different levels of competition. Our setup explicitly allows for these different mark-up trends.

walk (potentially with drift), as is sometimes assumed for analytical tractability, e.g., Alvarez, Lippi and Oskolkov (2022). This appendix also shows that, within expenditure items, there is no statistically significant relation between relative price trends and the number of periods that a product stays in the sample.

We then run regression (35) to estimate for each expenditure item k the coefficients $(\tilde{a}_k, b_k, c_k)$. The top panels in figure 2 depict the distributions of estimated coefficients across items and the bottom panels the corresponding distributions of p-values. Bins have a width of 2.5%. The intercepts $\tilde{a}_k$ are positive and highly statistically significant. The coefficients b_k are roughly centered around zero, but more often negative than positive and also often statistically significant. The most important finding is that the coefficients c_k on the quadratic term in equation (35) are overwhelmingly positive and statistically highly significant.[32] We quantify the productivity losses implied by these estimates in the next section.

6 Main Empirical Findings

To quantify the productivity losses $\mathcal{L}(\pi)$ associated with different aggregate inflation π, we first derive a convenient alternative representation of the aggregate loss in proposition 5:[33]

Proposition 7 *Suppose assumptions 1-4 hold and we observe an actual equilibrium with aggregate inflation π^{act} and item-level inflation rates $\{\pi_k^{act}\}_{k=1}^K$. The aggregate productivity loss with aggregate inflation rate $\pi = \pi^{act} + d$, with $d \gtrless 0$, is then given by*

$$\mathcal{L}(\pi) = \overline{L} + \frac{\theta}{2}\left(\sum_{k=1}^{K} \psi_k c_k\right)(\pi - \pi^*)^2 + O(3) \tag{36}$$

where

$$\pi^* \equiv \pi^{act} - \frac{\sum_{k=1}^K \psi_k c_k \int (\pi_k^{act} - \pi_{jk}^*)\, dj}{\sum_{k=1}^K \psi_k c_k} - \frac{1}{2}\frac{\sum_{k=1}^K \psi_k b_k}{\sum_{k=1}^K \psi_k c_k} \tag{37}$$

$$\overline{L} \equiv \frac{\theta}{2}\sum_{k=1}^{K} \psi_k \left(a_k + b_k \int_0^1 \tilde{\pi}_{jk}\, dj + c_k \left(\int_0^1 (\tilde{\pi}_{jk})^2\, dj - (\pi^*)^2\right)\right), \tag{38}$$

with $\tilde{\pi}_{jk} \equiv \pi_k^{act} - \pi_{jk}^* - \pi^{act}$.

Equation (36) shows that the losses associated with deviating from the inflation rate π^* are solely a function of the expenditure shares ψ_k and the coefficients c_k. These losses are positive, provided $\sum_{k=1}^K \psi_k c_k > 0$, as is the case in our data (top right panel in figure 2).

[32] The figure shows the estimated coefficients when expressing inflation in monthly rates.

[33] See appendix A.4.4 for the proof of proposition 7.

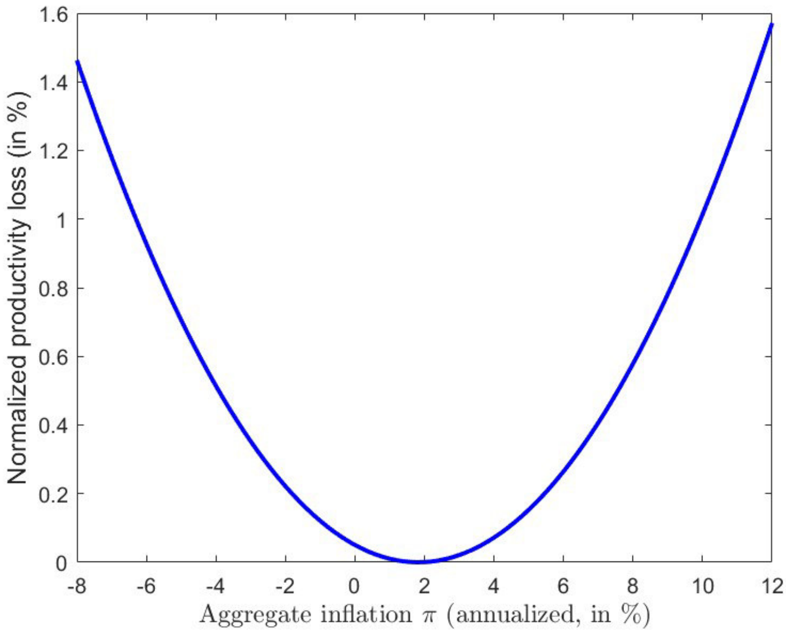

Figure 3: The normalized productivity loss $\mathcal{L}^N(\pi)$

The inflation rate π^* then minimizes the aggregate loss and depends on various sufficient statistics, but is *independent* of the parameters a_k, which cannot generally be recovered from the data (proposition 6).[34] The constant $\overline{L}$, however, depends on the parameters a_k and can thus be recovered from the data only in the special case without mark-up shocks.

Given the previous proposition, we shall consider the normalized loss function, which omits the constant $\overline{L}$:

$$\mathcal{L}^N(\pi) \equiv \frac{\theta}{2}\left(\sum_{k=1}^{K}\psi_k c_k\right)(\pi - \pi^*)^2 + O(3) \tag{39}$$

Figure 3 depicts the normalized loss $\mathcal{L}^N(\pi)$ when setting $\theta = 7$, in line with Golosov and Lucas (2007), Coibion, Gorodnichenko and Wieland (2012) and Nakamura et al. (2018). To construct this figure, we vary the aggregate annual inflation rate by up to ± 10 percentage points around the 2% average inflation rate in our data, and compute the corresponding productivity losses using our sufficient statistics. The productivity loss is minimized at an inflation rate of 1.8% (in annualized terms), which is in line with estimates in Adam and Weber (2023) who considered a setup allowing for significantly less heterogeneity.

Figure 3 shows that productivity losses remain quantitatively small for inflation rates that are within ± 1 p.p. of the optimal rate. However, productivity losses become quantitatively large when moving further away from the optimal rate: permanently increasing

[34]The optimal inflation rate π^* is also independent of the actual inflation rates observed in equilibrium: equation (37) implies that moving π^{act} and the $\{\pi_k^{act}\}_{k=1}^K$ by the same amount leaves π^* unchanged.

	Inflation expressed in		
	monthly rates	quarterly rates	annualized rates
π^*	0.15%	0.45%	1.8%
$\sum_{k=1}^{K} \psi_k c_k$	62.4	6.93	0.433

Table 2: Key statistics of the normalized loss function $\mathcal{L}^N(\pi)$

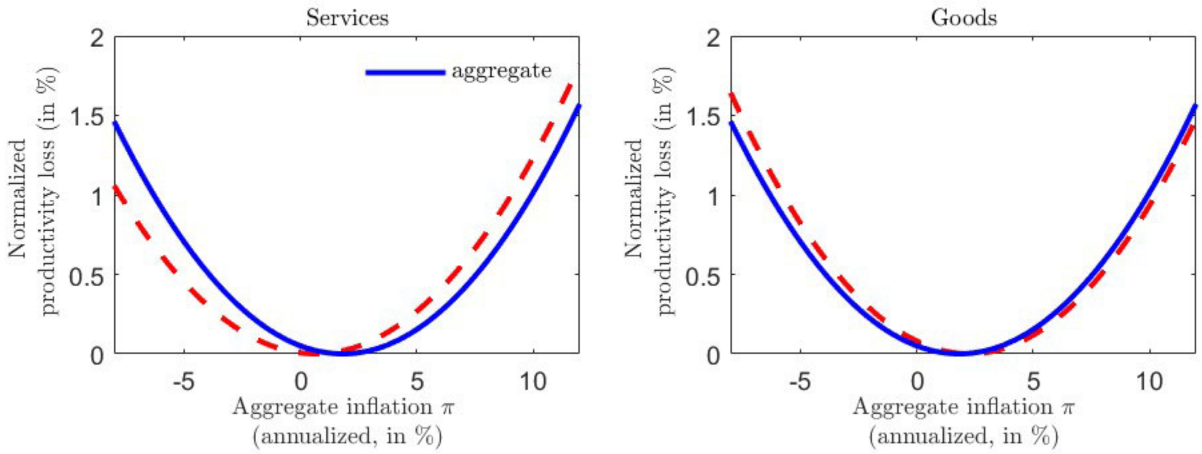

Figure 4: Productivity losses in services and goods vs. aggregate losses

inflation to +10%, which is approximately the level U.K. inflation reached in October 2022, leads to a productivity loss of 1.0% per period; conversely, moving inflation down to −6% leads to a loss of 0.92% per period. The estimated losses are strikingly different from the ones computed in Nakamura et al. (2018), who find virtually zero productivity losses due to inflation. One possible reason for this discrepancy is that their model calibration targets data moments that are not part of the sufficient statistics determining the productivity losses associated with different inflation rates. We investigate this further in section 7.

Table 2 reports the relevant sufficient statistics characterizing the normalized loss $\mathcal{L}^N(\pi)$, when inflation is expressed in monthly, quarterly or annual rates. A sticky-price model that seeks to capture the misallocation costs associated with deviations from the optimal inflation rate should use the estimated value of $\sum_k \psi_k c_k$ reported in this table as data target for the U.K. economy.

6.1 Subcategories of the Expenditure Basket

We now show that the productivity losses associated with deviations from the optimal inflation rate occur in a broad set of expenditure categories and are not concentrated in a few categories only.

The left panel in figure 4 compares the normalized productivity losses in the services

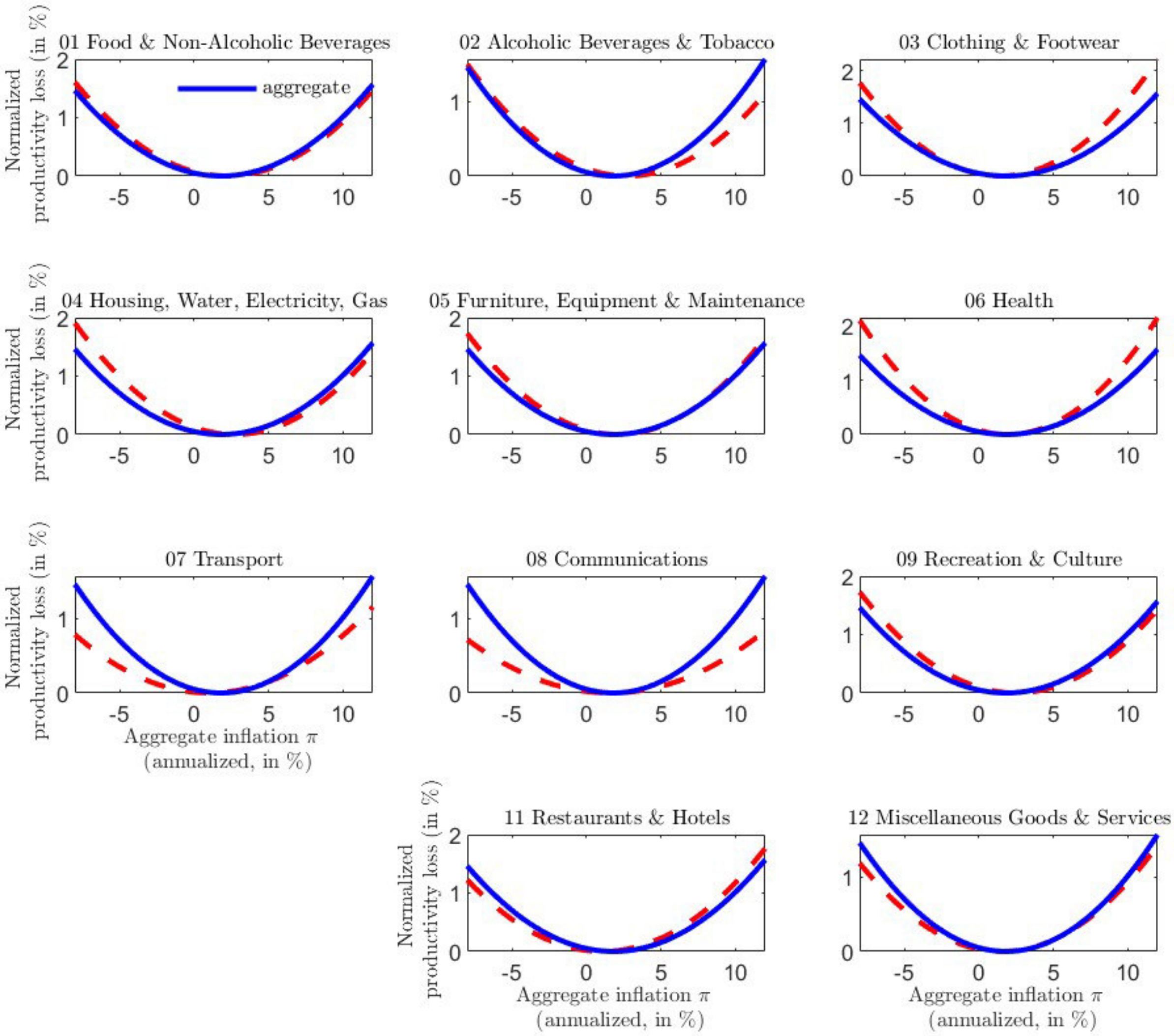

Figure 5: Productivity losses, ONS expenditure division vs. aggregate economy

category (dashed line) against the normalized aggregate losses (solid line).[35] The optimal aggregate inflation rate for services is approximately 0.6% and thus considerably lower than the optimal inflation rate for the aggregate economy (1.8%) because the desired inflation rates (π^*_{jk}) in services are on average lower than in the economy overall. However, once the difference in the level of the optimal inflation rate is taken into account, deviations from the optimum lead to productivity losses in services that are very similar to those of the aggregate economy.

The right panel of figure 4 shows the productivity losses in the goods category. The optimal inflation rate for goods is now higher (2.4%), but the losses associated with deviating from the optimum also align well with those for the aggregate economy.

These findings are quite surprising, especially because the average price adjustment frequency for goods (16.6% per month) is much higher than that for services (7.9% per month). This is often interpreted as goods prices being more flexible than service prices, which would imply lower misallocation costs of inflation with Calvo frictions, see proposition 4. However, this is inconsistent with the result that the misallocation costs of suboptimal inflation are very similar across the two subcategories.

Figure 5 repeats the previous exercise for a finer level of disaggregation using the so-called ONS expenditure divisions.[36] As with goods and services, most expenditure divisions display only variation in the optimal inflation rate, but exhibit similar losses for deviations from the optimum. Exceptions are 'Clothing and Footwear' and 'Health', where productivity losses associated with deviations from the optimum are larger, and 'Transport' and 'Communication', where these losses are smaller .

6.2 The Productivity Loss at the Optimal Inflation Rate

One additional object of interest is the productivity loss due to price stickiness. It can arise even in a situation when aggregate inflation is at its optimal level ($\pi = \pi^*$) and is captured by the constant $\overline{L}$ in proposition 7. We provide an upper bound for this loss and estimates for some of its components.

It follows from proposition 6 that the loss $\overline{L}$ is only identified in the special case without mark-up shocks, and figure 6 depicts the (non-normalized) loss $\mathcal{L}(\pi)$ in this case.[37] Misallocation costs at the optimal inflation rate ($\pi^* = 1.8\%$) amount to a staggering 2.3% of aggregate productivity, under the assumption that mark-up shocks are absent. Since

[35] We divide losses by the expenditure share of the subcategory, so that differences in expenditure weights do not affect our comparisons between categories.

[36] There are twelve expenditure divisions, but our sample does not cover the 10th division ('Education').

[37] We again assume $\theta = 7$. Note that the non-normalized loss is proportional to θ.

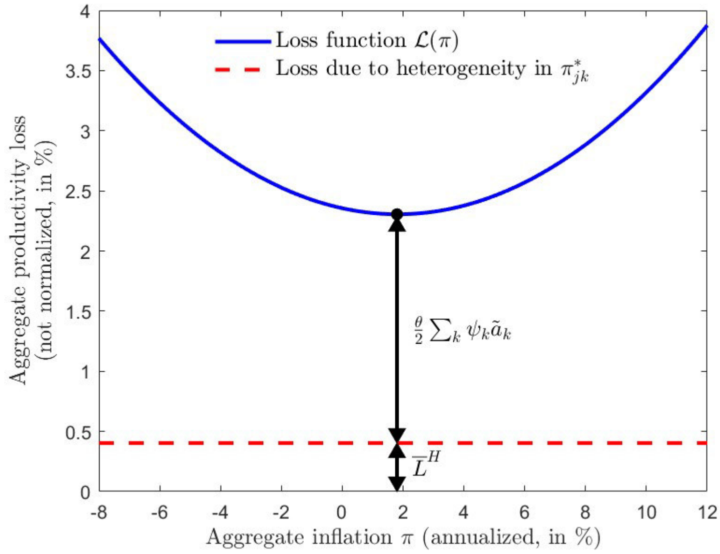

Figure 6: The level of the aggregate productivity loss $\mathcal{L}(\pi)$

these shocks are likely present in the data, actual losses are presumably lower.[38] Therefore, our approach identifies only an upper bound for the productivity loss $\overline{L} \leq 2.3\%$.

However, a component of the productivity loss $\overline{L}$ can be identified, independently of whether mark-up shocks are present: these are the misallocation costs arising from the fact that desired inflation rates π^*_{jk} are heterogeneous across products. This heterogeneity causes inflation to be suboptimal for nearly all products, even when aggregate inflation is at its optimal level ($\pi = \pi^*$).

From proposition 7 it follows that the losses associated with this heterogeneity are given by $\overline{L}^H \equiv \overline{L} - \frac{\theta}{2}\sum_{k=1}^{K} \psi_k a_k$, which no longer depends on the non-identifiable a_k. The dashed line in figure 6 depicts this estimate and shows that the misallocation costs $\overline{L}^H$ arising from heterogeneity in desired inflation rates amount to 0.4% of aggregate productivity.

7 The Importance of Targeting the Sufficient Statistics

We now calibrate a multi-sector structural model to our data, following a standard calibration strategy. This allows us to test (i) the ability of our sufficient statistics approach to accurately recover the normalized productivity loss $\mathcal{L}^N(\pi)$ generated by the model (without any approximation), and (ii) whether standard calibration targets imply productivity losses that are consistent with our empirical estimates. We find that our approach accurately recovers the productivity losses implied by the model, but the standard calibration strategy fails to generate the misallocation costs that we estimate from the data.

[38]Productivity losses could even be negative, see the discussion following proposition 4.

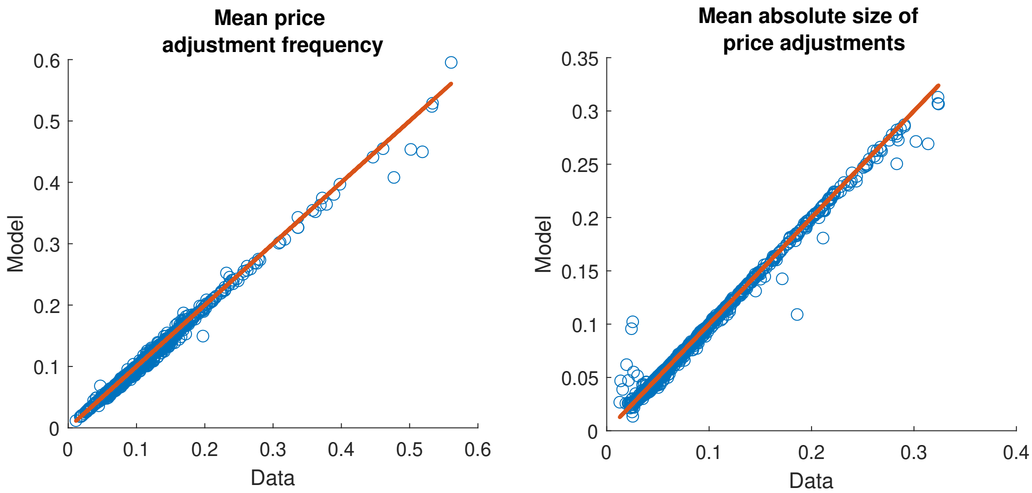

Figure 7: Targeted versus simulated moments across expenditure items

To illustrate these points, we calibrate item-specific Calvo-Plus pricing frictions for each of our more than 800 expenditure items. Under these frictions, firms can adjust their prices at any time by paying a fixed menu cost, or they can wait until the stochastic arrival of a costless adjustment opportunity. This provides a flexible framework allowing for both time- and state-dependent price adjustments. As in our theoretical framework in section 2, firms maximize profits given a CES demand function and their flexible price is affected by stationary idiosyncratic shocks.

We fix the distribution of relative price trends $\{\pi^*_{jk}\}$ in each expenditure item k to its empirical counterpart. Following the calibration approach of Nakamura and Steinsson (2010), we set the arrival rate of free adjustment opportunities equal to the rate of comparable product substitutions, and calibrate the menu cost parameter and the standard deviation of shocks to the flexible price to replicate the average absolute size of costly price adjustments and the average frequency of costly price adjustments in each expenditure item.[39] Figure 7 depicts the actual and simulated data moments and the 45 degree line (in red). The Calvo-Plus model successfully replicates the data moments for the vast majority of expenditure items.

We then simulate the calibrated model *preserving the small sample properties* of the price time series coming from our data. In particular, we replicate the empirical item-specific product length distribution, as observed in the data. Appendix D shows that our estimation approach successfully recovers the coefficients $\{b_k, c_k\}_{k=1}^K$ and the desired inflation rates $\{\pi^*_{jk}\}$. Figure 8 depicts the true normalized productivity loss $\mathcal{L}^N(\pi)$ implied by the calibrated model (solid blue line), without any approximations, and the normalized

[39]For tractability reasons, we only consider mark-up shocks. See appendix D for a detailed exposition.

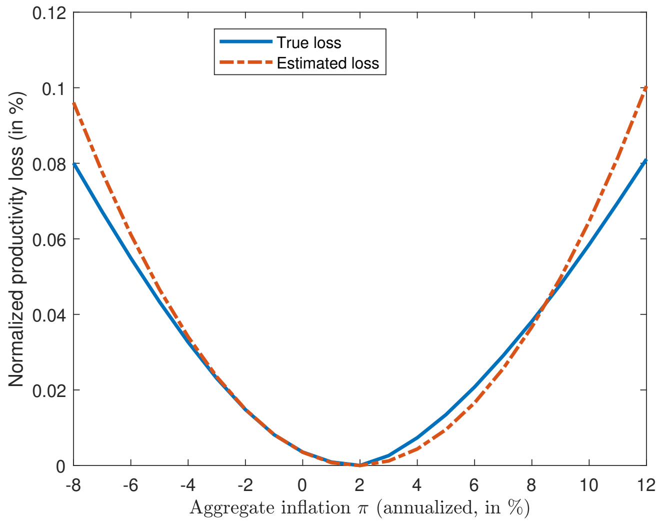

Figure 8: True model-implied vs. estimated productivity loss

productivity loss implied by our estimated statistics (dashed red line).

Figure 8 shows two important results. First, the two lines in the figure are closely aligned, which illustrates that our sufficient statistics approach performs rather well in recovering the losses implied by the calibrated sticky-price model. Second, productivity losses generated by the model are *an order of magnitude smaller* than those estimated from the actual data: the loss from a 10% inflation rate is now only 0.06%, while it is 1% in the data. This implies that the price time series generated by the calibrated Calvo-Plus model respond very differently to inflation than the price time series in the data.

The second finding is not due to a lack of flexibility of the price-setting model itself, but due to the data moments that are used as calibration targets. A standard calibration approach targets various average (or median) moments, such as average price adjustment frequency or the average absolute size of price adjustments, as in our exercise. However, the productivity losses in equation (36) do not depend on average moments of the price adjustment distribution, but on the coefficients $\{c_k\}_{k=1}^K$ that capture *how the misallocation costs react* to squared deviations of inflation from its desired value. We next show that in our data the standard calibration targets are quite uninformative about these coefficients.

Figure 9 depicts scatter plots of the frequency of price adjustments (left panel) and the mean absolute size of price adjustments (right panel) against the coefficient c_k (on the y-axis in both panels) across expenditure items. The point clouds make it clear that the standard data moments are not very informative about the coefficients c_k in the cross-section of items. In fact, linearly regressing c_k on the item-level price adjustment frequency and the item-level mean absolute size of price changes delivers an R^2 value of only 0.086.

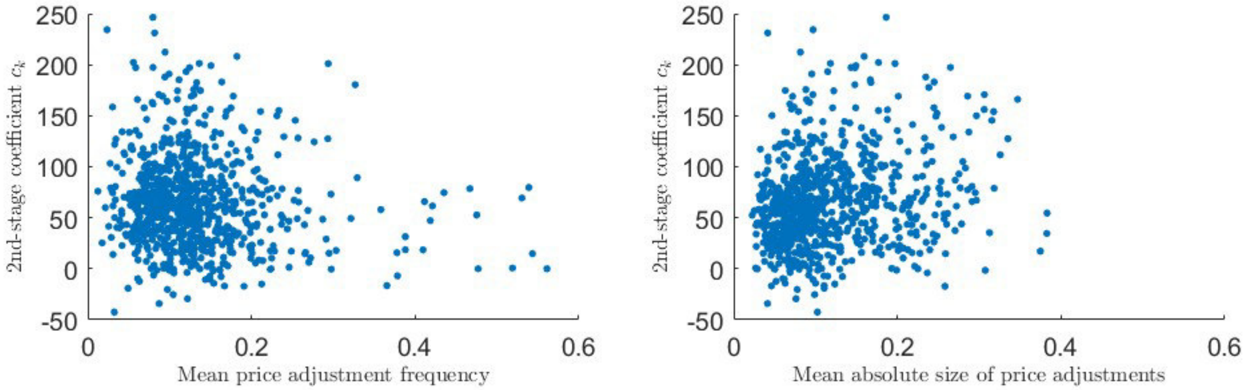

Figure 9: Standard calibration targets (x-axis) versus coefficient c_k (y-axis)

Adding the squares of the regressors and the interaction term increases the R^2 value to 0.096 only.

The logic underlying this finding is fairly simple: the standard calibration targets are *average* moments, whereas the coefficient c_k captures the *sensitivity* of misallocation costs with respect to inflation around the optimum (to second order). Except for the special case of a pure Calvo model (see proposition 4), there are no general results linking misallocation costs to standard calibration targets; and as the left panel in figure 9 makes clear, the predictions of the Calvo model are not born out in the data.

8 Robustness of the Empirical Findings

We now explore the robustness of our main quantitative finding in section 6 along a number of dimensions. The next two sections show that productivity losses are indeed symmetric and that our quantitative results do not depend on the precise treatment of sales prices. The following two sections address concerns about possible breaks in the time trend of relative prices and the potential effects of measurement error.

8.1 Are Productivity Losses Asymmetric?

Our baseline approach relies on proposition 3, which approximates the key moments of price gap distribution using a second-order Taylor expansion in the gap between actual and desired inflation.[40] This imposes from the outset that productivity losses are symmetric around the optimal inflation rate, see proposition 7.

We thus explore whether the symmetry of losses is robust to allowing for a cubic expansion in the gap between inflation and desired inflation. Specifically, we add the third-order Taylor term $(\pi_k - \pi^*_{jk})^3$ to the second-stage regression (35) and adjust the loss

[40] This is motivated by the fact that productivity in proposition 1 is itself approximated to second order in terms of the size of shocks and price gaps.

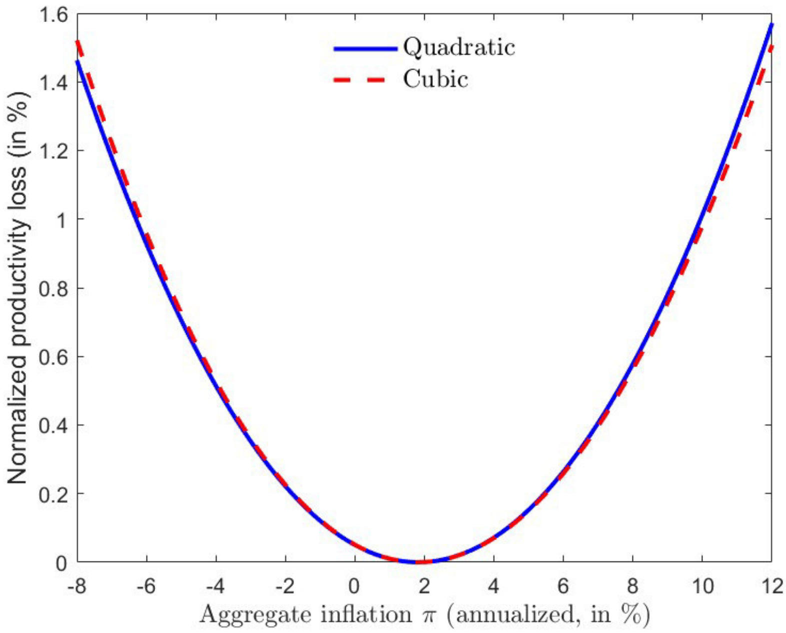

Figure 10: Productivity loss, quadratic versus cubic polynomial

function $\mathcal{L}(\pi)$ accordingly.[41] We then again normalize losses to zero at their minimum.

Figure 10 displays the resulting productivity loss and compares it with our baseline result: the losses associated with deviations from the optimal inflation rate π^* are indeed highly symmetric.

8.2 Do Results Depend on the Treatment of Sales Prices?

The treatment of sales prices can affect the way in which residual variation around the relative price trend of products varies with inflation. To test whether the choice of sales filter has an effect on our quantitative findings, we re-estimate the normalized productivity loss $\mathcal{L}^N(\pi)$ using a battery of different sales filters.

Our baseline results rely on the Nakamura-Steinsson filter A (NSA), but we now also consider the Nakamura-Steinsson filter B (NSB), the regular price (REG) of Kehoe and Midrigan (2015), and the regular price filter (RGF) in line with the approach of Kryvstov and Vincent (2021). Figure 11 shows that alternative sales filters leave estimated normalized productivity losses virtually unaffected.

[41] The overall approximation in equation (25) still remains second-order in the distribution of price gaps and shocks.

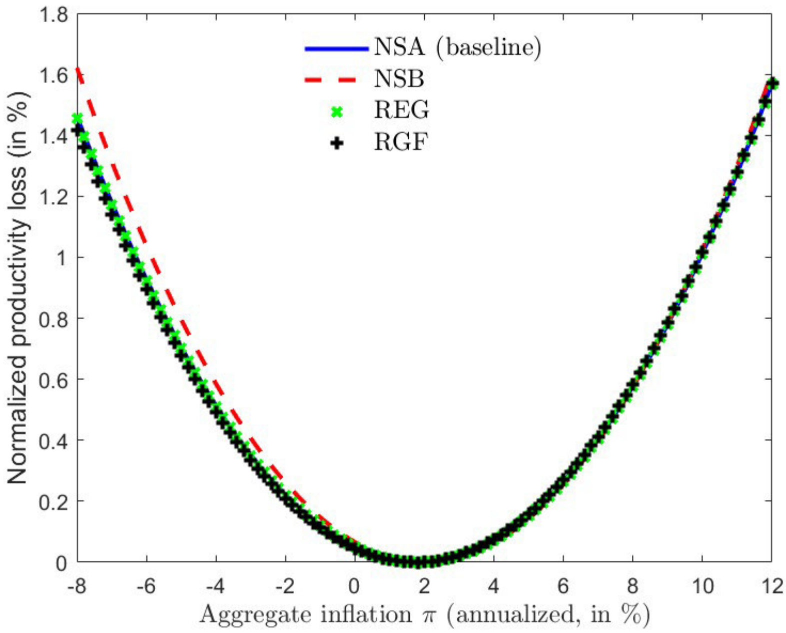

Figure 11: Normalized productivity loss $\mathcal{L}^N(\pi)$ for different sales filters

8.3 Possible Breaks in Relative Price Trends

One potential concern with our baseline estimation is that the time trend in relative prices may not be constant over time. In this case, the first-stage regression (34) would be mis-specified, leading to incorrect estimates of the relative price trend and the residual variance.

To address this concern in a conservative way, we use the Bai-Perron (1998) test statistic to determine the most likely break point in *each* product's price time series.[42] We then split *every* price time series at its most likely break point, independently of the value of the test statistic. The left panel of figure 12 displays the normalized loss $\mathcal{L}^N(\pi)$ when splitting all price time series as described and compares it to the loss under the baseline approach.[43] The estimated productivity losses drop only slightly: the productivity losses associated with a 10% inflation rate are then equal to 0.78% instead of 1.0%.[44] The effect of possible trend breaks on our results thus appears to be relatively muted.

[42] We search in the two inner quartiles of the price time series.

[43] As in the baseline approach, we use all price time series with at least 30 price observations.

[44] Part of this effect is likely due to the fact that price time series are now shorter than in the baseline sample; see the discussion in the next section.

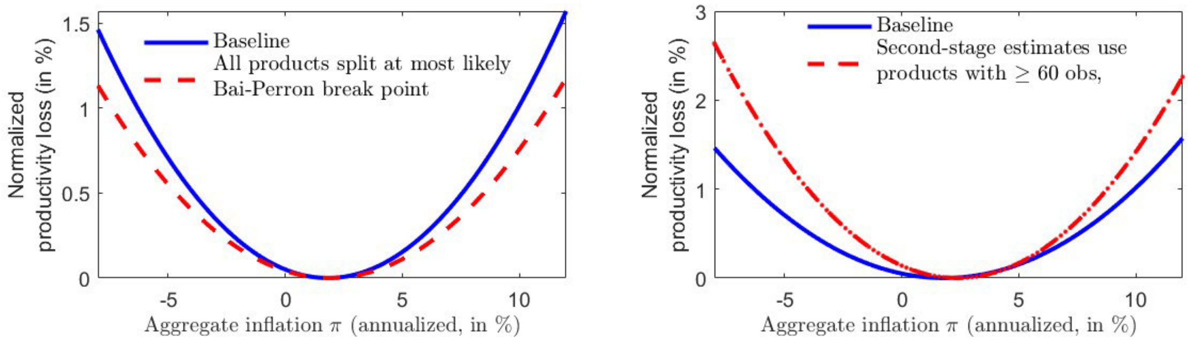

Figure 12: Effects of splitting price time series (left panel) or using longer price time series (right panel)

8.4 How Important Is First-Stage Measurement Error?

Another potential concern with our estimation approach is that the first-stage estimates of desired inflation π^*_{jk}, obtained from equation (34), are contaminated with measurement error. This may cause downward bias in the estimated second-stage coefficients c_k in equation (35) and thereby in the estimated loss $\mathcal{L}(\pi)$.[45]

To address this issue, we re-estimate the coefficients $\{b_k, c_k\}_{k=1}^{K}$ using only price time series with a minimum length of 60 monthly observations (instead of 30 observations as in the baseline). We then apply the estimated coefficients to all products considered in the baseline estimation, when computing the losses.

The resulting productivity loss is depicted in the right panel of figure 12. It shows that first-stage measurement error can lead to a substantial downward bias of the estimated loss. The loss associated with a 10% inflation rate is now 1.4% instead of 1.0% as in the baseline estimation.[46]

Summing up, the findings of this and the previous section imply that the aggregate productivity losses associated with a 10% inflation rate likely lie in a range between 0.8% and 1.4% of aggregate productivity.

9 Summary and Outlook

We propose a new sufficient-statistics approach to estimate the misallocation costs associated with alternative aggregate inflation rates. The framework applies to a broad class of

[45] See appendix C for details.

[46] This finding suggests that the lower losses found in the previous section, where we split the product price time series at the most likely break point, is partly the result of a larger first-stage measurement error.

sticky-price models and accommodates rich heterogeneity in pricing frictions and flexible-price dynamics at the product level. The sufficient statistics can be recovered from micro price data alone and imply that (i) misallocation costs are quantitatively large and (ii) optimal inflation is positive and low. Taken together, these findings substantially strengthen the normative case for targeting low and stable inflation.

Our results also carry important implications for model calibration. Standard calibration strategies target average data moments, yet these moments are largely uninformative about the sufficient statistics governing the response of misallocation to changes in aggregate inflation. Structural models calibrated in the conventional way therefore need not capture empirically relevant misallocation costs. Models designed for the study of optimal inflation should instead be disciplined directly by the sufficient statistics estimated from the data. This raises several questions for future research. What trade-offs arise when targeting standard moments versus sufficient statistics? Can existing sticky-price models match both? How does this calibration strategy alter implications for optimal monetary policy and aggregate dynamics?

Finally, we show that misallocation costs may remain sizable even when inflation is optimal. The magnitude of this effect depends on the role of idiosyncratic mark-up shocks, which cannot be separately identified using micro price data alone. Combining micro price data with information on firm-level costs or mark-ups could sharpen inference about the absolute level of productivity losses due to misallocation, beyond their sensitivity to inflation.

References

Acharya, Sushant, Edouard Challe, and Keshav Dogra (2023). "Optimal Monetary Policy According to HANK". In: *American Economic Review* 113(7), 1741–82.

Adam, Klaus and Henning Weber (2019). "Optimal Trend Inflation". In: *American Economic Review* 109(2), pp. 702–737.

— (2023). "Estimating the Optimal Inflation Target from Trends in Relative Prices". In: *American Economic Journal: Macroeconomics* 15.3, pp. 1–42.

Alvarez, Fernando, Martin Beraja, Martín Gonzalez-Rozada, and Pablo Andrés Neumeyer (2019). "From Hyperinflation to Stable Prices: Argentinas Evidence on Menu Cost Models". In: *The Quarterly Journal of Economics* 134.1, pp. 451–505. ISSN: 0033-5533.

Alvarez, Fernando, Francesco Lippi, and Aleksei Oskolkov (2022). "The Macroeconomics of Sticky Prices and Generalized Hazard Functions". In: *Quarterly Journal of Eonomics* 137, pp. 989–1038.

Bai, Jushan and Pierre Perron (1998). "Estimating and Testing Linear Models with Multiple Structural Changes". In: *Econometrica* 66.1, pp. 47–78.

Baley, Isaac and Andres Blanco (2021). "Aggregate Dynamics in Lumpy Economies". In: *Econometrica* 89, pp. 1235–1264.

Baqaee, David, Emmanuel Farhi, and Kunal Sangani (2024). "Supply Side Effects of Monetary Policy". In: *Journal of Political Economy* 132.4. Journals, pp. 1065–1112.

Bersvendsen, Tore and Jan Ditzen (2021). "Testing for slope heterogeneity in Stata". In: *The Stata Journal* 21.1, pp. 51–80.

Bils, Mark (2009). "Do Higher Prices for New Goods Reflect Quality Growth or Inflation?" In: *Quarterly Journal of Economics* 124, pp. 637–675.

Blanco, Andrés, Corina Boar, Callum J. Jones, and Virgiliu Midrigan (2024). "Nonlinear Inflation Dynamics in Menu Cost Economies". NBER Working Paper 32094.

Blomquist, Johan and Joakim Westerlund (2013). "Testing slope homogeneity in large panels with serial correlation". In: *Economics Letters* 121.3, pp. 374–378. ISSN: 0165-1765.

Calvo, Guillermo A. (1983). "Staggered Contracts in a Utility-Maximizing Framework". In: *Journal of Monetary Economics* 12, pp. 383–398.

Cavallo, Alberto, Francesco Lippi, and Ken Miyahara (2024). "Large Shocks Travel Fast". In: *American Economic Review: Insights* 6.4, pp. 558–74.

Coibion, Olivier, Yuriy Gorodnichenko, and Johannes Wieland (2012). "The Optimal Inflation Rate in New Keynesian Models: Should Central Banks Raise Their Inflation Targets in Light of the Zero Lower Bound?" In: *Review of Economic Studies* 79.4, pp. 1371–1406.

Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo (2011). "Reference prices, costs, and nominal rigidities". In: *American Economic Review* 101.1, pp. 234–62.

Gagliardone, Luca, Mark Gertler, Simone Lenzu, and Joris Tielens (2025). "Micro and Macro Cost-Price Dynamics In Normal Times and During Inflation Surges". NBER Working Paper 33478.

Galí, Jordi (2008). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications.* Princeton University Press.

— (2015). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New KeynesianFramework and Its Applications.* Second Edition, Princeton University Press.

Golosov, Mikhail and Robert E. Lucas (2007). "Menu Costs and Phillips Curves". In: *Journal of Political Economy* 115, pp. 171–199.

Holden, Tom D. (2025). "Rationing Under Sticky Prices". Deutsche Bundesbank, mimeo.

Itskhoki, Oleg and Benjamin Moll (2019). “Optimal Development Policies With Financial Frictions”. In: *Econometrica* 87.1, pp. 139–173.

Jaravel, Xavier (2018). “The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector”. In: *The Quarterly Journal of Economics* 134.2, pp. 715–783.

Kaplan, Greg and Sam Schulhofer-Wohl (2017). “Inflation at the household level”. In: *Journal of Monetary Economics* 91, pp. 19–38.

Karadi, Peter and Adam Reiff (2019). “Menu Costs, Aggregate Fluctuations, and Large Shocks”. In: *American Economic Journal: Macroeconomics* 11.3, pp. 111–46.

Kehoe, Patrick J. and Virgiliu Midrigan (2015). “Prices Are Sticky After All”. In: *Journal of Monetary Economics* 75, pp. 35–53.

King, Robert G., Charles I. Plosser, and Sergio T. Rebelo (May 1988). “Production, growth and business cycles: I. The basic neoclassical model”. In: *Journal of Monetary Economics* 21.2-3, pp. 195–232.

Kryvtsov, Oleksiy and Nicolas Vincent (2021). “The Cyclicality of Sales and Aggregate Price Flexibility”. In: *The Review of Economic Studies* 88.1, pp. 334–377. ISSN: 0034-6527.

La’O, Jennifer and Alireza Tahbaz-Salehi (2022). “Optimal Monetary Policy in Production Networks”. In: *Econometrica* 90.3, pp. 1295–1336.

Meier, Matthias and Timo Reinelt (2024). “Monetary Policy, Markup Dispersion, and Aggregate TFP”. In: *Review of Economics and Statistics* 106.4, pp. 1012–1027.

Midrigan, Virgiliu (2011). “Menu costs, multiproduct firms, and aggregate fluctuations”. In: *Econometrica* 79.4, pp. 1139–1180.

Moll, Benjamin (2014). “Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?” In: *American Economic Review* 104.10, 3186–3221.

Nakamura, Emi and Jón Steinsson (2008). “Five Facs about Prices: A Reevaluation of Menu Cost Models”. In: *Quarterly Journal of Economics* 123, pp. 1415–1464.

— (Aug. 2010). “Monetary Non-neutrality in a Multisector Menu Cost Model”. In: *The Quarterly Journal of Economics* 125.3, pp. 961–1013.

Nakamura, Emi, Jón Steinsson, Patrick Sun, and Daniel Villar (Aug. 2018). “The Elusive Costs of Inflation: Price Dispersion during the U.S. Great Inflation”. In: *The Quarterly Journal of Economics* 133.4, pp. 1933–1980.

Oskolkov, Aleksei and Francesco Lippi (2026). “Misallocation with Lumpy Investment”.

Rotemberg, Julio J. (1982). “Sticky Prices in the United States”. In: *Journal of Political Economy* 90, pp. 1187–1211.

Sara-Zaror, Francisca (2025). “Inflation, Price Dispersion, and Welfare: The Role of Consumer Search”. Available at SSRN: https://ssrn.com/abstract=4127502.

Sheremirov, Viacheslav (2020). “Price Dispersion and Inflation: New Facts and Theoretical Implications”. In: *Journal of Monetary Economics* 114, pp. 59–70. ISSN: 0304-3932.

Taylor, John B. (1999). “Staggered Price and Wage Setting in Macroeconomics”. In: *Handbook of Macroeconomics*. Ed. by John B. Taylor and Michael Woodford. Amsterdam: Elsevier, pp. 1009–1050.

“Testing slope homogeneity in large panels” (2008). In: *Journal of Econometrics* 142.1, pp. 50–93.

Woodford, Michael (2003). *Interest and Prices*. Princeton: Princeton University Press.

Appendix

A Theoretical Results

This section provides all the details of the theoretical model and derives all results in section 4.

A.1 Model

A.1.1 Demand structure

Aggregate consumption C_t is a Cobb-Douglas composite of item-level consumption C_{kt} of item $k = \{1, \ldots, K\}$ at time t given by

$$C_t = \prod_{k=1}^{K} C_{kt}^{\psi_k}, \tag{40}$$

where $\psi_k \geq 0$ denotes expenditure weight for item k and $\sum_{k=1}^{K} \psi_k = 1$. The set of items and the weights ψ_k are fixed over time. Cost-minimizing demand for item-level consumption satisfies

$$C_{kt} = \psi_k \left(\frac{P_{kt}}{P_t}\right)^{-1} C_t, \tag{41}$$

where the price level P_t aggregates item price levels P_{kt} according to

$$P_t = \prod_{k=1}^{K} \left(\frac{P_{kt}}{\psi_k}\right)^{\psi_k}, \tag{42}$$

and the aggregate inflation rate is defined as

$$\pi_t = \ln \frac{P_t}{P_{t-1}}. \tag{43}$$

Item-level consumption is a Dixit-Stiglitz composite of consumption of individual products $j \in [0, 1]$ in the item given by

$$C_{kt} = \left(\int_0^1 (Q_{jkt} C_{jkt})^{\frac{\theta-1}{\theta}} \, \mathrm{dj}\right)^{\frac{\theta}{\theta-1}}, \tag{44}$$

where C_{jkt} denotes quantity consumed (physical units) of product j and $\theta > 1$ denotes the elasticity of substitution between products. The product-specific demand shifter (quality) evolves according to

$$Q_{jkt} = Q_{jk}^{0} \varepsilon_{jkt}^{G} \tag{45}$$

and is thus determined by the initial level Q^0_{jk} drawn at the date when the product enters the market and an idiosyncratic shock denoted by ε^G_{jkt}.[47] The shock is a time t realization of a stationary product-specific stochastic process $\mathfrak{g}_{jk}$, which itself is drawn upon product entry from a time-invariant distribution of stationary stochastic processes.[48]

Let P_{jkt} denote the price of a physical unit of product j of quality level Q_{jkt} in item k. Cost-minimizing demand for product j is given by

$$\frac{Q_{jkt}C_{jkt}}{C_{kt}} = \left(\frac{P_{jkt}/Q_{jkt}}{P_{kt}}\right)^{-\theta}, \tag{46}$$

with quality-adjusted item price level satisfying

$$P_{kt} = \left(\int_0^1 (P_{jkt}/Q_{jkt})^{1-\theta}\, \mathrm{dj}\right)^{\frac{1}{1-\theta}}, \tag{47}$$

and the item-level inflation rate is defined as

$$\pi_{kt} = \ln \frac{P_{kt}}{P_{kt-1}}. \tag{48}$$

For market clearing, consumption C_{jkt} of product j in item k must be equal to production Y_{jkt} with a corresponding requirement at item level. Imposing market clearing at product level ($C_{jkt} = Y_{jkt}$) and item level ($C_{kt} = Y_{kt}$) yields optimal product demand

$$\frac{Q_{jkt}Y_{jkt}}{Y_{kt}} = \left(\frac{P_{jkt}/Q_{jkt}}{P_{kt}}\right)^{-\theta}, \tag{49}$$

$$Y_{kt} = \psi_k \left(\frac{P_{kt}}{P_t}\right)^{-1} Y_t. \tag{50}$$

Analogously to (50), we define ψ_{jkt} to be the expenditure weight of product j in item k at time t, given by:

$$\psi_{jkt} \equiv \frac{P_{jkt}Y_{jkt}}{P_{kt}Y_{kt}} = \left(\frac{P_{jkt}/Q_{jkt}}{P_{kt}}\right)^{1-\theta} \tag{51}$$

A.1.2 Firm technology

Production of product j in item k uses constant returns-to-scale technology

$$Y_{jkt} = A_{kt}G^{-1}_{jkt}L_{jkt}, \tag{52}$$

where A_{kt} denotes the common and exogenous level of productivity in item k, and L_{jkt} denotes labor input. We let item-level productivity grow at an item-specific gross rate γ_k:

$$A_{kt+1} = \gamma_k A_{kt}$$

[47] Here we anticipate that these shocks affect both productivity and quality and denote them with the inverse productivity G superscript.

[48] For example, this allows that different products within an item are subject to stationary shocks of different magnitude and persistence.

Firm-specific inverse productivity is given by

$$G_{jkt} = G^0_{jk} \overline{G}^{s_{jkt}}_{jk} \varepsilon^G_{jkt}, \tag{53}$$

where G^0_{jk} is the initial level drawn at the time of product entry, $\overline{G}_{jk}$ is the product-specific trend component in inverse productivity, s_{jkt} is the time since entry, and ε^G_{jkt} is the inverse productivity/quality idiosyncratic shock drawn from a stationary process $\mathfrak{g}_{jk}$.

A.1.3 Item-level and aggregate technology

Substituting product demand (49) into firm technology (52) yields

$$L_{jkt} = G_{jkt} Q^{\theta-1}_{jkt} \left(\frac{P_{jkt}}{P_{kt}}\right)^{-\theta} \frac{Y_{kt}}{A_{kt}},$$

and aggregating over firms in item k yields

$$\int_0^1 L_{jkt} \text{ dj } = \int_0^1 G_{jkt} Q^{\theta-1}_{jkt} \left(\frac{P_{jkt}}{P_{kt}}\right)^{-\theta} \text{ dj } \frac{Y_{kt}}{A_{kt}}.$$

Denoting $L_{kt} = \int L_{jkt} \text{ dj}$ and defining item-level endogenous productivity as

$$\Delta_{kt} = \int_0^1 G_{jkt} Q^{\theta-1}_{jkt} \left(\frac{P_{jkt}}{P_{kt}}\right)^{-\theta} \text{ dj} \tag{54}$$

yields item-level technology

$$Y_{kt} = \frac{A_{kt}}{\Delta_{kt}} L_{kt}. \tag{55}$$

To derive aggregate technology, we denote $L_t = \sum_{k=1}^K L_{kt}$ and first substitute for L_{kt} using item-level technology (55) and second for Y_{kt} using demand function (50). This yields

$$L_t = Y_t \sum_{k=1}^K \psi_k \left(\frac{P_{kt}}{P_t}\right)^{-1} \left(\frac{A_{kt}}{\Delta_{kt}}\right)^{-1}.$$

Aggregate exogenous productivity is defined as follows:

$$A_t \equiv \prod_{k=1}^K (\psi_k A_{kt})^{\psi_k} \tag{56}$$

and grows at an exogenous gross rate γ:

$$A_{t+1} = \gamma A_t \qquad \gamma = \prod_{k=1}^K \gamma_k^{\psi_k}$$

Multiplying L_t from above by A_t yields:

$$A_t L_t = Y_t \sum_{k=1}^K \psi_k \left(\frac{P_{kt}}{P_t} \frac{A_{kt}}{A_t}\right)^{-1} \Delta_{kt}, \tag{57}$$

and rearranging equation (57) provides aggregate technology

$$Y_t = \frac{A_t}{\Delta_t} L_t, \tag{58}$$

where aggregate endogenous productivity is given by:

$$\Delta_t = \sum_{k=1}^{K} \psi_k \left(\frac{P_{kt}}{P_t} \frac{A_{kt}}{A_t} \right)^{-1} \Delta_{kt}. \tag{59}$$

For later convenience, let product labor input share within items and item labor input share in the aggregate be denoted by ϕ_{jkt} and ϕ_{kt}, respectively:

$$\phi_{jkt} \equiv \frac{L_{jkt}}{L_{kt}} \qquad \phi_{kt} \equiv \frac{L_{kt}}{L_t} \tag{60}$$

A.1.4 Profits and Markups

Nominal period profits of firm j in item k are given by

$$D_{jkt} = (1 + \tau_{jkt}) P_{jkt} Y_{jkt} - W_t L_{jkt} \tag{61}$$

where Y_{jkt} satisfies demand (49) and τ_{jkt} is an exogenous idiosyncratic sales tax/subsidy. Using firm's technology (52) we can rewrite profits as follows:

$$D_{jkt} = \left((1 + \tau_{jkt}) P_{jkt} - \frac{W_t}{A_{kt}} G_{jkt} \right) Y_{jkt}$$

We define the markup of firm j as price over marginal costs according to

$$\mu_{jkt} = \frac{P_{jkt}}{W_t \frac{G_{jkt}}{A_{kt}}}. \tag{62}$$

In particular, with flexible prices,

$$\mu^f_{jkt} = \frac{1}{1 + \tau_{jkt}} \frac{\theta}{\theta - 1} \tag{63}$$

$$P^*_{jkt} = \mu^f_{jkt} \frac{W_t}{A_{kt}} G_{jkt} \tag{64}$$

In the following we refer to μ^f_{jkt} as desired markups and to P^*_{jkt} as the flexible price.[49] Analogous to inverse productivity G_{jkt}, we assume that desired markups follow:

$$\mu^f_{jkt} = \mu^{f,0}_{jk} (\overline{\mu}^f_{jk})^{s_{jkt}} \varepsilon^{\mu}_{jkt}, \tag{65}$$

with initial level $\mu^{f,0}_{jk}$, trend component $\overline{\mu}^f_{jk}$, and idiosyncratic shock ε^{μ}_{jkt} realized from a potentially product-specific stationary process $\mathfrak{m}_{jk}$. As with inverse productivity/quality

[49] To be more precise, this is the price that maximizes firm's period profits in a sticky-price equilibrium, since it uses the corresponding equilibrium wage W_t.

shocks, desired markup shock process $\mathfrak{m}_{jk}$ is drawn upon product entry from a time-invariant distribution of stationary stochastic processes. Using expenditure weight definition (51) and rewriting equation (62) yields

$$P_{kt}\frac{\psi_{jkt}}{\mu_{jkt}} = P_{kt}\frac{P_{jkt}Y_{jkt}}{P_{kt}Y_{kt}}\mu_{jkt}^{-1} = W_t\frac{G_{jkt}}{A_{kt}}\frac{Y_{jkt}}{Y_{kt}},$$

and integrating across firms in the item yields

$$P_{kt}\int_0^1 \frac{\psi_{jkt}}{\mu_{jkt}}\,\mathrm{dj} \ = \frac{W_t}{A_{kt}}\int_0^1 G_{jkt}\frac{Y_{jkt}}{Y_{kt}}\,\mathrm{dj}\ . \tag{66}$$

First, equation (54) and product demand (46) imply that the RHS integral is equal to Δ_{kt}. Second, we define the item-level markup μ_{kt} as expenditure weighted harmonic mean of product-level markups according to

$$\mu_{kt} = \left(\int_0^1 \frac{\psi_{jkt}}{\mu_{jkt}}\,\mathrm{dj}\right)^{-1}. \tag{67}$$

Using both implications, we rearrange equation (66) to obtain

$$P_{kt} = \mu_{kt}\frac{W_t}{A_{kt}/\Delta_{kt}}. \tag{68}$$

Thus, at the item level, average price is equal to average markup times average marginal costs. We correspondingly define the aggregate markup μ_t as expenditure weighted harmonic mean of item-level markups according to

$$\mu_t = \left(\sum_{k=1}^{K}\frac{P_{kt}Y_{kt}}{P_tY_t}\,\frac{1}{\mu_{kt}}\right)^{-1},$$

which after substituting product demand (50) yields

$$\mu_t = \left(\sum_{k=1}^{K}\psi_k\frac{1}{\mu_{kt}}\right)^{-1}. \tag{69}$$

Rearranging equation (68) according to

$$\frac{P_{kt}Y_{kt}}{P_tY_t}\frac{1}{\mu_{kt}} = \frac{W_t}{P_tA_t}\frac{Y_{kt}/A_{kt}}{Y_t/A_t}\Delta_{kt} \tag{70}$$

and summing over items yields

$$\sum_{k=1}^{K}\frac{P_{kt}Y_{kt}}{P_tY_t}\frac{1}{\mu_{kt}} = \frac{W_t}{P_tA_t}\sum_{k=1}^{K}\frac{Y_{kt}/A_{kt}}{Y_t/A_t}\Delta_{kt}.$$

Using product demand (50) yields

$$\sum_{k=1}^{K}\psi_k\frac{1}{\mu_{kt}} = \frac{W_t}{P_tA_t}\sum_{k=1}^{K}\psi_k\left(\frac{P_{kt}A_{kt}}{P_tA_t}\right)^{-1}\Delta_{kt}.$$

Using markup in equation (69) and productivity definition (59) further yields

$$P_t = \mu_t \frac{W_t}{A_t/\Delta_t}. \tag{71}$$

Finally, note that using (50), (55), (58), (68) and (71), we obtain:

$$\begin{aligned}\psi_k \frac{\mu_t}{\mu_{kt}} &= \frac{P_{kt}Y_{kt}}{P_tY_t}\frac{P_tA_t}{W_t\Delta_t}\frac{W_t\Delta_{kt}}{P_{kt}A_{kt}} \\ &= \frac{A_t}{Y_t\Delta_t}\frac{Y_{kt}\Delta_{kt}}{A_{kt}} = \frac{L_{kt}}{L_t} = \phi_{kt}\end{aligned} \tag{72}$$

which links item labor input share to its expenditure share and its markup relative to the aggregate one.

A.1.5 Relationship between item and aggregate distortions

Dividing item-level and aggregate pricing equations (68) and (71), respectively, by each other yields

$$\frac{P_{kt}}{P_t} = \frac{\mu_{kt}}{\mu_t}\frac{A_t/\Delta_t}{A_{kt}/\Delta_{kt}}, \tag{73}$$

and shows that relative prices are equal to relative markups times inverse relative productivities. Detrending relative prices yields

$$\frac{P_{kt}}{P_t}\frac{A_{kt}}{A_t} = \frac{\mu_{kt}\Delta_{kt}}{\mu_t\Delta_t}. \tag{74}$$

Taking exponent ψ_k and aggregating over items yields

$$\prod_{k=1}^{K}\left(\frac{P_{kt}}{P_t}\frac{A_{kt}}{A_t}\right)^{\psi_k} = \prod_{k=1}^{K}\left(\frac{\mu_{kt}\Delta_{kt}}{\mu_t\Delta_t}\right)^{\psi_k}.$$

Multiplying price level equation (42) by the definition (56) of the aggregate growth trend yields

$$1 = \prod_{k=1}^{K}\left(\frac{P_{kt}}{P_t}\frac{A_{kt}}{A_t}\right)^{\psi_k}. \tag{75}$$

We thus obtain

$$1 = \prod_{k=1}^{K}\left(\frac{\mu_{kt}\Delta_{kt}}{\mu_t\Delta_t}\right)^{\psi_k}$$

and hence the relationship

$$\mu_t\Delta_t = \prod_{k=1}^{K}\left(\mu_{kt}\Delta_{kt}\right)^{\psi_k} \tag{76}$$

between aggregate and item-level distortions.

A.1.6 Household preferences

We use growth-consistent KPR preferences (the same preferences as in Adam and Weber (2019)), which are given by

$$U(C_t, L_t) = \frac{[C_t \cdot V(L_t)]^{1-\sigma} - 1}{1-\sigma}.$$

We assume $\sigma > 0$ and that $V(L_t)$ is such that period utility is strictly concave in (C_t, L_t) and that Inada conditions are satisfied. The flow budget constraint of the household reads:

$$P_t C_t = W_t L_t + T_t$$

where T_t includes firm profits and government transfers/taxes.[50] First-order conditions imply:

$$\frac{V'(L_t)}{V(L_t)} = -\frac{W_t}{P_t C_t} \tag{77}$$

A.1.7 Pricing Frictions

Each firm j maximizes the expected discounted stream of real period profits:

$$E_t \sum_{i=0}^{\infty} \beta^i (1-\delta_k)^i \frac{U_{Ct+i}}{U_{Ct}} \frac{D_{jkt+i}}{P_{t+i}}$$

minus possible costs associated with price adjustments, given the household's discount factor and item-level rate of exogenous exit δ_k, and subject to unmodeled and potentially product-specific pricing frictions $\mathfrak{f}_{jk}$. Frictions $\mathfrak{f}_{jk}$ are defined implicitly as the modeling element that is required to uniquely determine the distribution of firm's adjustment times (conditional on the history of shocks), given the information on the dynamics of aggregate and item-level variables, firm's shock processes $\mathfrak{g}_{jk}$ and $\mathfrak{m}_{jk}$, and firm's inflation gap $\hat{\pi}_{jk} \equiv \pi_k - \pi^*_{jk}$.[51] These frictions are elements of set $\mathfrak{F}$, which is the set of all pricing frictions satisfying assumption 1. For each product, $\mathfrak{f}_{jk}$ is drawn independently from inflation gaps $\hat{\pi}_{jk}$ within the item k, in accordance with assumption 2.2.

Assumption 1.2 guarantees that, given the distribution of adjustment times, at every adjustment time $t = T_{jkh}$, firm solves the following problem:

$$\max_{P_{jkt}} E_t \sum_{i=0}^{S_{jkh}-1} \beta^i (1-\delta_k)^i \frac{U_{Ct+i}}{U_{Ct}} \left((1+\tau_{jkt+i}) \frac{P_{jkt}}{P_{t+i}} - \frac{W_{t+i}}{A_{kt+i} P_{t+i}} G_{jkt+i} \right) Y_{jkt+i}$$

where S_{jkh} denotes the duration of price spell following adjustment h. Demand Y_{jkt+i} depends on the firm's price via product demand in equations (49) and (50). Augmenting

[50] Without loss of generality we ignore the presence of government bonds and money demand for simplicity.

[51] See appendix A.5 for more details on the determinants of price gap distributions.

the equation and dividing it by Y_t yields

$$\max_{P_{jkt}} E_t \sum_{i=0}^{S_{jkh}-1} \beta^i (1-\delta_k)^i \frac{U_{Ct+i}}{U_{Ct}} \frac{Y_{t+i}}{Y_t} \Bigg((1+\tau_{jkt+i}) \frac{P_{jkt}}{P_{kt+i}} - \frac{W_{t+i}}{P_{kt+i} A_{kt+i}} G_{jkt+i} \Bigg) \frac{Y_{jkt+i}}{Y_{kt+i}} \frac{P_{kt+i} Y_{kt+i}}{P_{t+i} Y_{t+i}}$$

Using item and product demand functions (49) and (50), we can rewrite:

$$\max_{P_{jkt}} \psi_k E_t \sum_{i=0}^{S_{jkh}-1} \beta^i (1-\delta_k)^i \frac{U_{Ct+i}}{U_{Ct}} \frac{Y_{t+i}}{Y_t} \Bigg((1+\tau_{jkt+i}) \frac{P_{jkt}}{P_{kt+i}} - \frac{W_{t+i}}{P_{kt+i} A_{kt+i}} G_{jkt+i} \Bigg) \left(\frac{P_{jkt}}{P_{kt+i}} \right)^{-\theta} Q_{jkt+i}^{\theta-1}$$

Taking the derivative wrt P_{jkt} yields optimality condition

$$E_t \sum_{i=0}^{S_{jkh}-1} \beta^i (1-\delta_k)^i \frac{U_{Ct+i}}{U_{Ct}} \frac{Y_{t+i}}{Y_t} \Bigg(\left(\mu_{jkt+i}^f \right)^{-1} \left(\frac{P_{jkt}}{P_{kt+i}} \right)^{1-\theta} - \frac{W_{t+i}}{P_{kt+i} A_{kt+i}} G_{jkt+i} \left(\frac{P_{jkt}}{P_{kt+i}} \right)^{-\theta} \Bigg) Q_{jkt+i}^{\theta-1} = 0 \tag{78}$$

where we have used the definition of desired markup (63).

A.1.8 Price Gaps

We define the price gap as

$$z_{jkt} \equiv \frac{P_{jkt}}{P_{jkt}^*} \tag{79}$$

where P_{jkt}^* denotes the price that maximizes firm's period profits (the flexible price defined in (64)):

$$P_{jkt}^* = \mu_{jkt}^f \frac{W_t}{A_{kt}} G_{jkt}$$

In the following, we rewrite the key equations of the model in terms of price gaps.

Real wage. For convenience, define scaled real wage and relative item price levels as follows:

$$p_{kt} \equiv \frac{P_{kt} A_{kt}}{P_t A_t}, \quad k = 1, \dots K \tag{80}$$

$$w_t \equiv \frac{W_t}{P_t A_t} \tag{81}$$

Then, equation (64) implies:

$$\frac{P_{jkt}}{P_{kt}} = \mu_{jkt}^f G_{jkt} \frac{w_t}{p_{kt}} z_{jkt}, \tag{82}$$

Dividing by Q_{jkt}, taking exponent $1-\theta$ of the equation and aggregating across firms yields

$$\int_0^1 \left(\frac{P_{jkt}/Q_{jkt}}{P_{kt}}\right)^{1-\theta} \mathrm{dj} = \left(\frac{w_t}{p_{kt}}\right)^{1-\theta} \int_0^1 \left(\mu^f_{jkt} \frac{G_{jkt}}{Q_{jkt}} z_{jkt}\right)^{1-\theta} \mathrm{dj}.$$

By the quality-adjusted item price level in equation (47), the LHS of the previous equation is equal to one, which yields

$$1 = \left(\frac{w_t}{p_{kt}}\right)^{1-\theta} \int_0^1 \left(\mu^f_{jkt} \frac{G_{jkt}}{Q_{jkt}} z_{jkt}\right)^{1-\theta} \mathrm{dj} \tag{83}$$

Item-level endogenous productivity Δ_{kt}. Substituting (82) into equation (54) yields

$$\Delta_{kt} = \left(\frac{w_t}{p_{kt}}\right)^{-\theta} \int_0^1 \left(\frac{G_{jkt}}{Q_{jkt}}\right)^{1-\theta} \left(\mu^f_{jkt} z_{jkt}\right)^{-\theta} \mathrm{dj}\ . \tag{84}$$

Expenditure weights and markups. Using expenditure weight definition (51), and the expressions for desired price (64) and price gap (79), write:

$$\psi_{jkt} = \left(\frac{w_t}{p_{kt}} \mu^f_{jkt} \frac{G_{jkt}}{Q_{jkt}} z_{jkt}\right)^{1-\theta} \tag{85}$$

Rearranging equation (64) and using equation (79) yields

$$\mu_{jkt} = \frac{P_{jkt}}{W_t \frac{G_{jkt}}{A_{kt}}} = \mu^f_{jkt} z_{jkt},$$

We use the above two results to derive an expression for the item-level markup :

$$(\mu_{kt})^{-1} = \int_0^1 \frac{\psi_{jkt}}{\mu_{jkt}} \mathrm{dj} \ = \left(\frac{w_t}{p_{kt}}\right)^{1-\theta} \int_0^1 \left(\frac{G_{jkt}}{Q_{jkt}}\right)^{1-\theta} \left(\mu^f_{jkt} z_{jkt}\right)^{-\theta} \mathrm{dj}\ . \tag{86}$$

Note also that expenditure-weighted markup can be written as:

$$\frac{\psi_{jkt}}{\mu_{jkt}} = \frac{P_{jkt} Y_{jkt}}{P_{kt} Y_{kt}} \mu^{-1}_{jkt} = \frac{w_t}{p_{kt}} \Delta_{kt} \frac{L_{jkt}}{L_{kt}} = \frac{w_t}{p_{kt}} \Delta_{kt} \phi_{jkt}$$

and therefore labor input share can be expressed as:

$$\phi_{jkt} = \frac{1}{\Delta_{kt}} \left(\frac{w_t}{p_{kt}}\right)^{-\theta} \left(\frac{G_{jkt}}{Q_{jkt}}\right)^{1-\theta} \left(\mu^f_{jkt} z_{jkt}\right)^{-\theta} \tag{87}$$

A.1.9 Balanced growth path

Our analysis involves three different economies: the main economy of interest – stochastic sticky-price economy, the benchmark economy – stochastic flex-price economy, and the approximation point – deterministic flex-price economy. Note that all the derivations and results above apply to all three economies, with the latter two being special cases of the general framework. In particular, setting price rigidities $\mathfrak{f}_{jk}$ for each product to an element of $\mathfrak{F}$ corresponding to flexible prices, results in price gaps z_{jkt} being equal to one for all products at all times and recovers the stochastic flex-price economy. Additionally letting $\mathfrak{g}_{jk}$ and $\mathfrak{m}_{jk}$ be degenerate processes implies $\ln \varepsilon^{G}_{jkt} = \ln \varepsilon^{\mu}_{jkt} = 0$ for for all products at all times, and therefore recovers the deterministic flex-price economy.

For a meaningful approximation of productivity loss, we require that all three economies are on the same balanced growth path, meaning that all aggregate and item-level variables grow at the same rate across the three economies, and therefore any differences are in levels only. In particular, we consider balanced growth paths, on which aggregate and item-level labor, L_t and L_{kt}, are constant over time, and aggregate and item-level output grow at the corresponding rates of exogenous productivity growth, γ and γ_k:

$$Y_{t+1} = \gamma Y_t \qquad Y_{kt+1} = \gamma_k Y_{kt}$$

The next proposition establishes a necessary and sufficient condition for a BGP that applies to each of the three economies.

Proposition 8 (BGP) *An economy is on a balanced growth path with constant aggregate and item-level labor, and with aggregate and item-level output growing at corresponding rates of exogenous productivity growth, if and only if for all $t \geq 0$ and all $k \in \{1, \dots, K\}$ the integrals:*

$$\int_0^1 \left(\frac{G_{jkt}}{Q_{jkt}}\right)^{1-\theta} \left(\mu^{f}_{jkt} z_{jkt}\right)^{-\theta} \, dj \tag{88}$$

$$\int_0^1 \left(\mu^{f}_{jkt} \frac{G_{jkt}}{Q_{jkt}} z_{jkt}\right)^{1-\theta} \, dj \tag{89}$$

are item-specific (potentially distinct) constants.

As discussed above, in flex-price economies, $z_{jkt} = 1$ for all products and all time periods, and in the deterministic economy fundamental processes do not have stochastic components, but the general form of conditions (88) and (89) still applies. The proof of proposition 8, provided in section A.4.4, also shows that the following objects are constant over time on a balanced growth path: endogenous productivity at the item and aggregate level ($\Delta_{kt} = \Delta_k$, $\Delta_t = \Delta$), item-level and aggregate markups ($\mu_{kt} = \mu_k$, $\mu_t = \mu$), scaled real wage $w_t = w$, and relative item price levels $p_{kt} = p_k$. The latter result implies that the difference between aggregate and item-level inflation is constant over time:

$$\pi_k = \pi + \ln \frac{\gamma}{\gamma_k}$$

In the following we discuss the assumptions that we impose to ensure BGP in all three economies.

First, we adopt the following condition on the joint distribution of fundamental intercepts and trends, which reproduces (88) and (89) for the deterministic flex-price economy and therefore ensures its BGP by proposition 8.

Condition 1 *We assume that for all $t \geq 0$ and all $k \in \{1, \ldots, K\}$ the integrals:*

$$\int_0^1 \left(\frac{G^0_{jk} \overline{G}^{s_{jkt}}_{jk}}{Q^0_{jk}} \right)^{1-\theta} \left(\mu^{f,0}_{jk} (\overline{\mu}^f_{jk})^{s_{jkt}} \right)^{-\theta} \, dj$$

$$\int_0^1 \left(\frac{G^0_{jk} \overline{G}^{s_{jkt}}_{jk}}{Q^0_{jk}} \mu^{f,0}_{jk} (\overline{\mu}^f_{jk})^{s_{jkt}} \right)^{1-\theta} \, dj$$

are item-specific (potentially distinct) constants.

Second, we consider the stochastic flex-price economy. The next lemma states that under the assumptions of our setup, BGP condition 1 also implies a balanced growth path in that economy.

Lemma 2 *Suppose assumption 2.1 and BGP condition 1 hold. Then the stochastic flex-price economy is on a balanced growth path, in which aggregate and item-level variables grow at the same rate as in the deterministic flex-price economy.*

The formal proof is provided in section A.4.4, and the key idea is that since shock process distributions from which products draw upon entry are time-invariant, (88) and (89) are satisfied for a stochastic flex-price economy if these conditions are satisfied for a deterministic one.

Finally, assumption 1.1 ensures that the stochastic sticky-price economy is also on a balanced growth path, with aggregate and item-level variables featuring the same growth rates as in the two flex-price economies. More formally, this implies that pricing frictions induce price gap distributions such that:

$$\int_0^1 \left(\varepsilon^{\mu}_{jkt} z_{jkt} \right)^{-\theta} \, \mathrm{dj} \tag{90}$$

$$\int_0^1 \left(\varepsilon^{\mu}_{jkt} z_{jkt} \right)^{1-\theta} \, \mathrm{dj} \tag{91}$$

are both constant over time in each item k.[52]

A.1.10 Summary

Here we summarize the main equations that we need for the subsequent approximation, imposing the balanced growth path. From (83), (84) and (86), the scaled real wage,

[52] We omit the proof of this claim as it is not essential for our analysis, but it can be obtained using assumption 2.1 and following the steps of proofs of lemma 2 and lemma 3, stated below.

endogenous productivity, and average markup in each item are given by:

$$1 = \left(\frac{w}{p_k}\right)^{1-\theta} \int_0^1 \left(\mu^f_{jkt}\frac{G_{jkt}}{Q_{jkt}}z_{jkt}\right)^{1-\theta} \mathrm{dj} \tag{92}$$

$$\Delta_k = \left(\frac{w}{p_k}\right)^{-\theta} \int_0^1 \left(\frac{G_{jkt}}{Q_{jkt}}\right)^{1-\theta} \left(\mu^f_{jkt}z_{jkt}\right)^{-\theta} \mathrm{dj} \tag{93}$$

$$(\mu_k)^{-1} = \left(\frac{w}{p_k}\right)^{1-\theta} \int_0^1 \left(\frac{G_{jkt}}{Q_{jkt}}\right)^{1-\theta} \left(\mu^f_{jkt}z_{jkt}\right)^{-\theta} \mathrm{dj} \tag{94}$$

Product-level expenditure and labor input shares (85) and (87) become:

$$\psi_{jkt} = \left(\frac{w}{p_k}\right)^{1-\theta} \left(\mu^f_{jkt}\frac{G_{jkt}}{Q_{jkt}}z_{jkt}\right)^{1-\theta} \tag{95}$$

$$\phi_{jkt} = \frac{1}{\Delta_k}\left(\frac{w}{p_k}\right)^{-\theta} \left(\frac{G_{jkt}}{Q_{jkt}}\right)^{1-\theta} \left(\mu^f_{jkt}z_{jkt}\right)^{-\theta} \tag{96}$$

Since aggregate labor is constant and aggregate output grows at rate γ on a BGP, and using market clearing ($C_t = Y_t$), we obtain:

$$\frac{U_{Ct+i}}{U_{Ct}}\frac{Y_{t+i}}{Y_t} = \left(\frac{C_{t+i}}{C_t}\right)^{-\sigma}\frac{Y_{t+i}}{Y_t} = \left(\frac{Y_{t+i}}{Y_t}\right)^{1-\sigma} = \left(\gamma^{1-\sigma}\right)^i \tag{97}$$

Such that firm's optimal reset price condition (78) becomes:

$$E_t \sum_{i=0}^{S_{jkh}-1} \left[\beta\left(1-\delta_k\right)\gamma^{1-\sigma}\right]^i \left(\left(\mu^f_{jkt+i}\right)^{-1}\left(\frac{P_{jkt}}{P_{kt+i}}\right)^{1-\theta} - \frac{w}{p_k}G_{jkt+i}\left(\frac{P_{jkt}}{P_{kt+i}}\right)^{-\theta}\right) Q^{\theta-1}_{jkt+i} = 0 \tag{98}$$

Using the definition of P^*_{jkt} (64), the reset price can be written as

$$\frac{P_{jkt}}{P_{kt+i}} = \mu^f_{jkt+i}\frac{w}{p_k}G_{jkt+i}z_{jkt+i}$$

Plugging into the optimality condition above, dividing through by w/p_k and using (63) yields

$$E_t \sum_{i=0}^{S_{jkh}-1} \left[\beta\gamma^{1-\sigma}(1-\delta_k)\right]^i \left(\left(\mu^f_{jkt+i}\right)^{-\theta}(z_{jkt+i})^{1-\theta} - \left(\mu^f_{jkt+i}z_{jkt+i}\right)^{-\theta}\right)\left(\frac{G_{jkt+i}}{Q_{jkt+i}}\right)^{1-\theta} = 0 \tag{99}$$

Equations (45), (53) and (65) imply

$$Q_{jkt+i} = \left[Q^0_{jk}\right]\varepsilon^G_{jkt+i}$$

$$G_{jkt+i} = G^0_{jk}\overline{G}^{s_{jkt}+i}_{jk}\varepsilon^G_{jkt+i} = \left[G^0_{jk}\overline{G}^{s_{jkt}}_{jk}\right]\overline{G}^i_{jk}\varepsilon^G_{jkt+i}$$

$$\mu^f_{jkt+i} = \mu^{f,0}_{jk}\left(\overline{\mu}^f_{jk}\right)^{s_{jkt}+i}\varepsilon^\mu_{jkt+i} = \left[\mu^{f,0}_{jk}\left(\overline{\mu}^f_{jk}\right)^{s_{jkt}}\right]\left(\overline{\mu}^f_{jk}\right)^i\varepsilon^\mu_{jkt+i}$$

where the terms in square brackets are independent of i. Substituting all three processes into the optimality condition and eliminating terms independent of i yields

$$E_t \sum_{i=0}^{S_{jkh}-1} \left[\beta\gamma^{1-\sigma}(1-\delta_k)\overline{G}_{jk}^{1-\theta}\left(\overline{\mu}_{jk}^f\right)^{-\theta}\right]^i \left(\left(\varepsilon_{jkt+i}^{\mu}\right)^{-\theta} z_{jkt+i}^{1-\theta} - \left(\varepsilon_{jkt+i}^{\mu} z_{jkt+i}\right)^{-\theta}\right) = 0 \quad (100)$$

A.2 Productivity Loss Approximation

This section derives the results in propositions 1 and 2 in the main text. To ease notation, we omit the time subscript for variables that are constant along the BGP and substitute integrals with expectation operators such that for any variable x_{jkt} with a stationary within-item distribution:

$$\mathbb{E}_k[x] \equiv \int_0^1 x_{jkt} \mathrm{dj}$$

Combining (69) and (76) yields:

$$\Delta = \left(\sum_{k=1}^{K} \psi_k \frac{1}{\mu_k}\right) \prod_{k=1}^{K} (\mu_k \Delta_k)^{\psi_k} \quad (101)$$

where equations (92), (93) and (94) imply:

$$\mu_k \Delta_k = \left(\frac{w}{p_k}\right)^{-1} \quad (102)$$

$$(\mu_k)^{-1} = \left(\frac{w}{p_k}\right)^{1-\theta} \mathbb{E}_k\left[G^{1-\theta}\left(\mu^f z\right)^{-\theta} Q^{\theta-1}\right] \quad (103)$$

$$\frac{w}{p_k} = \left(\mathbb{E}_k\left[\left(G\mu^f z\right)^{1-\theta} Q^{\theta-1}\right]\right)^{\frac{1}{\theta-1}} \quad (104)$$

We denote the stochastic flex-price economy variables with a superscript f, and the corresponding equations for this economy are:

$$\Delta^f = \left(\sum_{k=1}^{K} \psi_k \frac{1}{\mu_k^f}\right) \prod_{k=1}^{K} \left(\mu_k^f \Delta_k^f\right)^{\psi_k} \quad (105)$$

$$\mu_k^f \Delta_k^f = \left(\frac{w^f}{p_k^f}\right)^{-1} \quad (106)$$

$$(\mu_k^f)^{-1} = \left(\frac{w^f}{p_k^f}\right)^{1-\theta} \mathbb{E}_k\left[G^{1-\theta}\left(\mu^f\right)^{-\theta} Q^{\theta-1}\right] \quad (107)$$

$$\frac{w^f}{p_k^f} = \left(\mathbb{E}_k\left[\left(G\mu^f\right)^{1-\theta} Q^{\theta-1}\right]\right)^{\frac{1}{\theta-1}} \quad (108)$$

Note that flex-price expressions are identical to the sticky-price ones, once price gaps z are set to one. We are interested in productivity loss defined as:

$$\ln\frac{\Delta}{\Delta^f} = \sum_{k=1}^{K} \psi_k \ln\frac{\mu_k \Delta_k}{\mu_k^f \Delta_k^f} - \left(\ln\sum_{k=1}^{K} \psi_k \frac{1}{\mu_k^f} - \ln\sum_{k=1}^{K} \psi_k \frac{1}{\mu_k}\right) \quad (109)$$

In the following, we approximate productivity loss $\ln \frac{\Delta}{\Delta^f}$ around the deterministic flexible-price economy.[53] First, we decompose products of G, μ^f, z and Q into products of intercepts, trends and stationary shock components. Define:

$$\begin{aligned}
c_{jk}^{f,0} &= (1-\theta)(\ln G_{jk}^0 - \ln Q_{jk}^0 + \ln \mu_{jk}^{f,0}) \\
c_{jk}^{f,1} &= (1-\theta)(\ln \bar{G}_{jk} + \ln \bar{\mu}_{jk}^f) \\
c_{jk}^{g,0} &= (1-\theta)(\ln G_{jk}^0 - \ln Q_{jk}^0) - \theta \ln \mu_{jk}^{f,0} \\
c_{jk}^{g,1} &= (1-\theta)\ln \bar{G}_{jk} - \theta \ln \bar{\mu}_{jk}^f \\
\ln f(\ln z, \ln \varepsilon^\mu) &= (1-\theta)\left(\ln z + \ln \varepsilon^\mu\right) \\
\ln g(\ln z, \ln \varepsilon^\mu) &= -\theta \left(\ln z + \ln \varepsilon^\mu\right)
\end{aligned}$$

such that:

$$\begin{aligned}
\mathbb{E}_k\left[\left(G\mu^f z\right)^{1-\theta} Q^{\theta-1}\right] &= \mathbb{E}_k\left[e^{c^{f,0}+sc^{f,1}} f\left(\ln z, \ln \varepsilon^\mu\right)\right] \\
\mathbb{E}_k\left[G^{1-\theta}\left(\mu^f z\right)^{-\theta} Q^{\theta-1}\right] &= \mathbb{E}_k\left[e^{c^{g,0}+sc^{g,1}} g\left(\ln z, \ln \varepsilon^\mu\right)\right]
\end{aligned}$$

where s is the time since product entry. The following lemma establishes that we can separate deterministic and stochastic components.

Lemma 3 *Under assumption 2:*

$$\begin{aligned}
\ln \mathbb{E}_k\left[\left(G\mu^f z\right)^{1-\theta} Q^{\theta-1}\right] &= \ln \mathbb{E}_k\left[e^{c^{f,0}+sc^{f,1}}\right] + \ln \mathbb{E}_k\left[f\left(\ln z, \ln \varepsilon^\mu\right)\right] \\
\ln \mathbb{E}_k\left[G^{1-\theta}(\mu^f z)^{-\theta} Q^{\theta-1}\right] &= \ln \mathbb{E}_k\left[e^{c^{g,0}+sc^{g,1}}\right] + \ln \mathbb{E}_k\left[g\left(\ln z, \ln \varepsilon^\mu\right)\right]
\end{aligned}$$

The proof is provided in section A.4.4. From Lemma 3 it follows:

$$\ln \frac{\mu_k \Delta_k}{\mu_k^f \Delta_k^f} = \frac{1}{1-\theta}\Big(\ln \mathbb{E}_k\left[f\left(\ln z, \ln \varepsilon^\mu\right)\right] - \ln \mathbb{E}_k\left[f\left(0, \ln \varepsilon^\mu\right)\right]\Big) \tag{110}$$

To approximate μ_k^{-1}, denote by $\alpha_k = \frac{\mathbb{E}_k\left[e^{c^{g,0}+sc^{g,1}}\right]}{\mathbb{E}_k\left[e^{c^{f,0}+sc^{f,1}}\right]}$ the item-level (inverse) markup in the deterministic flex-price economy. Note that assumption 4 implies that $\alpha_k = \alpha$ and is constant across items, since labor input share ϕ_k is linked to item-level markup μ_k by (72) and is equal to expenditure share ψ_k by the assumption. We can then approximate $\frac{1}{\mu_k} = e^{-\ln \mu_k}$ around $\ln \mu_k = -\ln \alpha$:

$$\frac{1}{\mu_k} = e^{-\ln \mu_k} = \alpha\left(1 - (\ln \mu_k + \ln \alpha) + \frac{1}{2}(\ln \mu_k + \ln \alpha)^2\right) + O((\ln \mu_k + \ln \alpha)^3) \tag{111}$$

[53] In fact, our results also apply to an approximation around a stochastic flexible-price economy, as long as there are no markup shocks and the only source of uncertainty are productivity/quality shocks. We omit this qualification in the main text to ease exposition.

and a similar approximation applies to the average item-level markup in the stochastic flex-price economy μ_k^f. Using Lemma 3, their log-values can be expressed as:

$$\ln \mu_k = -\ln \alpha + \ln \mathbb{E}_k \left[f\left(\ln z, \ln \varepsilon^\mu\right)\right] - \ln \mathbb{E}_k \left[g\left(\ln z, \ln \varepsilon^\mu\right)\right] \tag{112}$$

$$\ln \mu_k^f = -\ln \alpha + \ln \mathbb{E}_k \left[f\left(0, \ln \varepsilon^\mu\right)\right] - \ln \mathbb{E}_k \left[g\left(0, \ln \varepsilon^\mu\right)\right] \tag{113}$$

In the following we approximate (110), (112) and (113) around the deterministic flex-price economy (setting $\ln z = \ln \varepsilon^\mu = 0$). The details on the expansion method and derivations are provided in section A.4.1. The resulting equations are:

$$\ln \frac{\mu_k \Delta_k}{\mu_k^f \Delta_k^f} = \mathbb{E}_k[\ln z] + \frac{(1-\theta)}{2}\Big(Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu]\Big) + O(\mathbb{E}_k[||(\ln z, \ln \varepsilon^\mu)||^3]) \tag{114}$$

$$\begin{aligned} \ln \mu_k = &-\ln \alpha + \mathbb{E}_k[\ln z] + \frac{1-2\theta}{2}\Big(Var_k(\ln z) + \mathbb{E}_k[(\ln \varepsilon^\mu)^2] + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu]\Big) \\ &+ O(\mathbb{E}_k[||(\ln z, \ln \varepsilon^\mu)||^3]) \end{aligned} \tag{115}$$

$$\ln \mu_k^f = -\ln \alpha + \frac{1-2\theta}{2}\mathbb{E}_k[(\ln \varepsilon^\mu)^2] + O(\mathbb{E}_k[||(\ln z, \ln \varepsilon^\mu)||^3]) \tag{116}$$

Combining the above equations delivers:

$$\ln \frac{\Delta_k}{\Delta_k^f} = \frac{\theta}{2}\Big(Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu]\Big) + O(\mathbb{E}_k[||(\ln z, \ln \varepsilon^\mu)||^3]) \tag{117}$$

which proves proposition 1 in the main text. Plugging (115) into (111) we obtain:

$$\begin{aligned} \frac{1}{\mu_k} = \alpha\Bigg[&1 - \mathbb{E}_k[\ln z] - \frac{1-2\theta}{2}\Big(Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu] + \mathbb{E}_k[(\ln \varepsilon^\mu)^2]\Big) \\ &+ O(\mathbb{E}_k[||(\ln z, \ln \varepsilon^\mu)||^3]) \\ &+ \frac{1}{2}\Bigg(\mathbb{E}_k[\ln z] + \frac{1-2\theta}{2}\Big(Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu] + \mathbb{E}_k[(\ln \varepsilon^\mu)^2]\Big) \\ &+ O(\mathbb{E}_k[||(\ln z, \ln \varepsilon^\mu)||^3])\Bigg)^2 \\ &+ O\Bigg(\Big(\mathbb{E}_k[\ln z] + \frac{1-2\theta}{2}\Big(Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu] + \mathbb{E}_k[(\ln \varepsilon^\mu)^2]\Big) \\ &+ O(\mathbb{E}_k[||(\ln z, \ln \varepsilon^\mu)||^3])\Big)^3\Bigg)\Bigg] \end{aligned}$$

As shown in section A.4.2, $(\mathbb{E}[\ln z])^2 = O\left(\mathbb{E}[||(\ln z, \ln \varepsilon^\mu)||^4]\right)$ and therefore all terms from line three onward go into the residual. The expression becomes:

$$\begin{aligned} \frac{1}{\mu_k} = \alpha\Big[&1 - \mathbb{E}_k[\ln z] - \frac{1-2\theta}{2}\Big(Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu] + \mathbb{E}_k[(\ln \varepsilon^\mu)^2]\Big)\Big] \\ &+ O(\mathbb{E}_k[||(\ln z, \ln \varepsilon^\mu)||^3]) \end{aligned} \tag{118}$$

It remains to approximate $\ln \sum_{k=1}^{K} \psi_k \frac{1}{\mu_k}$ around the point $\frac{1}{\mu_k} = \alpha$ for all $k \in \{1, \ldots, K\}$:

$$\ln \sum_{k=1}^{K} \psi_k \frac{1}{\mu_k} = \ln \sum_{k=1}^{K} \psi_k \alpha + \frac{\sum_{k=1}^{K} \psi_k \left(\frac{1}{\mu_k} - \alpha\right)}{\sum_{k=1}^{K} \psi_k \alpha} - \frac{1}{2} \frac{\sum_{k=1}^{K} \psi_k^2 \left(\frac{1}{\mu_k} - \alpha\right)^2}{\left(\sum_{k=1}^{K} \psi_k \alpha\right)^2}$$
$$- \frac{1}{2} \frac{\sum_{k=1}^{K} \sum_{l \neq k} \psi_k \psi_l \left(\frac{1}{\mu_k} - \alpha\right)\left(\frac{1}{\mu_l} - \alpha\right)}{\left(\sum_{k=1}^{K} \psi_k \alpha\right)^2} + O\left(\left|\left|\frac{1}{\mu} - \alpha\right|\right|^3\right)$$

Using (118), rewrite the linear term from above as:

$$\frac{\sum_{k=1}^{K} \psi_k \left(\frac{1}{\mu_k} - \alpha\right)}{\sum_{k=1}^{K} \psi_k \alpha} = \sum_{k=1}^{K} \psi_k \Bigg(- \mathbb{E}_k[\ln z] - \frac{1-2\theta}{2}\Big(Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^{\mu}] + \mathbb{E}_k[(\ln \varepsilon^{\mu})^2]\Big)$$
$$+ O(\mathbb{E}_k[||(\ln z, \ln \varepsilon^{\mu})||^3])\Bigg)$$

For the arguments established in sections A.4.2 and A.4.3, quadratic and interaction terms go into the residual:

$$\frac{1}{2} \frac{\sum_{k=1}^{K} \psi_k^2 \left(\frac{1}{\mu_k} - \alpha\right)^2}{\left(\sum_{k=1}^{K} \psi_k \alpha\right)^2} = O\left(\sum_{k=1}^{K} \mathbb{E}_k[||(\ln z, \ln \varepsilon^{\mu})||^3]\right)$$
$$\frac{1}{2} \frac{\sum_{k=1}^{K} \sum_{l \neq k} \psi_k \psi_l \left(\frac{1}{\mu_k} - \alpha\right)\left(\frac{1}{\mu_l} - \alpha\right)}{\left(\sum_{k=1}^{K} \psi_k \alpha\right)^2} = O\left(\sum_{k=1}^{K} \mathbb{E}_k[||(\ln z, \ln \varepsilon^{\mu})||^3]\right)$$

and in addition:

$$O\left(\left|\left|\frac{1}{\mu} - \alpha\right|\right|^3\right) \leq O\left(\sum_{k=1}^{K} \mathbb{E}_k[||(\ln z, \ln \varepsilon^{\mu})||^3]\right)$$

Combining these results and following analogous steps for μ_k^f, we obtain:

$$\ln \sum_{k=1}^{K} \psi_k \frac{1}{\mu_k^f} - \ln \sum_{k=1}^{K} \psi_k \frac{1}{\mu_k} = \sum_{k=1}^{K} \psi_k \mathbb{E}_k[\ln z] + \frac{1-2\theta}{2} \sum_{k=1}^{K} \psi_k \Big(Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^{\mu}]\Big)$$
$$+ O\left(\sum_{k=1}^{K} \mathbb{E}_k[||(\ln z, \ln \varepsilon^{\mu})||^3]\right) \quad (119)$$

Plugging (114) and (119) this into (109) delivers:

$$\ln \frac{\Delta}{\Delta^f} = \frac{\theta}{2} \sum_{k=1}^{K} \psi_k \Big(Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^{\mu}]\Big) + O\left(\sum_{k=1}^{K} \mathbb{E}_k[||(\ln z, \ln \varepsilon^{\mu})||^3]\right) \quad (120)$$

which, together with (117) proves proposition 2 in the main text.

A.3 Recovering Productivity Loss from the Data

This section derives the results in propositions 3 and 6 in the main text. Given a value for elasticity of substitution θ, equation (120) shows that recovering productivity loss from the data amounts to estimating

$$Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu] \tag{121}$$

for each item and summing over all of them using expenditure weights. As a first step, note that we can condition (121) on inflation gap $\hat{\pi} = \pi_k - \pi^*$, price rigidity $\mathfrak{f}$ and shock processes $\mathfrak{g}$ and $\mathfrak{m}$ using the between-within variance decomposition:

$$\begin{aligned} Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu] &= \mathbb{E}_k\big[Var_k(\ln z|\hat{\pi}, \mathfrak{f}, \mathfrak{g}, \mathfrak{m}) + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu|\hat{\pi}, \mathfrak{f}, \mathfrak{g}, \mathfrak{m}]\big] \\ &+ \underbrace{\mathbb{E}_k\big[\,(\mathbb{E}_k[\ln z|\hat{\pi}, \mathfrak{f}, \mathfrak{g}, \mathfrak{m}] - \mathbb{E}_k[\ln z])^2\,\big]}_{O(\mathbb{E}_k[||(\ln z, \ln \varepsilon^\mu)||^4])} \end{aligned}$$

where the second line falls into the residual, as shown in section A.4.2.

Suppose now that a researcher observes quality-adjusted relative price $p_{jkt} = P_{jkt}/(Q_{jkt}P_{kt})$:

$$\begin{aligned} \ln p_{jkt} &= \ln p^*_{jkt} + \ln z_{jkt} \\ &= \ln p^{*,0}_{jk} - s_{jkt}\pi^*_{jk} + \ln \varepsilon^\mu_{jkt} + \ln z_{jkt} \end{aligned}$$

Where $\pi^*_{jk} = -(\ln \overline{G}_{jk} + \ln \overline{\mu}^f_{jk})$ is the negative of the trend of flexible relative price of product j in item k, and $\ln p^{*,0}_{jk} = \ln G^0_{jk} + \ln \mu^{f,0}_{jk} - \ln Q^0_{jk} + \ln(w/p_k)$ is the flex-price intercept drawn at the time of entry. A linear trend estimate recovers the optimal inflation rate π^*_{jk} and therefore inflation gap $\hat{\pi}_{jk}$. Residuals of the linear regression take the form:

$$u_{jkt} = \ln \varepsilon^\mu_{jkt} + \big(\ln z_{jkt} - \mathbb{E}_{jk}[\ln z_{jkt}|p^{*,0}_{jk}, \pi^*_{jk}]\big)$$

Their cross-time variance for product j in sector k is then given by:

$$Var^{time}_{jk}(u) = Var^{time}_{jk}(\ln z) + 2\mathbb{E}^{time}_{jk}[\ln z \ln \varepsilon^\mu] + Var^{time}_{jk}(\ln \varepsilon^\mu)$$

Note that since distribution of price gaps $\ln z_{jkt}$ only depends on frictions $\mathfrak{f}_{jk}$, shock processes $\mathfrak{m}_{jk}$ and $\mathfrak{g}_{jk}$, and inflation gap $\hat{\pi}_{jk}$,[54] ergodicity implies that across-time moments of product j in item k are equal to cross-sectional moments in item k conditional on $\mathfrak{f}_{jk}$, $\mathfrak{m}_{jk}$, $\mathfrak{g}_{jk}$ and $\hat{\pi}_{jk}$:

$$\begin{aligned} Var^{time}_{jk}(u) &= Var_k(u|\hat{\pi}_{jk}, \mathfrak{f}_{jk}, \mathfrak{g}_{jk}, \mathfrak{m}_{jk}) \\ &= Var_k(\ln z|\hat{\pi}_{jk}, \mathfrak{f}_{jk}, \mathfrak{g}_{jk}, \mathfrak{m}_{jk}) + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu|\hat{\pi}_{jk}, \mathfrak{f}_{jk}, \mathfrak{g}_{jk}, \mathfrak{m}_{jk}] \\ &+ Var_k(\ln \varepsilon^\mu|\mathfrak{m}_{jk}) \end{aligned}$$

[54] See section A.5 for more details.

For convenience, define:

$$F_k(\hat{\pi}, \mathfrak{f}, \mathfrak{g}, \mathfrak{m}) = Var_k(\ln z|\hat{\pi}, \mathfrak{f}, \mathfrak{g}, \mathfrak{m}) + 2\mathbb{E}_k[\ln z \ln \varepsilon^{\mu}|\hat{\pi}, \mathfrak{f}, \mathfrak{g}, \mathfrak{m}]$$

which is identified up to a constant (in $\hat{\pi}$) by the cross-time variance of residuals of quality-adjusted relative prices $Var_{jk}^{time}(u)$. A Taylor expansion of $F_k(\cdot)$ in $\hat{\pi}$ yields:

$$F_k(\hat{\pi}, \mathfrak{f}, \mathfrak{g}, \mathfrak{m}) = F_k(0, \mathfrak{f}, \mathfrak{g}, \mathfrak{m}) + \sum_{n=1}^{N} \frac{\partial^n F_k(0, \mathfrak{f}, \mathfrak{g}, \mathfrak{m})}{(\partial \hat{\pi})^n} \frac{1}{n!} \hat{\pi}^n + O(\hat{\pi}^{N+1})$$

such that, for $N \geq 2$, unconditional moments become:

$$\begin{aligned} Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^{\mu}] &= \mathbb{E}_k\big[F_k(\hat{\pi}, \mathfrak{f}, \mathfrak{g}, \mathfrak{m})\big] + O\left(\mathbb{E}_k[||(\ln z, \ln \varepsilon^{\mu})||^4]\right) \\ &= \mathbb{E}_k\left[F_k(0, \mathfrak{f}, \mathfrak{g}, \mathfrak{m}) + \sum_{n=1}^{N} \frac{\partial^n F_k(0, \mathfrak{f}, \mathfrak{g}, \mathfrak{m})}{(\partial \hat{\pi})^n} \frac{1}{n!} \hat{\pi}^n\right] \\ &+ O(\mathbb{E}_k[||(\ln z, \ln \varepsilon^{\mu}, \hat{\pi})||^3]) \end{aligned}$$

Assumption 2.2 ensures that coefficients in the Taylor approximation of $F_k(\cdot)$ are independent of $\hat{\pi}$ and therefore:

$$\begin{aligned} Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^{\mu}] = \mathbb{E}_k\left[F_k(0, \mathfrak{f}, \mathfrak{g}, \mathfrak{m})\right] + \sum_{n=1}^{N} \mathbb{E}_k\left[\frac{\partial^n F_k(0, \mathfrak{f}, \mathfrak{g}, \mathfrak{m})}{(\partial \hat{\pi})^n} \frac{1}{n!}\right] \mathbb{E}_k\left[\hat{\pi}^n\right] \\ + O(\mathbb{E}_k[||(\ln z, \ln \varepsilon^{\mu}, \hat{\pi})||^3]) \end{aligned}$$

which proves proposition 3 in the main text, written as a special case of the above equation with $N = 2$. The coefficients $\mathbb{E}_k\left[\frac{\partial^n F_k(0,\mathfrak{f},\mathfrak{g},\mathfrak{m})}{(\partial \hat{\pi})^n} \frac{1}{n!}\right]$ can then be recovered from a cross-sectional regression of residual across-time variances $Var_{jk}^{time}(u)$ on inflation gaps $\hat{\pi}_{jk}$. With this, the aggregate productivity loss can be written as:

$$\begin{aligned} \ln \frac{\Delta}{\Delta^f} = \frac{\theta}{2}\left(\sum_{k=1}^{K} \psi_k \mathbb{E}_k\left[F_k(0, \mathfrak{f}, \mathfrak{g}, \mathfrak{m})\right] + \sum_{k=1}^{K} \psi_k \sum_{n=1}^{N} \mathbb{E}_k\left[\frac{\partial^n F_k(0, \mathfrak{f}, \mathfrak{g}, \mathfrak{m})}{(\partial \hat{\pi})^n} \frac{1}{n!}\right] \mathbb{E}_k\left[\hat{\pi}^n\right]\right) \\ + O\left(\sum_{k=1}^{K} \mathbb{E}_k[||(\ln z, \ln \varepsilon^{\mu}, \hat{\pi})||^3]\right) \end{aligned} \tag{122}$$

and the sufficient statistics from the main text are given by:

$$\begin{aligned} a_k &= \mathbb{E}_k\left[F_k(0, \mathfrak{f}, \mathfrak{g}, \mathfrak{m})\right] \\ b_k &= \mathbb{E}_k\left[\frac{\partial F_k(0, \mathfrak{f}, \mathfrak{g}, \mathfrak{m})}{\partial \hat{\pi}}\right] \\ c_k &= \frac{1}{2}\mathbb{E}_k\left[\frac{\partial^2 F_k(0, \mathfrak{f}, \mathfrak{g}, \mathfrak{m})}{(\partial \hat{\pi})^2}\right] \end{aligned}$$

A.4 Auxiliary Results

A.4.1 Expansion Method

Suppose we are interested in expansion of $\ln \mathbb{E}[f(x)]$ around the $x = 0$ point, where x is an N-dimensional vector of random variables and $f : \mathbb{R}^N \to \mathbb{R}$ such that $f(0) = 1$. We will first approximate $f(x)$ around $x = 0$, then take expectations, and finally approximate the logarithm. First, expand $f(x)$:

$$f(x) = 1 + \nabla f(0) \cdot x + \frac{1}{2} x' H_f(0) x + h(x)$$

where $\nabla f(0)$ and $H_f(0)$ are the gradient and Hessian evaluated at $x = 0$, and $h(x) = O(||x||^3)$. Second, take expectations:

$$\mathbb{E}[f(x)] = 1 + \nabla f(0) \cdot \mathbb{E}[x] + \frac{1}{2} trace(H_f(0)\Sigma_x) + \mathbb{E}[h(x)]$$

where Σ_x is the matrix of second moments of x (its element (i,j) is $\mathbb{E}[x_i x_j]$). Since $|h(x)| \leq C||x||^3$ for some $C > 0$ and x sufficiently small ($\forall x : ||x|| < \delta$ for some $\delta > 0$), it follows from triangle inequality that:

$$|\mathbb{E}[h(x)]| \leq \mathbb{E}[|h(x)|] \leq C\mathbb{E}[||x||^3]$$

and therefore $\mathbb{E}[h(x)] = O(\mathbb{E}[||x||^3])$. Third, approximate $\ln \mathbb{E}[f(x)]$ around $\mathbb{E}[f(x)] = 1$ as:

$$\begin{aligned}
\ln \mathbb{E}[f(x)] &= \mathbb{E}[f(x)] - 1 - \frac{1}{2}\left(\mathbb{E}[f(x)] - 1\right)^2 + O((\mathbb{E}[f(x)] - 1)^3) \\
&= \nabla f(0) \cdot \mathbb{E}[x] + \frac{1}{2} trace(H_f(0)\Sigma_x) + \mathbb{E}[h(x)] \\
&- \frac{1}{2}\left(\nabla f(0) \cdot \mathbb{E}[x] + \frac{1}{2} trace(H_f(0)\Sigma_x) + \mathbb{E}[h(x)]\right)^2 \\
&+ O\left(\left(\nabla f(0) \cdot \mathbb{E}[x] + \frac{1}{2} trace(H_f(0)\Sigma_x) + \mathbb{E}[h(x)]\right)^3\right) \\
&= \nabla f(0) \cdot \mathbb{E}[x] + \frac{1}{2} trace(H_f(0)\Sigma_x) - \frac{1}{2}\left(\nabla f(0) \cdot \mathbb{E}[x]\right)^2 + O(\mathbb{E}[||x||^3])
\end{aligned}$$

The result requires that all the terms appearing in the third and fourth lines of the above equation are at least of order of $\mathbb{E}[||x||^3]$, except for $(\mathbb{E}[x])^2$. To see this, first note that $\mathbb{E}[x] = O(\mathbb{E}[||x||])$ and $\Sigma_x = O(E[||x||^2])$. Therefore, expanding the third and fourth lines, one would obtain a term with $(\mathbb{E}[x])^2$ and terms of the form:

$$O(\mathbb{E}[||x||^r] \cdot \mathbb{E}[||x||^s])$$

with pairs $\{r, s\}$ such that $r \geq 1$, $s \geq 1$ and $r + s >= 3$. By Hölder's inequality, for $q \in \{r, s\}$:

$$\mathbb{E}[||x||^q] \leq \left(\mathbb{E}[||x||^{\frac{q}{p}}]\right)^p$$

for any $p \in (0,1)$. Set $p = \frac{q}{r+s}$, then:

$$\mathbb{E}[||x||^r] \cdot \mathbb{E}[||x||^s] \leq \left(\mathbb{E}[||x||^{r+s}]\right)^{\frac{r}{r+s}} \cdot \left(\mathbb{E}[||x||^{r+s}]\right)^{\frac{s}{r+s}} = \mathbb{E}[||x||^{r+s}]$$

Since $r + s \geq 3$, all of these terms fall into the $O(\mathbb{E}[||x||^3])$ residual. Therefore:

$$\ln \mathbb{E}[f(x)] = \nabla f(0) \cdot \mathbb{E}[x] + \frac{1}{2} trace(H_f(0)\Sigma_x) - \frac{1}{2} \left(\nabla f(0) \cdot \mathbb{E}[x]\right)^2 + O(\mathbb{E}[||x||^3])$$

In the following, whenever we write $O(f_1(x)) \leq O(f_2(x))$ we mean that $h(x) = O(f_1(x))$ implies $h(x) = O(f_2(x))$.

Application to our setting. In our setting, $x = [\ln z, \ln \varepsilon^\mu]'$ is the vector of two random variables, distributed according to the sticky-price equilibrium joint distribution. Recall:

$$f(\ln z, \ln \varepsilon^\mu) = e^{(1-\theta)(\ln z + \ln \varepsilon^\mu)}$$
$$g(\ln z, \ln \varepsilon^\mu) = e^{-\theta(\ln z + \ln \varepsilon^\mu)}$$

In the following we use the short-hand notation $f(x)$ and $g(x)$. Then for $\ln \mathbb{E}[f(x)]$:

$$\mathbb{E}\left[f(0)\right] = 1$$
$$\nabla f(0) = (1-\theta)[1,1]$$
$$H_f(0) = (1-\theta)^2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$
$$\nabla f(0) \cdot \mathbb{E}[x] = (1-\theta)\mathbb{E}[\ln z]$$
$$trace(H_f(0)\Sigma_x) = (1-\theta)^2 \Big(\mathbb{E}[(\ln z)^2] + \mathbb{E}[(\ln \varepsilon^\mu)^2] + 2\mathbb{E}[\ln z \ln \varepsilon^\mu]\Big)$$

For $\ln \mathbb{E}[g(x)]$:

$$\mathbb{E}\left[g(0)\right] = 1$$
$$\nabla g(0) = -\theta[1,1]$$
$$H_g(0) = \theta^2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$
$$\nabla g(0) \cdot \mathbb{E}[x] = -\theta\mathbb{E}[\ln z]$$
$$trace(H_g(0)\Sigma_x) = \theta^2 \Big(\mathbb{E}[(\ln z)^2] + \mathbb{E}[(\ln \varepsilon^\mu)^2] + 2\mathbb{E}[\ln z \ln \varepsilon^\mu]\Big)$$

Altogether:

$$\ln \mathbb{E}\left[f(\ln z, \ln \varepsilon^\mu)\right] = (1-\theta)\mathbb{E}[\ln z] + \frac{(1-\theta)^2}{2}\Big(Var(\ln z) + \mathbb{E}[(\ln \varepsilon^\mu)^2] + 2\mathbb{E}[\ln z \ln \varepsilon^\mu]\Big)$$
$$+ O(\mathbb{E}[||(\ln z, \ln \varepsilon^\mu)||^3])$$
$$\ln \mathbb{E}\left[g(\ln z, \ln \varepsilon^\mu\right] = -\theta\mathbb{E}[\ln z] + \frac{\theta^2}{2}\Big(Var(\ln z) + \mathbb{E}[(\ln \varepsilon^\mu)^2] + 2\mathbb{E}[\ln z \ln \varepsilon^\mu]\Big)$$
$$+ O(\mathbb{E}[||(\ln z, \ln \varepsilon^\mu)||^3])$$

A.4.2 Optimal Reset Price Implications

Lemma 4 *In the no-discounting limit and due to ergodicity, the firm's optimal reset price condition* (100) *implies:*

$$\mathbb{E}_k\left[e^{-\theta\ln\varepsilon^{\mu}+(1-\theta)\ln z}\right]=\mathbb{E}_k\left[e^{-\theta(\ln\varepsilon^{\mu}+\ln z)}\right] \tag{123}$$

The proof is provided in section A.4.4. Define:

$$\ln\tilde{g}(\ln z,\ln\varepsilon^{\mu})=(1-\theta)\ln z-\theta\ln\varepsilon^{\mu}$$

Then, expanding $\ln\mathbb{E}[\tilde{g}(\ln z,\ln\varepsilon^{\mu})]$ we obtain:

$$\mathbb{E}\left[\tilde{g}(0)\right]=1$$
$$\nabla\tilde{g}(0)=[1-\theta,-\theta]$$
$$\nabla\tilde{g}(0)\cdot\mathbb{E}[x]=(1-\theta)\mathbb{E}[\ln z]$$

such that:

$$\ln\mathbb{E}\left[\tilde{g}(x)\right]=(1-\theta)\mathbb{E}[\ln z]+O(\mathbb{E}[||(\ln z,\ln\varepsilon^{\mu})||^2])$$

A first order expansion of (123) yields:

$$\mathbb{E}_k[\ln z]=O\left(\mathbb{E}_k[||(\ln z,\ln\varepsilon^{\mu})||^2]\right)$$

And therefore it follows from Hölder's inequality that:

$$(\mathbb{E}_k[\ln z])^2=O\left(\mathbb{E}_k[||(\ln z,\ln\varepsilon^{\mu})||^4]\right)$$

Note that since firm's FOC holds at individual level, the above result applies also to conditional expectations of $\ln z$:

$$(\mathbb{E}_k[\ln z|\hat{\pi},\mathfrak{f},\mathfrak{g},\mathfrak{m}])^2=O\left(\mathbb{E}_k[||(\ln z,\ln\varepsilon^{\mu})||^4|\hat{\pi},\mathfrak{f},\mathfrak{g},\mathfrak{m}]\right)$$

A.4.3 Multi-sector residuals

First, note that since $\frac{1}{\mu_k}-\alpha=O(\mathbb{E}_k[||x||])$ with $x=(\ln z,\ln\varepsilon^{\mu})$, it follows that:

$$O\left(\left|\left|\frac{1}{\mu}-\alpha\right|\right|^3\right)\leq O\left(||(\mathbb{E}[||x||])||^3\right)$$

where $(\mathbb{E}[||x||])$ is a vector $(\mathbb{E}_1[||x||],\ldots,\mathbb{E}_k[||x||])$. Using Hölder's inequality, we can write:

$$\begin{aligned}\frac{1}{\sqrt{K}}||(\mathbb{E}[||x||])||&=\left(\frac{1}{K}\sum_{k=1}^{K}(\mathbb{E}_k[||x||])^2\right)^{\frac{1}{2}}\\&\leq\left(\frac{1}{K}\sum_{k=1}^{K}(\mathbb{E}_k[||x||])^{2\cdot\frac{3}{2}}\right)^{\frac{1}{2}\cdot\frac{2}{3}}=\left(\frac{1}{K}\sum_{k=1}^{K}(\mathbb{E}_k[||x||])^3\right)^{\frac{1}{3}}\end{aligned}$$

It follows that:

$$||(\mathbb{E}[||x||])||^3 \leq \sqrt{K}\sum_{k=1}^{K}(\mathbb{E}_k[||x||])^3 \leq \sqrt{K}\sum_{k=1}^{K}\mathbb{E}_k[||x||^3]$$

implying:

$$||(\mathbb{E}[||x||])||^3 = O\left(\sum_{k=1}^{K}\mathbb{E}_k[||x||^3]\right)$$

and therefore:

$$O\left(\left|\left|\frac{1}{\mu}-\alpha\right|\right|^3\right) \leq O\left(\sum_{k=1}^{K}\mathbb{E}_k[||x||^3]\right)$$

Second, in the cross-terms between sectors, all terms will be of the form:

$$O(\mathbb{E}_k[||x||^r]\cdot\mathbb{E}_l[||x||^s])$$

with pairs $\{r,s\}$ such that $r \geq 2$, $s \geq 2$. For pairs such that $\max\{r,s\} \geq 3$, it follows immediately that $O(\mathbb{E}_k[||x||^r]\cdot\mathbb{E}_l[||x||^s]) \leq O(\mathbb{E}_k[||x||^3]+\mathbb{E}_l[||x||^3])$. For pairs such that $\max\{r,s\} < 3$, apply Hölder's inequality:

$$\begin{aligned}\mathbb{E}_k[||x||^r]\cdot\mathbb{E}_l[||x||^s] &\leq (\mathbb{E}_k[||x||^{r\frac{3}{r}}])^{\frac{r}{3}}\cdot(\mathbb{E}_l[||x||^{s\frac{3}{s}}])^{\frac{s}{3}} \\ &= (\mathbb{E}_k[||x||^3])^{\frac{r}{3}}\cdot(\mathbb{E}_l[||x||^3])^{\frac{s}{3}} \leq \max\{(\mathbb{E}_k[||x||^3])^{\frac{r+s}{3}},(\mathbb{E}_l[||x||^3])^{\frac{r+s}{3}}\} \\ &= O\left(\mathbb{E}_k[||x||^3]+\mathbb{E}_l[||x||^3]\right)\end{aligned}$$

Since all pairs have $r+s \geq 4$. Therefore all the cross-terms fall into the $O\left(\sum_{k=1}^{K}\mathbb{E}_k[||x||^3]\right)$ residual.

A.4.4 Additional Lemmas and Proofs

Proof of Proposition 8.

Necessity. Suppose an economy is on a BGP with $L_t = L$, $L_{kt} = L_k$, $Y_{t+1} = \gamma Y_t$ and $Y_{kt+1} = \gamma_k Y_{kt}$. Households' FOC (77) and aggregate market clearing $C_t = Y_t$ imply that $W_t/(P_tY_t)$ is constant over time. Combining (58) and (71), we obtain:

$$\frac{1}{\mu_t} = \frac{W_t}{P_tY_t}L$$

which implies that aggregate markup $\mu_t = \mu$ is constant on the BGP. Equation (72):

$$\frac{L_k}{L} = \phi_k = \psi_k\frac{\mu}{\mu_{kt}}$$

implies that sector-level markups $\mu_{kt} = \mu_k$ are constant as well. With aggregate and item-level output growing at the rates of exogenous productivity, equations (55) and (58):

$$Y_{kt} = \frac{A_{kt}}{\Delta_{kt}}L_k \qquad Y_t = \frac{A_t}{\Delta_t}L$$

imply that endogenous productivity is constant both at the item and aggregate level: $\Delta_{kt} = \Delta_k$, $\Delta_t = \Delta$. Combining (84) and (86), we obtain:

$$\mu_{kt}\Delta_{kt} = \left(\frac{w_t}{p_{kt}}\right)^{-1} \tag{124}$$

Equation (124) then implies that w_t/p_{kt} is constant over time and therefore from (83) and (86):

$$\begin{aligned} 1 &= \left(\frac{w_t}{p_{kt}}\right)^{1-\theta} \int_0^1 \left(\mu^f_{jkt}\frac{G_{jkt}}{Q_{jkt}}z_{jkt}\right)^{1-\theta} \mathrm{dj} \\ (\mu_k)^{-1} &= \left(\frac{w_t}{p_{kt}}\right)^{1-\theta} \int_0^1 \left(\frac{G_{jkt}}{Q_{jkt}}\right)^{1-\theta} \left(\mu^f_{jkt}z_{jkt}\right)^{-\theta} \mathrm{dj}\ . \end{aligned}$$

it follows that the two integrals are constant over time. In addition, since w_t/p_{kt} is constant for all $k \in \{1, \ldots, K\}$, and using (42) and (56), we obtain:

$$w_t = \frac{W_t}{P_t A_t} = \frac{\prod_{k=1}^K (W_t)^{\psi_k}}{\prod\limits_{k=1}^K \left(\frac{P_{kt}}{\psi_k}\right)^{\psi_k} \prod\limits_{k=1}^K (\psi_k A_{kt})^{\psi_k}} = \prod_{k=1}^K \left(\frac{W_t}{P_{kt}A_{kt}}\right)^{\psi_k} = \prod_{k=1}^K \left(\frac{w_t}{p_{kt}}\right)^{\psi_k} \tag{125}$$

such that $w_t = w$ is constant over time and therefore $p_{kt} = p_k$ is constant as well for all k.

Sufficiency. Suppose the two integrals are constant over time for each item $k \in \{1, \ldots, K\}$. Then from (83), (84) and (86) it follows that Δ_{kt}, w_t/p_{kt} and μ_{kt} are all constant over time for each k. From (125) it follows that w_t and p_{kt} are constant for all k as well. Using equations (69) and (71):

$$\begin{aligned} \mu_t &= \left(\sum_{k=1}^K \psi_k \frac{1}{\mu_k}\right)^{-1} \\ \frac{1}{\mu_t} &= \frac{W_t}{P_t A_t}\Delta_t = w\Delta_t \end{aligned}$$

implies that $\mu_t = \mu$ and $\Delta_t = \Delta$ are constant over time. Combining households' FOC (77), aggregate market clearing ($C_t = Y_t$) and (58) yields:

$$\frac{V'(L_t)}{V(L_t)} = -\frac{W_t}{P_t Y_t} = -\frac{W_t \Delta_t}{P_t A_t L_t} = -\frac{w\Delta}{L}$$

which implies that aggregate labor is constant. From (72) it follows that item-level labor input shares and therefore labor inputs are constant. Finally, (55) and (58) provide that item-level and aggregate output grow at corresponding rates of exogenous productivity growth, concluding the proof. ■

Proof of Lemma 2. Recall from equations (85) and (87) that in the stochastic flex-price economy:

$$\psi_{jkt} = \left(\frac{w_t}{p_{kt}}\right)^{1-\theta} \left(\frac{G_{jk}^0 \overline{G}_{jk}^{s_{jkt}}}{Q_{jk}^0} \mu_{jk}^{f,0} (\overline{\mu}_{jk}^f)^{s_{jkt}} \varepsilon_{jkt}^{\mu}\right)^{1-\theta}$$

$$\phi_{jkt} = \frac{1}{\Delta_{kt}} \left(\frac{w_t}{p_{kt}}\right)^{-\theta} \left(\frac{G_{jk}^0 \overline{G}_{jk}^{s_{jkt}}}{Q_{jk}^0}\right)^{1-\theta} \left(\mu_{jk}^{f,0} (\overline{\mu}_{jk}^f)^{s_{jkt}} \varepsilon_{jkt}^{\mu}\right)^{-\theta}$$

Recall that assumption 2.1 imposes that deterministic flex-price expenditure weights and labor input shares are mean independent from markup shock processes, which can be expressed as:

$$\mathbb{E}_{kt}\left[\left(\frac{G_{jk}^0 \overline{G}_{jk}^{s_{jkt}}}{Q_{jk}^0} \mu_{jk}^{f,0} (\overline{\mu}_{jk}^f)^{s_{jkt}}\right)^{1-\theta} \middle| \mathfrak{m}_{jk} = \mathfrak{m}\right] = \mathbb{E}_{kt}\left[\left(\frac{G_{jk}^0 \overline{G}_{jk}^{s_{jkt}}}{Q_{jk}^0} \mu_{jk}^{f,0} (\overline{\mu}_{jk}^f)^{s_{jkt}}\right)^{1-\theta}\right]$$

$$\mathbb{E}_{kt}\left[\left(\frac{G_{jk}^0 \overline{G}_{jk}^{s_{jkt}}}{Q_{jk}^0}\right)^{1-\theta} \left(\mu_{jk}^{f,0} (\overline{\mu}_{jk}^f)^{s_{jkt}}\right)^{-\theta} \middle| \mathfrak{m}_{jk} = \mathfrak{m}\right] = \mathbb{E}_{kt}\left[\left(\frac{G_{jk}^0 \overline{G}_{jk}^{s_{jkt}}}{Q_{jk}^0}\right)^{1-\theta} \left(\mu_{jk}^{f,0} (\overline{\mu}_{jk}^f)^{s_{jkt}}\right)^{-\theta}\right]$$

This implies that in the stochastic flex-price economy:

$$\begin{aligned}\mathbb{E}_{kt}\left[\left(\frac{G_{jk}^0 \overline{G}_{jk}^{s_{jkt}}}{Q_{jk}^0} \mu_{jk}^{f,0} (\overline{\mu}_{jk}^f)^{s_{jkt}} \varepsilon_{jkt}^{\mu}\right)^{1-\theta}\right] &= \mathbb{E}_{kt}\left[\mathbb{E}_{kt}\left[\left(\frac{G_{jk}^0 \overline{G}_{jk}^{s_{jkt}}}{Q_{jk}^0} \mu_{jk}^{f,0} (\overline{\mu}_{jk}^f)^{s_{jkt}} \varepsilon_{jkt}^{\mu}\right)^{1-\theta} \middle| \mathfrak{m}_{jk} = \mathfrak{m}\right]\right] \\ &= \mathbb{E}_{kt}\left[\left(\frac{G_{jk}^0 \overline{G}_{jk}^{s_{jkt}}}{Q_{jk}^0} \mu_{jk}^{f,0} (\overline{\mu}_{jk}^f)^{s_{jkt}}\right)^{1-\theta}\right] \mathbb{E}_{kt}\left[\left(\varepsilon_{jkt}^{\mu}\right)^{1-\theta}\right]\end{aligned}$$

because markup shock realizations are independent from intercepts and trends, given $\mathfrak{m}$. Since the first expectation in the last line is constant over time by the BGP condition 1 and the second expectation is constant over time because of the time-invariant distribution of markup shock processes, it follows that

$$\int_0^1 \left(\frac{G_{jk}^0 \overline{G}_{jk}^{s_{jkt}}}{Q_{jk}^0} \mu_{jk}^{f,0} (\overline{\mu}_{jk}^f)^{s_{jkt}} \varepsilon_{jkt}^{\mu}\right)^{1-\theta} \mathrm{dj} \tag{126}$$

is constant over time for each item k. The same logic ensures that

$$\int_0^1 \left(\frac{G_{jk}^0 \overline{G}_{jk}^{s_{jkt}}}{Q_{jk}^0}\right)^{1-\theta} \left(\mu_{jk}^{f,0} (\overline{\mu}_{jk}^f)^{s_{jkt}} \varepsilon_{jkt}^{\mu}\right)^{-\theta} \mathrm{dj} \tag{127}$$

is constant over time as well. This means that the necessary and sufficient condition for a BGP in a stochastic flex-price economy is satisfied, according to proposition 8. ■

Proof of Lemma 3. Note that from (85) and (87) it follows that expenditure and labor input shares can be written as:

$$\psi_{jkt} = \left(\frac{w}{p_k}\right)^{1-\theta} e^{c_{jk}^{f,0}+s_{jkt}c_{jk}^{f,1}} f(\ln z_{jkt}, \ln \varepsilon_{jkt}^{\mu})$$
$$\phi_{jkt} = \frac{1}{\Delta_k}\left(\frac{w}{p_k}\right)^{-\theta} e^{c_{jk}^{g,0}+s_{jkt}c_{jk}^{g,1}} g(\ln z_{jkt}, \ln \varepsilon_{jkt}^{\mu})$$

Since in a deterministic flex-price economy $\ln z_{jkt} = \ln \varepsilon_{jkt}^{\mu} = 0$, assumption 2 implies that $e^{c_{jk}^{f,0}+s_{jkt}c_{jk}^{f,1}}$ and $e^{c_{jk}^{g,0}+s_{jkt}c_{jk}^{g,1}}$ are mean independent from $\hat{\pi}_{jk}$, $\mathfrak{f}_{jk}$, $\mathfrak{g}_{jk}$ and $\mathfrak{m}_{jk}$. As discussed in section A.5, price gap distribution of product j depends only on these four objects, which makes price gaps $\ln z_{jkt}$ conditionally independent from $e^{c_{jk}^{f,0}+s_{jkt}c_{jk}^{f,1}}$ and $e^{c_{jk}^{g,0}+s_{jkt}c_{jk}^{g,1}}$, given those four objects. Therefore:

$$\begin{aligned}
\mathbb{E}_k\left[\left(G\mu^f z\right)^{1-\theta} Q^{\theta-1}\right] &= \mathbb{E}_k\left[e^{c^{f,0}+sc^{f,1}} f\left(\ln z, \ln \varepsilon^{\mu}\right)\right] \\
&= \mathbb{E}_k\left[\mathbb{E}_k\left[e^{c^{f,0}+sc^{f,1}} f\left(\ln z, \ln \varepsilon^{\mu}\right)\middle| z, \varepsilon^{\mu}, \hat{\pi}, \mathfrak{f}, \mathfrak{g}, \mathfrak{m}\right]\right] \\
&= \mathbb{E}_k\left[f\left(\ln z, \ln \varepsilon^{\mu}\right) \mathbb{E}_k\left[e^{c^{f,0}+sc^{f,1}}\middle| z, \varepsilon^{\mu}, \hat{\pi}, \mathfrak{f}, \mathfrak{g}, \mathfrak{m}\right]\right] \\
&= \mathbb{E}_k\left[e^{c^{f,0}+sc^{f,1}}\right] \mathbb{E}_k\left[f\left(\ln z, \ln \varepsilon^{\mu}\right)\right]
\end{aligned}$$

and a similar result holds for $\mathbb{E}\left[G^{1-\theta}(\mu^f z)^{-\theta} Q^{\theta-1}\right]$. Therefore:

$$\begin{aligned}
\ln \mathbb{E}_k\left[\left(G\mu^f z\right)^{1-\theta} Q^{\theta-1}\right] &= \ln \mathbb{E}_k\left[e^{c^{f,0}+sc^{f,1}}\right] + \ln \mathbb{E}_k\left[f\left(\ln z, \ln \varepsilon^{\mu}\right)\right] \\
\ln \mathbb{E}_k\left[G^{1-\theta}(\mu^f z)^{-\theta} Q^{\theta-1}\right] &= \ln \mathbb{E}_k\left[e^{c^{g,0}+sc^{g,1}}\right] + \ln \mathbb{E}_k\left[g\left(\ln z, \ln \varepsilon^{\mu}\right)\right]
\end{aligned}$$

Finally, note that from proposition 8 it follows that all expectations in the above equations are constant over time since both the deterministic flex-price economy and the stochastic flex-price economy are on the BGP. ■

Proof of Lemma 4. Recall the first-order condition with respect to the reset price (100) and consider the no-discounting limit:

$$\mathbb{E}_0\left[\sum_{t=0}^{T-1}\left(\varepsilon_{jkt}^{\mu}\right)^{-\theta} z_{jkt}^{1-\theta}\middle|\hat{\pi}_{jk}, \mathfrak{g}_{jk}, \mathfrak{m}_{jk}, \mathfrak{f}_{jk}\right] = \mathbb{E}_0\left[\sum_{t=0}^{T-1}\left(\varepsilon_{jkt}^{\mu} z_{jkt}\right)^{-\theta}\middle|\hat{\pi}_{jk}, \mathfrak{g}_{jk}, \mathfrak{m}_{jk}, \mathfrak{f}_{jk}\right] \quad (128)$$

where we denote the (stochastic) first stopping time by T and explicitly write expectations as conditional on reset-time information and all the idiosyncratic elements relevant for reset-price determination. Note that the above condition applies to any reset state and any product j and therefore also holds unconditionally within an item:

$$\mathbb{E}_0\left[\sum_{t=0}^{T-1}\left(\varepsilon_{jkt}^{\mu}\right)^{-\theta} z_{jkt}^{1-\theta}\right] = \mathbb{E}_0\left[\sum_{t=0}^{T-1}\left(\varepsilon_{jkt}^{\mu} z_{jkt}\right)^{-\theta}\right] \quad (129)$$

Ergodicity and Lemma 5 imply:

$$\mathbb{E}_k\left[\left(\varepsilon^{\mu}\right)^{-\theta} z^{1-\theta}\right] = \frac{\mathbb{E}_0\left[\sum_{t=0}^{T-1}\left(\varepsilon^{\mu}_{jkt}\right)^{-\theta} z_{jkt}^{1-\theta}\right]}{\mathbb{E}_0[T]}$$

$$\mathbb{E}_k\left[\left(\varepsilon^{\mu} z\right)^{-\theta}\right] = \frac{\mathbb{E}_0\left[\sum_{t=0}^{T-1}\left(\varepsilon^{\mu}_{jkt} z_{jkt}\right)^{-\theta}\right]}{\mathbb{E}_0[T]}$$

It follows that:

$$\mathbb{E}_k\left[\left(\varepsilon^{\mu}\right)^{-\theta} z^{1-\theta}\right] = \mathbb{E}_k\left[\left(\varepsilon^{\mu} z\right)^{-\theta}\right]$$

■

Lemma 5 *For an ergodic process x_t with a strong Markov property, it holds that:*

$$\mathbb{E}[x] = \frac{\mathbb{E}_0\left[\sum_{t=0}^{T_1-1} x_t\right]}{\mathbb{E}_0\left[T_1\right]}$$

where $\mathbb{E}[x]$ denotes cross-sectional expectation of x, $\mathbb{E}_0[\cdot]$ denotes reset-time expectation, and T_1 denotes the first stopping time.

Proof of Lemma 5. The following proof adapts the proof of Auxiliary Theorem 2 of Baley and Blanco (2021) to a discrete time setting. Let x_t be an ergodic process, T_i be the i-th adjustment time, $T_0 \equiv 0$, N_T be the number of adjustments before T. Then:

$$\mathbb{E}[x] = \lim_{T\to\infty} \frac{\sum_{t=0}^{T-1} x_t}{T} = \lim_{T\to\infty} \frac{\sum_{i=1}^{N_T}\sum_{t=T_{i-1}}^{T_i-1} x_t + \sum_{t=T_{N_T}}^{T-1} x_t}{\sum_{i=1}^{N_T}\left(T_i - T_{i-1}\right) + \left(T - T_{N_T}\right)}$$

$$= \lim_{T\to\infty} \frac{\frac{\sum_{i=1}^{N_T}\sum_{t=T_{i-1}}^{T_i-1} x_t}{N_T} + \frac{\sum_{t=T_{N_T}}^{T-1} x_t}{N_T}}{\frac{\sum_{i=1}^{N_T}(T_i - T_{i-1})}{N_T} + \frac{\left(T - T_{N_T}\right)}{N_T}}$$

Note that the last terms in the numerator and denominator vanish as T and therefore N_T go to infinity. The first two terms converge to expected values, yielding:

$$\mathbb{E}[x] = \frac{\mathbb{E}_0\left[\sum_{t=0}^{T_1-1} x_t\right]}{\mathbb{E}_0\left[T_1\right]}$$

where T_1 is the first stopping time and we can use unconditional expectations starting from $t = 0$ due to the strong Markov property of x_t. ■

Proof of Proposition 4. Consider the continuous-time limit of a discrete-time Calvo (1983) model by letting the length of time periods shrink to zero. Let productivity and markup shocks $\ln \varepsilon_t^{\mu} \sim ii\mathcal{N}(0, \sigma_{\mu}^2)$ and $\ln \varepsilon_t^{G} \sim ii\mathcal{N}(0, \sigma_G^2)$ and let λ denote the price adjustment intensity.[55] In the limit without discounting (assumption 3), the firm first-order condition (100) implies:

[55] See Moll (2014), Itskhoki and Moll (2019) and Holden (2025) for examples of modeling i.i.d. processes in continuous time.

$$\mathbb{E}\int_0^T e^{(1-\theta)\ln z_t - \theta \ln \varepsilon_t^\mu} dt = \mathbb{E}\int_0^T e^{-\theta(\ln z_t + \ln \varepsilon_t^\mu)} dt$$

with $\ln z_t = \ln z_0 - \hat{\pi}t - (\ln \varepsilon_t^\mu - \ln \varepsilon_0^\mu) - (\ln \varepsilon_t^G - \ln \varepsilon_0^G)$ in between adjustments. Using the distribution of adjustment times and collecting terms, one obtains:

$$\mathbb{E}\int_0^\infty e^{-\lambda t + (1-\theta)\ln z_t - \theta \ln \varepsilon_t^\mu} dt = \mathbb{E}\int_0^\infty e^{-\lambda t - \theta(\ln z_t + \ln \varepsilon_t^\mu)} dt$$

$$e^{\ln z_0 + \ln \varepsilon_0^\mu + \ln \varepsilon_0^G}\mathbb{E}\int_0^\infty e^{-(\lambda + (1-\theta)\hat{\pi})t - (1-\theta)\ln \varepsilon_t^G - \ln \varepsilon_t^\mu} dt = \mathbb{E}\int_0^\infty e^{-(\lambda - \theta\hat{\pi})t + \theta \ln \varepsilon_t^G} dt$$

such that optimal reset price gap is given by:

$$\ln z_0(\hat{\pi}, \ln \varepsilon_0^G, \ln \varepsilon_0^\mu) = \ln \widehat{\gamma}_2(\hat{\pi}, \ln \varepsilon_0^G) - \ln \widehat{\gamma}_1(\hat{\pi}, \ln \varepsilon_0^G, \ln \varepsilon_0^\mu) - \ln \varepsilon_0^\mu - \ln \varepsilon_0^G$$

$$\widehat{\gamma}_1(\hat{\pi}, \ln \varepsilon_0^G, \ln \varepsilon_0^\mu) = \mathbb{E}\left[\int_0^\infty e^{-(\lambda + (1-\theta)\hat{\pi})t - (1-\theta)\ln \varepsilon_t^G - \ln \varepsilon_t^\mu} dt \,\middle|\, \ln \varepsilon_0^G, \ln \varepsilon_0^\mu\right]$$

$$\widehat{\gamma}_2(\hat{\pi}, \ln \varepsilon_0^G) = \mathbb{E}\left[\int_0^\infty e^{-(\lambda - \theta\hat{\pi})t + \theta \ln \varepsilon_t^G} dt \,\middle|\, \ln \varepsilon_0^G\right]$$

Note that for all $t > 0$:

$$\mathbb{E}[e^{-\ln \varepsilon_t^\mu} | \ln \varepsilon_0^G, \ln \varepsilon_0^\mu] = e^{\frac{\sigma_\mu^2}{2}}$$

$$\mathbb{E}[e^{-(1-\theta)\ln \varepsilon_t^G} | \ln \varepsilon_0^G, \ln \varepsilon_0^\mu] = e^{(1-\theta)^2 \frac{\sigma_G^2}{2}}$$

$$\mathbb{E}[e^{\theta \ln \varepsilon_t^G} | \ln \varepsilon_0^G, \ln \varepsilon_0^\mu] = e^{\theta^2 \frac{\sigma_G^2}{2}}$$

such that:

$$\widehat{\gamma}_1(\hat{\pi}, \ln \varepsilon_0^G, \ln \varepsilon_0^\mu) = \frac{e^{(1-\theta)^2 \frac{\sigma_G^2}{2} + \frac{\sigma_\mu^2}{2}}}{\lambda + (1-\theta)\hat{\pi}}$$

$$\widehat{\gamma}_2(\hat{\pi}, \ln \varepsilon_0^G) = \frac{e^{\theta^2 \frac{\sigma_G^2}{2}}}{\lambda - \theta\hat{\pi}}$$

And therefore:

$$\ln z_0(\hat{\pi}, \ln \varepsilon_0^G, \ln \varepsilon_0^\mu) = \ln \frac{\lambda + (1-\theta)\hat{\pi}}{\lambda - \theta\hat{\pi}} + (2\theta - 1)\frac{\sigma_G^2}{2} - \frac{\sigma_\mu^2}{2} - \ln \varepsilon_0^\mu - \ln \varepsilon_0^G$$

For convenience, define $\ln \tilde{z}_0(\hat{\pi}) = \ln z_0(\hat{\pi}, \ln \varepsilon_0^G, \ln \varepsilon_0^\mu) + \ln \varepsilon_0^\mu + \ln \varepsilon_0^G$ such that in between adjustments:

$$\ln z_t = \ln \tilde{z}_0(\hat{\pi}) - \hat{\pi}t - \ln \varepsilon_t^\mu - \ln \varepsilon_t^G$$

Let $F(\hat{\pi}) = Var\left(\ln z|\hat{\pi}\right) + 2\mathbb{E}\left[\ln z \ln \varepsilon^{\mu}|\hat{\pi}\right]$ and consider the two terms separately:

$$\begin{aligned}
Var(\ln z|\hat{\pi}) &= \mathbb{E}[(\ln z)^2|\hat{\pi}] - (\mathbb{E}[\ln z|\hat{\pi}])^2 \\
&= \frac{\mathbb{E}\left[\int_0^T (\ln z_t)^2 dt \,\big|\hat{\pi}\right]}{\mathbb{E}[T]} - \left(\frac{\mathbb{E}\left[\int_0^T \ln z_t dt \,\big|\hat{\pi}\right]}{\mathbb{E}[T]}\right)^2 \\
&= \lambda \mathbb{E}\left[\int_0^\infty e^{-\lambda t} \left(\ln z_t\right)^2 dt \,\Big|\hat{\pi}\right] - \lambda^2 \left(\mathbb{E}\left[\int_0^\infty e^{-\lambda t} \ln z_t dt \,\Big|\hat{\pi}\right]\right)^2 \\
2\mathbb{E}[\ln z \ln \varepsilon^{\mu}|\hat{\pi}] &= 2\frac{\mathbb{E}\left[\int_0^T \ln z_t \ln \varepsilon_t^{\mu} dt \,\big|\hat{\pi}\right]}{\mathbb{E}[T]} \\
&= 2\lambda \mathbb{E}\left[\int_0^\infty e^{-\lambda t} \ln z_t \ln \varepsilon_t^{\mu} dt \,\Big|\hat{\pi}\right]
\end{aligned}$$

Note that due to i.i.d. properties of the shocks and independence of $\ln \tilde{z}_0$ from the reset-time shock values, conditional expectations given the reset-time shock values are identical to the unconditional ones. Then the moments become:

$$\begin{aligned}
\mathbb{E}\left[\int_0^\infty e^{-\lambda t} \ln z_t dt \,\Big|\hat{\pi}\right] &= \mathbb{E}\left[\int_0^\infty e^{-\lambda t} \left(\ln \tilde{z}_0(\hat{\pi}) - \hat{\pi} t - \ln \varepsilon_t^{\mu} - \ln \varepsilon_t^{G}\right) dt\right] \\
&= \frac{\ln \tilde{z}_0(\hat{\pi})}{\lambda} - \frac{\hat{\pi}}{\lambda^2} \\
\mathbb{E}\left[\int_0^\infty e^{-\lambda t} \left(\ln z_t\right)^2 dt \,\Big|\hat{\pi}\right] &= \mathbb{E}\left[\int_0^\infty e^{-\lambda t} \left(\ln \tilde{z}_0(\hat{\pi}) - \hat{\pi} t\right)^2 dt\right] \\
&\quad - 2\mathbb{E}\left[\int_0^\infty e^{-\lambda t} \left(\ln \tilde{z}_0(\hat{\pi}) - \hat{\pi} t\right)\left(\ln \varepsilon_t^{\mu} + \ln \varepsilon_t^{G}\right) dt\right] \\
&\quad + \mathbb{E}\left[\int_0^\infty e^{-\lambda t} \left(\ln \varepsilon_t^{\mu} + \ln \varepsilon_t^{G}\right)^2 dt\right] \\
&= \frac{\left(\ln \tilde{z}_0(\hat{\pi})\right)^2}{\lambda} - \frac{2\hat{\pi} \ln \tilde{z}_0(\hat{\pi})}{\lambda^2} + \frac{2\hat{\pi}^2}{\lambda^3} + \frac{\sigma_G^2 + \sigma_\mu^2}{\lambda} \\
\mathbb{E}\left[\int_0^\infty e^{-\lambda t} \ln z_t \ln \varepsilon_t^{\mu} dt \,\Big|\hat{\pi}\right] &= \mathbb{E}\left[\int_0^\infty e^{-\lambda t} \left(\ln \tilde{z}_0(\hat{\pi}) - \hat{\pi} t - \ln \varepsilon_t^{\mu} - \ln \varepsilon_t^{G}\right) \ln \varepsilon_t^{\mu} dt\right] \\
&= -\frac{\sigma_\mu^2}{\lambda}
\end{aligned}$$

Altogether:

$$\begin{aligned}
F(\hat{\pi}) &= \lambda \mathbb{E}\left[\int_0^\infty e^{-\lambda t} \left(\ln z_t\right)^2 dt \,\Big|\hat{\pi}\right] - \lambda^2 \left(\mathbb{E}\left[\int_0^\infty e^{-\lambda t} \ln z_t dt \,\Big|\hat{\pi}\right]\right)^2 \\
&\quad + 2\lambda \mathbb{E}\left[\int_0^\infty e^{-\lambda t} \ln z_t \ln \varepsilon_t^{\mu} dt \,\Big|\hat{\pi}\right] \\
&= \left(\sigma_G^2 - \sigma_\mu^2\right) + \frac{\hat{\pi}^2}{\lambda^2}
\end{aligned}$$

And therefore coefficients in equation (29) in the main text are $a_k = \sigma_G^2 - \sigma_\mu^2$, $b_k = 0$ and $c_k = \frac{1}{\lambda^2}$. ■

Proof of Proposition 7. Letting $\pi = \pi^{act} + d$ and $\pi_k = \pi_k^{act} + d$, we have from proposition 5:

$$\begin{aligned}
\mathcal{L}(\pi) &= \mathcal{L}(\pi^{act} + d) \\
&= \frac{\theta}{2} \sum_{k=1}^{K} \psi_k \left(a_k + b_k \int_0^1 \left(\pi_k^{act} - \pi_{jk}^* + d\right) \mathrm{dj} + c_k \int_0^1 \left(\pi_k^{act} - \pi_{jk}^* + d\right)^2 \mathrm{dj} \right) + O(3) \\
&= \frac{\theta}{2} \sum_{k=1}^{K} \psi_k \left(a_k + b_k \int_0^1 \left(\pi_k - \pi_{jk}^*\right) \mathrm{dj} + c_k \int_0^1 \left(\pi_k - \pi_{jk}^*\right)^2 \mathrm{dj} \right) + O(3) \\
&= \frac{\theta}{2} \sum_{k=1}^{K} \psi_k \left(a_k + b_k \int_0^1 \left(\pi + \ln \frac{\gamma}{\gamma_k} - \pi_{jk}^*\right) \mathrm{dj} + c_k \int_0^1 \left(\pi + \ln \frac{\gamma}{\gamma_k} - \pi_{jk}^*\right)^2 \mathrm{dj} \right) + O(3),
\end{aligned}$$

where the last line used equation (19). Rewriting the last expression in vertex form delivers

$$\mathcal{L}(\pi) = \overline{L} + \frac{\theta}{2} \left(\sum_{k=1}^{K} \psi_k c_k \right) \cdot (\pi - \pi^*)^2 + O(3) \tag{130}$$

where

$$\begin{aligned}
\pi^* &= -\frac{\sum_{k=1}^{K} \psi_k c_k \left(\ln \frac{\gamma}{\gamma_k} - \int_0^1 \pi_{jk}^* \, dj \right)}{\sum_{k=1}^{K} \psi_k c_k} - \frac{\sum_{k=1}^{K} \psi_k b_k}{2 \sum_{k=1}^{K} \psi_k c_k} && (131) \\
&= \pi^{act} - \frac{\sum_{k=1}^{K} \psi_k c_k \left(\pi_k^{act} - \int_0^1 \pi_{jk}^* \, dj \right)}{\sum_{k=1}^{K} \psi_k c_k} - \frac{\sum_{k=1}^{K} \psi_k b_k}{2 \sum_{k=1}^{K} \psi_k c_k}, && (132)
\end{aligned}$$

where the last line used equation (19), and where

$$\overline{L} = \frac{\theta}{2} \sum_{k=1}^{K} \psi_k \left(a_k + b_k \int_0^1 \left(\ln \frac{\gamma}{\gamma_k} - \pi_{jk}^* \right) dj + c_k \left(\int_0^1 \left(\ln \frac{\gamma}{\gamma_k} - \pi_{jk}^* \right)^2 dj - (\pi^*)^2 \right) \right) \tag{133}$$

with $\ln \frac{\gamma}{\gamma_k} = \pi_k^{act} - \pi^{act}$, again due to equation (19). ∎

A.5 Determinants of Price Gap Distribution

The distribution of price gaps $\ln z_{jkt}$ of firm j is determined by three objects: (i) distribution of adjustment times, conditional on histories of shocks, (ii) distribution of reset price gaps, conditional on adjustment times, and (iii) dynamics of price gaps between adjustments. In the following, we discuss what these three objects depend on, in reverse order.

Dynamics between adjustments. Since nominal prices are constant in between adjustments, price gaps follow the inverse of the flexible price:

$$\Delta \ln z_{jkt} \equiv \ln z_{jkt} - \ln z_{jkt-1} = -\hat{\pi}_{jk} - \Delta \ln \varepsilon_{jkt}^{G} - \Delta \ln \varepsilon_{jkt}^{\mu}$$

and therefore only depend on $\hat{\pi}_{jk}$, $\mathfrak{g}_{jk}$ and $\mathfrak{m}_{jk}$. Note that the only way aggregate variables can affect price gap dynamics in between adjustments is via changes in $\hat{\pi}_{jk}$ driven by changes in aggregate inflation rate: $\hat{\pi}_{jk} = \pi_k - \pi^*_{jk} = \pi + \ln\gamma - \ln\gamma_k - \pi^*_{jk}$.

Reset price gaps. By assumption 1.2, reset price gaps are determined by maximizing expected discounted profits until the next adjustment time and satisfy the following FOC in the no-discounting limit (see section A.1.8):

$$E_t \sum_{i=0}^{S_{jkh}-1} \left(\varepsilon^{\mu}_{jkt+i}\right)^{-\theta} z^{1-\theta}_{jkt+i} - \left(\varepsilon^{\mu}_{jkt+i} z_{jkt+i}\right)^{-\theta} = 0 \tag{134}$$

where S_{jkh} is the length of the h's price spell. Reset price gaps therefore depend on $\hat{\pi}_{jk}$, $\mathfrak{g}_{jk}$ and $\mathfrak{m}_{jk}$, and additionally on the distribution of adjustment times $\{t + S_{jkh}, t + S_{jkh} + S_{jkh+1}, ...\}$. Note that adjustment friction $\mathfrak{f}_{jk}$ does not directly affect reset price gap distribution, but only indirectly through the distribution of adjustment times. Note also that aggregate variables can affect the distribution of reset price gaps in two ways only: via (i) changes in $\hat{\pi}_{jk}$ driven by changes in aggregate inflation rate and (ii) the effects of aggregate variables on adjustment time distribution.

Adjustment times. By definition of $\mathfrak{f}_{jk}$, given the dynamics of aggregate and item-level variables, the distribution of adjustment times depends on adjustment friction $\mathfrak{f}_{jk}$ and the remaining idiosyncratic objects that determine reset price gaps and dynamics in between adjustments: $\hat{\pi}_{jk}$, $\mathfrak{m}_{jk}$ and $\mathfrak{g}_{jk}$. Assumption 1.3, in turn, ensures that any change in aggregate inflation affects the distribution of adjustment times (and therefore of price gaps) via its effects on the distribution of $\hat{\pi}_{jk}$-s only. As a result, for any two products i and j, belonging to the same item and with identical fundamentals ($\mathfrak{m}$, $\mathfrak{g}$ and $\mathfrak{f}$) except for $\hat{\pi}_{ik} = \hat{\pi}_{jk} - d$ for some $d \in \mathbb{R}$, price gap distribution of product j in an economy with aggregate inflation level π is identical to the price gap distribution of product i in an economy with adjusted aggregate inflation level $\pi + d$.

A.6 Admissible flex-price distributions

First, proposition 9 states that our setup allows for arbitrary distributions of flex-price intercepts, trends and shock processes within items, as long as they satisfy assumption 2 and are balanced growth path consistent. Second, we discuss the limitations on flex-price distributions, imposed by assumption 2 and the BGP requirement.

Proposition 9 *If quality-adjusted flexible relative price dynamics have the structure of* (12), *are consistent with balanced growth path according to condition 1, and their distri-*

butions within items satisfy assumption 2, then joint within-item distributions of quality-adjusted flexible relative price intercepts, trends and shock processes are otherwise unrestricted, for any collection of exit rates across items.

Proof of Proposition 9. Suppose we are given a distribution of exit rates $\{\delta_k\}_{k=1}^K$ and distributions of relative flexible quality-adjusted price intercepts, trends and shock processes, satisfying assumption 2 and balanced growth path condition 1. Recall that quality-adjusted flex-price intercepts, trends and shocks are given, respectively, by:

$$\ln \mu_{jk}^{f,0} + \ln G_{jk}^0 - \ln Q_{jk}^0 + \ln \frac{w}{p_k} \tag{135}$$

$$\ln \overline{\mu}_{jk}^f + \ln \overline{G}_{jk} \tag{136}$$

$$\ln \varepsilon_{jkt}^{\mu} \tag{137}$$

where we treat w/p_k as given. In the following we show that we can find a collection of distributions of fundamental intercepts, trends and shock processes that satisfies all of the assumptions of our setup and reproduces the given distribution of flex-price components, taking into account a given distribution of exit rates. There are two conditions that need to be verified: (i) no-discounting limit assumption 3 and (ii) assumption 4.

First, note that exit rate δ_k pins down the distribution of ages s, as well as a linear combination of productivity and markup trends that ensures the no-discounting limit:

$$c_{jk}^{g,1} = (1-\theta)\ln \overline{G}_{jk} - \theta \ln \overline{\mu}_{jk}^f \stackrel{!}{=} -\ln\left(\beta\gamma^{1-\sigma}(1-\delta_k)\right) \equiv c_k^{g,1}$$

Given the distribution of flex-price trends $\ln \overline{\mu}_{jk}^f + \ln \overline{G}_{jk}$, this uniquely pins down productivity and markup trends for each product.

Second, the distribution of quality-adjusted flex-price shock processes uniquely pins down the distribution of markup shock processes $\mathfrak{m}_{jk}$ and leaves the distribution of productivity/quality shock processes $\mathfrak{g}_{jk}$ unrestricted. Hence, it only remains to specify the distributions of fundamental intercepts. Assumption 4 requires that, given the inverse of the average item-level markup in the deterministic flex-price economy α, at the item level it must hold that:

$$\frac{\mathbb{E}_k\left[e^{c^{g,0}+sc^{g,1}}\right]}{\mathbb{E}_k\left[e^{c^{f,0}+sc^{f,1}}\right]} = \frac{\mathbb{E}_k\left[e^{c^{g,0}+sc^{g,1}}\right]}{\mathbb{E}_k\left[e^{(1-\theta)(\ln G^0 - \ln Q^0 + \ln \mu^{f,0}) + s(1-\theta)\left(\ln \overline{\mu}^f + \ln \overline{G}\right)}\right]} \stackrel{!}{=} \alpha$$

Note that the denominator is pinned down by the distribution of flex-price intercepts and trends, whereas the distribution of ages s, as well as coefficient $c^{g,1}$, are both pinned down by exit rate δ_k. This puts a restriction on the distribution of $c_{jk}^{g,0}$ within an item. Let

$c^{g,0}_{jk}$ be constant across all products within an item and equal to the value that satisfies assumption 4:

$$c^{g,0}_{jk} = (1-\theta)(\ln G^0_{jk} - \ln Q^0_{jk}) - \theta \ln \mu^{f,0}_{jk} \overset{!}{=} \ln \left(\frac{\alpha \mathbb{E}_k \left[e^{c^{f,0}+sc^{f,1}} \right]}{\mathbb{E}_k \left[e^{sc^{g,1}} \right]} \right) \tag{138}$$

Altogether, given the premise of the proposition, intercepts of productivity, quality and desired markups have to satisfy two linearly independent conditions (135) and (138), leaving one degree of freedom. As a result, for any given collection of item-specific exit rates, there exist distributions of fundamental processes that satisfy the assumptions of our setup and produce a given joint distribution of flex-price intercepts, trends and shock processes.

■

In the following we discuss restrictions that assumptions of the proposition impose on flex-price distributions. Clearly, flexible prices need to follow a linear trend in logs and be subject to stationary shocks, as in (12). In addition, the distributions of flex-price trends and shock processes, from which firms draw upon entry, need to be time-invariant. The BGP condition 1 requires that expectations:

$$\mathbb{E}_k \left[e^{c^{g,0}+sc^{g,1}} \right] \tag{139}$$

$$\mathbb{E}_k \left[e^{(1-\theta)(\ln G^0 - \ln Q^0 + \ln \mu^{f,0}) + s(1-\theta)\left(\ln \overline{\mu}^f + \ln \overline{G}\right)} \right] \tag{140}$$

are constant over time for each item k. The first one does not impose any restrictions and holds for a constant $c^{g,0}_{jk}$, as in the proof of proposition 9. The second one, in contrast, restricts time-variation in the distribution of flex-price intercepts from which firms draw upon entry. In particular, these distributions are allowed to vary over time, as long as (140) remains constant over time, given the time-invariant distribution of flex-price trends. Assumption 2.1 requires that:

$$\mathbb{E}_k \left[e^{c^{g,0}+sc^{g,1}} \middle| \ln \bar{\mu}^f + \ln \bar{G}, \mathfrak{f}, \mathfrak{g}, \mathfrak{m} \right] \overset{!}{=} \mathbb{E}_k \left[e^{c^{g,0}+sc^{g,1}} \right]$$

$$\mathbb{E}_k \left[e^{(1-\theta)(\ln G^0 - \ln Q^0 + \ln \mu^{f,0}) + s(1-\theta)\left(\ln \overline{\mu}^f + \ln \overline{G}\right)} \middle| \ln \bar{\mu}^f + \ln \bar{G}, \mathfrak{f}, \mathfrak{g}, \mathfrak{m} \right] \overset{!}{=} \mathbb{E}_k \left[e^{c^{f,0}+sc^{f,1}} \right]$$

The first line again does not impose any restrictions on flex-price distributions and can be trivially satisfied, e.g. given a constant $c^{g,0}_{jk}$ as in the proof of proposition 9. The second line, however, requires that flex-price intercepts and trends are not independently distributed across products within an item, but instead ensure that average expenditure weights ψ^d_{jkt} at the point of approximation[56] are independent from flex-price trends, shock processes and pricing frictions. Finally. assumption 2.2 requires that flex-price trends are

[56]Which is the deterministic flex-price economy.

drawn independently from flex-price shock processes.

Altogether, our setup imposes minimal restrictions on the joint distributions of flex-price intercepts, trends and shock processes. It effectively only (i) requires time-invariance of the distributions of trends and shock processes, and limits time-variation in the distribution of intercepts to ensure BGP; (ii) imposes independence of trends from shock processes; and (iii) requires some correlation between trends and intercepts to preserve equal product weighting in expectation at the point of approximation. We stress that the last restriction of equal product weighting in expectation is only required for the deterministic flex-price economy, which is our point of approximation. In the stochastic sticky-price economy, average expenditure weights *do* depend on the inflation gap, pricing frictions and shock processes, via the endogenous distribution of price gaps.

A.7 Example of a random menu cost model

Suppose a firm is subject to a random menu cost model, in which firm's price adjustment cost κ_{jkt} is drawn from some exogenous stationary stochastic process and is scaled by after-tax flex-price sales $(1+\tau_{jkt})P^*_{jkt}Y^*_{jkt}$. The firms' objective is to maximize the discounted stream of profits minus adjustment costs:

$$E_t \sum_{i=0}^{\infty} \beta^i (1-\delta_k)^i \frac{U_{Ct+i}}{U_{Ct}} \left(\frac{D_{jkt+i}}{P_{t+i}} - (1+\tau_{jkt+i}) \frac{P^\star_{jkt+i} Y^\star_{jkt+i}}{P_{t+i}} \mathbb{I}_{jkt+i} \kappa_{jkt+i} \right)$$

where indicator function $\mathbb{I}_{jkt+i}$ is equal to one if firm adjusts in period $t+i$ and is zero otherwise. Following the same steps as in section A.1.7, we obtain:

$$\psi_k E_t \sum_{i=0}^{\infty} \left[\beta\gamma^{1-\sigma}(1-\delta_k)\right]^i Q^{\theta-1}_{jkt+i} \Bigg[\left((1+\tau_{jkt+i}) \frac{P_{jkt+i}}{P_{kt+i}} - \frac{w}{p_k} G_{jkt+i} \right) \left(\frac{P_{jkt+i}}{P_{kt+i}} \right)^{-\theta} - (1+\tau_{jkt+i}) \left(\frac{P^*_{jkt+i}}{P_{kt+i}} \right)^{1-\theta} \mathbb{I}_{jkt+i} \kappa_{jkt+i} \Bigg]$$

Using price gap notation (79), expression for desired price (64), BGP equations (80) and (81), and the definition of desired markup (63), we can write:

$$\psi_k \left(\frac{w}{p_k} \right)^{1-\theta} E_t \sum_{i=0}^{\infty} \left[\beta\gamma^{1-\sigma}(1-\delta_k)\right]^i \left(\mu^f_{jkt+i} \right)^{-\theta} \left(\frac{G_{jkt+i}}{Q_{jkt+i}} \right)^{1-\theta} \Bigg[\left(\frac{\theta}{\theta-1} z^{1-\theta}_{jkt+i} - z^{-\theta}_{jkt+i} \right) - \frac{\theta}{\theta-1} \mathbb{I}_{jkt+i} \kappa_{jkt+i} \Bigg]$$

Removing the scaling factor independent of i yields:

$$E_t \sum_{i=0}^{\infty} \left[\beta\gamma^{1-\sigma}(1-\delta_k)\right]^i \left(\mu^f_{jkt+i}\right)^{-\theta} \left(\frac{G_{jkt+i}}{Q_{jkt+i}}\right)^{1-\theta} \Bigg[\left(\frac{\theta}{\theta-1} z^{1-\theta}_{jkt+i} - z^{-\theta}_{jkt+i}\right) - \frac{\theta}{\theta-1}\mathbb{I}_{jkt+i}\kappa_{jkt+i}\Bigg]$$

which shows that in this setup the only aggregate variable that matters for firm's pricing decisions is aggregate inflation, which only affects the drift of price gaps in between adjustments, given by (the negative of) inflation gap $\hat{\pi}_{jk} = \pi_k - \pi^*_{jk} = \pi + \ln\gamma - \ln\gamma_k - \pi^*_{jk}$. Clearly, as long as shock realizations and price gaps at entry are drawn from respective stationary distributions, the distribution of adjustment times does not depend on product age. Therefore, the setup satisfies assumption 1.3 in the main text.

B Testing Key Assumptions of the Approach

This appendix tests key assumptions underlying our theoretical approach in the data.

B.1 The Presence of Heterogeneous Relative Price Trends

Since our empirical approach relies on the presence of heterogeneous desired inflation rates π^*_{jk} across products j within an item k, we formally test the null hypothesis of identical desired inflation rates in each item. Specifically, for each item $k \in \{1, \ldots, K\}$, we test the null of a common slope considering regression (34):

$$\ln p_{jkt} = \alpha_{jk} - t \cdot \beta_{jk} + u_{jkt}$$
$$H_0: \ \beta_{jk} = \beta_{ik} \text{ for all } i, j$$

using the approach of Blomquist and Westerlund (2013) who extend the slope homogeneity test of Pesaran and Yamagata (2008) to a setting with auto-correlated errors. The test is designed for panel data models where the cross-sectional dimension is large compared to the time series dimension, as in our setup. The HAC estimator of the covariance matrix is constructed with three alternative kernels: Bartlett, quadratic spectral (QS) and truncated (see Bersvendsen and Ditzen (2021) for details).

Expenditure-weighted % share of items with p-value < 0.01.

	Bartlett	QS	Truncated
Bandwidth = 6	100	98.8	97.1
Bandwidth = 12	100	97	95.3

Table 3: Trend Homogeneity Test

Table 3 summarizes expenditure-weighted % shares of items with p-values less than 0.01 for the three kernels, using bandwidth values of 6 and 12 months. It shows that the null of no within-item trend heterogeneity is strongly rejected in the vast majority of items in our sample. Increasing the minimal number of products per item from 50 (baseline)

Expenditure-weighted % share of items with p-value < 0.01.

	Bartlett	QS	Truncated
Bandwidth = 6	100	99.9	99.3
Bandwidth = 12	100	99.6	98.8

Table 4: Trend Homogeneity Test (> 100 products per item)

to 100 (covers 91% of baseline expenditure share) strengthens the result even further, see table 4. This shows that desired inflation rates are indeed heterogeneous across products within an item.

B.2 The Absence of a Random Walk in Flexible Prices

Our empirical analysis assumes that the shocks to the flexible relative price $\ln p^*_{jkt}$ in equation (12) are stationary. This appendix shows that our data strongly rejects the alternative hypothesis that the shocks follow instead a random walk, i.e., that the flexible relative price is given by

$$\ln p^*_{jkt} = \ln p^{*,0}_{jk} - s_{jkt} \cdot \pi^*_{jk} + \ln \varepsilon^{\mu}_{jkt},$$

with $\ln \varepsilon^{\mu}_{jkt}$ being a random walk.

To test for the presence of a random walk, we leverage the following insight: since the random walk is a martingale, the optimal reset price $\ln p^{opt}_{jkt_n}$ in price-adjustment periods t_{jkn} $(n = 1, 2, ...N_{jk})$ will involve a constant gap to the flexible relative price $\ln p^*_{jkt}$. Therefore, the change between two reset prices is given by

$$\ln p^{opt}_{jkt_{n+1}} - \ln p^{opt}_{jkt_n} = -\,\pi^*_{jk} \cdot (t_{n+1} - t_n) + \ln e_{jkn+1} \tag{141}$$

where

$$\ln e_{jkn+1} \equiv \ln \varepsilon^{\mu}_{jkt_{n+1}} - \ln \varepsilon^{\mu}_{jkt_n}.$$

With a random walk in $\ln \varepsilon^{\mu}$, the residuals $\ln e$ are uncorrelated over time, which can be tested. To do so, we re-scale residuals according to

$$\ln e^s_{jkn+1} = \frac{\ln e_{jkn+1}}{\sqrt{t_{n+1} - t_n}}$$

to make them homoskedastic under the null hypothesis of a random walk and then compute the autocorrelations $\widehat{Corr}_k = \widehat{Cov}_k / \widehat{Var}_k$ of the re-scaled residuals within each item k using the variance and covariance estimates for all products with $N_{jk} > 3$:

$$\widehat{Var}_k = \sum_j \left(\frac{N_{jk} - 2}{\sum_l (N_{lk} - 2)} \sum_{n=2}^{N_{jk}} \frac{\left(\ln e^s_{jkn}\right)^2}{N_{jk} - 2} \right)$$
$$\widehat{Cov}_k = \sum_j \left(\frac{N_{jk} - 3}{\sum_l (N_{lk} - 3)} \sum_{n=2}^{N_{jk}-1} \frac{\ln e^s_{jkn} \ln e^s_{jkn+1}}{N_{jk} - 3} \right)$$

The top left panel in figure 13 depicts the estimated autocorrelations for all expenditure items k in our baseline sample. Almost all of the autocorrelations are negative and most of them sizably so, which is not consistent with $\ln \varepsilon^{\mu}_{jkt}$ following a random walk. The right panel in the figure reports the bootstrapped p-values for the autocorrelation being weakly larger than zero, as implied by the random walk, and shows that these values are very low.

We then repeat the analysis when exogenously imposing $\pi^*_{jk} = 0$ for all products in equation (141). This is motivated by the possibility that the estimated time trends π^*_{jk} could be purely spurious in the presence of a random walk in $\ln \varepsilon^{\mu}_{jkt}$. The auto-correlations of the resulting residuals are shown in the lower left panel of figure 13 and continue to be negative. The bootstrapped p-values - shown in the lower right panel of the figure - also remain very low.

We then repeat the above analysis for all alternative sales filters considered also in section 8.2. Autocorrelations continue to be negative and the bootstrapped p-values for a weakly positive auto-correlation remain low. Figure 14 illustrates this for the Kehoe-Midrigan regular price filter (RGF).

Based on these findings, we can conclude that in our data unobserved shocks to the flexible price do *not* follow a random walk.

B.3 Relative Price Trends and Length of Product Life

Our approach assumes that the product exit probability δ_k in expenditure item k is the same across all products and thus independent of products' relative price trends. This implies that the length of a product life should not systematically depend on the relative price trend. We test this assumption by considering the regression

$$\ln T_{jk} = \chi_k + \chi_1 \cdot \pi^*_{jk} + \epsilon_{jk}, \tag{142}$$

where T_{jk} denotes the maximum observed age of the product and χ_k an item fixed effect. The coefficient of interest is χ_1, which we estimate using all 234,047 products in our sample. Specifically, if products whose relative price increases over time have a systematically shorter life in our sample, we would observe $\chi_1 > 0$. However, the regression produces a

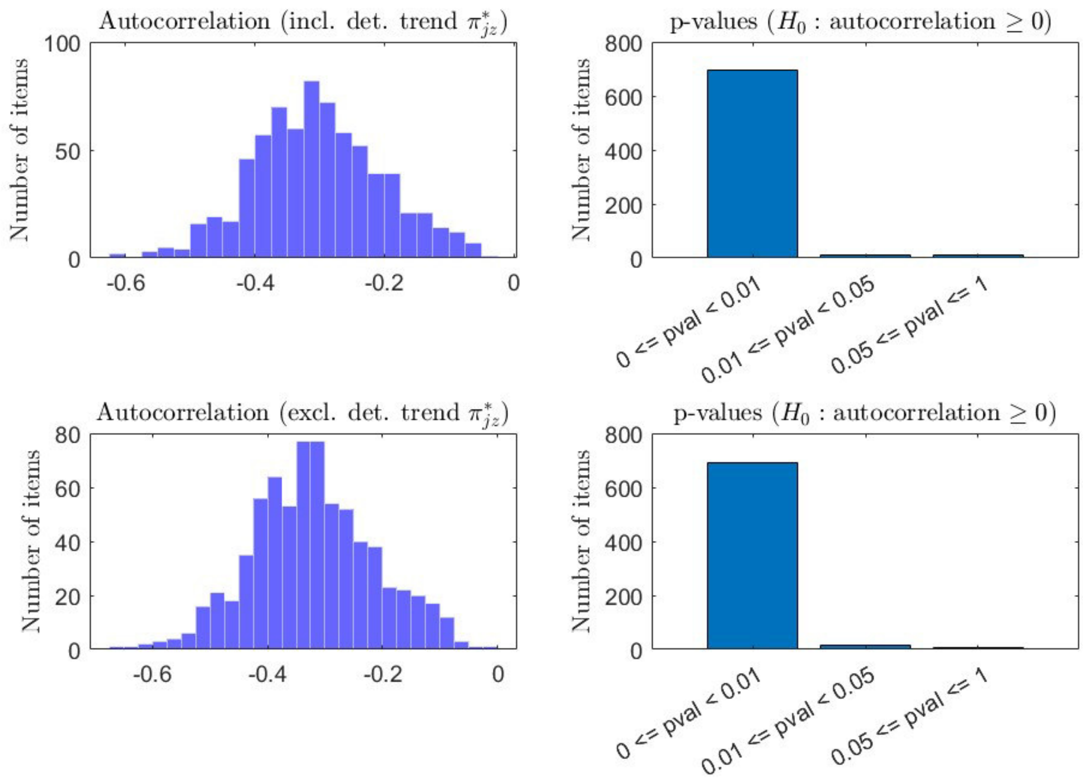

Figure 13: Autocorrelation of residuals (left panel) and bootstrapped p-values (right panel): random walk implies autocorrelation of zero (baseline data sample)

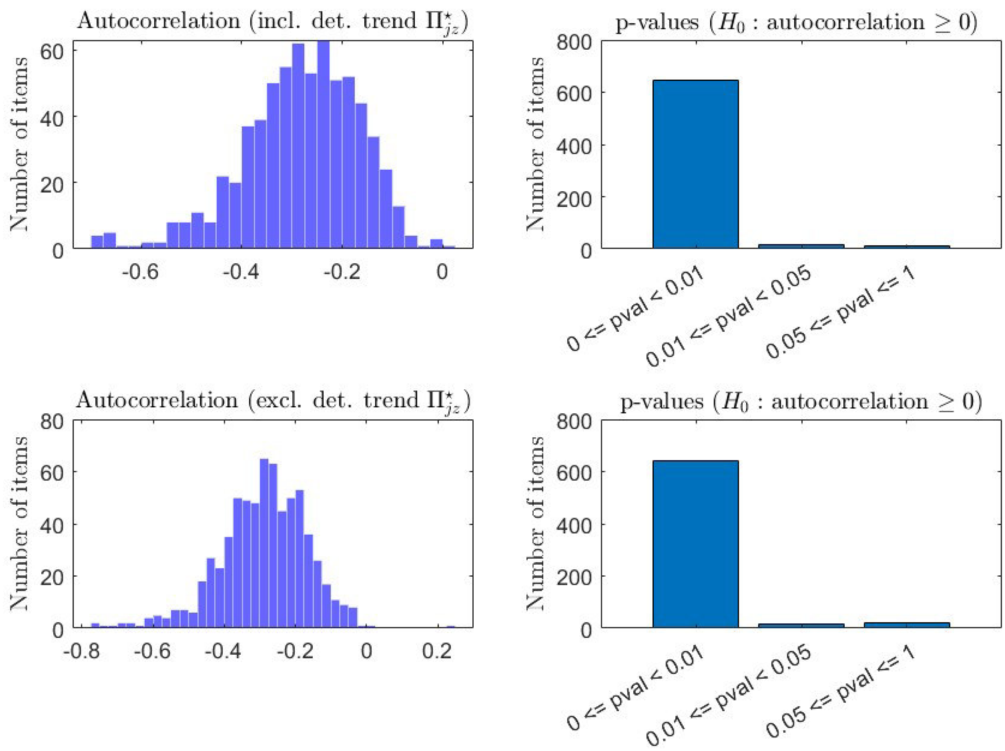

Figure 14: Autocorrelation of residuals (left panel) and bootstrapped p-values (right panel): random walk implies autocorrelation of zero (Kehoe-Midrigan regular price filter)

statistically insignificant negative coefficient $\chi_1 = -0.2374$, with a t-statistic of 1.4. In line with our theoretical assumption, there is thus no evidence that products whose relative price increases at a faster rate face systematically higher exit risk.

C Details of the Regression Approach

This appendix discusses econometric details associated with the estimation of the sufficient statistics (b_k, c_k) based on proposition 6.

For every expenditure item k, we estimate by OLS a first-stage seemingly unrelated regression (SUR) system consisting of equation (34) and the nominal price regression

$$\ln P_{jkt}/Q_{jkt} = \breve{\alpha}_{jk} + t \cdot (\pi_k^{act} - \pi_{jk}^*) + \breve{u}_{jkt} \tag{143}$$

for every product j. The estimation of equation (34) produces unbiased estimates $\widehat{\pi}_{jk}^*$ of the coefficient π_{jk}^* and of the residual variance in equation (34),

$$\widehat{VAR}_{jk}(u_{jkt}) = \frac{1}{T_{jk} - 2} \sum_t \left(\widehat{u}_{jkt}\right)^2,$$

where T_{jk} denotes the number of price observations for product j in item k and $\widehat{u}_{jkt}$ the estimated residuals of equation (34). Summary statistics of the first-stage estimates of equation (34) are reported in figure 1. Similarly, the estimation of equation (143) provides an unbiased estimate $(\widehat{\pi_k^{act} - \pi_{jk}^*})$ of the coefficient $(\pi_k^{act} - \pi_{jk}^*)$. On econometric grounds, we prefer to use this latter estimate in our second-stage equation below, because it avoids having to deal with separate first-stage estimates of π_{jk}^* and π_k^{act}.[57]

To ensure that our results are not driven by outliers, e.g., associated with errors in price collection, within each expenditure item, we eliminate all products falling into the top 5% of the distribution of residual variances $\widehat{VAR}_{jk}(u_{jkt})$ and the top 5% of estimated squared inflation gaps $(\widehat{\pi_k^{act} - \pi_{jk}^*})^2$.

We then run for every item k the second-stage regression (35) across products j in the item, replacing $VAR_{jk}(u_{jkt})$ and $(\pi_k^{act} - \pi_{jk}^*)$ by their corresponding first-stage estimates. This delivers the OLS estimates of the coefficients $(\tilde{a}_k, b_k, c_k)$ depicted in figure 2.

First-stage estimates are subject to estimation error, and this will cause the second-stage estimates of b_k and c_k to be biased towards zero. To see why, note that OLS estimation implies that the estimation error in $\widehat{VAR}_{jk}(u_{jkt})$ is independent of the estimation error in $(\widehat{\pi_k^{act} - \pi_{jk}^*})$, whenever residuals are normally distributed.[58] As a result, the estimation error in $\widehat{VAR}_{jk}(u_{jkt})$ is also independent of that in $((\widehat{\pi_k^{act} - \pi_{jk}^*}))^2$. Together, this implies

[57] Although econometrically desirable, it is not quantitatively important for estimated misallocation costs.

[58] In the more general case with non-normal residuals, the measurement error would still be orthogonal.

that the estimation error on the left-hand side of the second-stage regression is harmless because it gets absorbed by the second-stage residual. The estimation error on the right-hand side of the second-stage regression is such that the error contained in $(\widehat{\pi_k^{act} - \pi_{jk}^*})$ and $((\widehat{\pi_k^{act} - \pi_{jk}^*}))^2$ is orthogonal by construction, so that one does not have to worry about the covariance of measurement errors in the regressors. Therefore, the second-stage estimates of b_k and c_k individually suffer from standard attenuation bias.

D Details of the Calibrated Multi-Item Model from Section 7

In this appendix, we calibrate a multi-sector model to our baseline sample and compare sufficient statistics estimated using our approach with the ones computed directly using the model's solution. Most importantly, we preserve the small-sample property of individual price series on which we perform our estimation.

D.1 Setup

The model features K sectors indexed by $k \in \{1, \ldots, K\}$, each corresponding to an expenditure item in our data. We assume that each sector/item is populated by products that face the same (item-specific) price rigidity $\mathfrak{f}_{jk} = \mathfrak{f}_k$ and draw idiosyncratic shocks from the same (item-specific) processes $\mathfrak{g}_{jk} = \mathfrak{g}_k$ and $\mathfrak{m}_{jk} = \mathfrak{m}_k$. The only source of ex-ante product heterogeneity within an item are product-specific optimal relative price trends, determined by product-specific trends in productivity and mark-up. We write and solve the model numerically in continuous time. To simplify numerical solution, we assume that firms are subject to markup shocks only, which follow a mean-reverting process:[59]

$$d \ln \varepsilon_{jkt}^{\mu} = -\eta \ln \varepsilon_{jkt}^{\mu} dt + \sigma_k dW_{jkt}$$

where $\ln \varepsilon_{jkt}^{\mu}$ is the stochastic part of product j-s desired markup, η is the mean-reversion parameter common across all sectors, σ_k is the sector-specific volatility of idiosyncratic markup shocks and W_{jkt} is the product-specific Wiener process. We use the Calvo-Plus framework, in which a firm may reset its product's price at any time subject to an item-specific menu cost κ_k or wait for a free adjustment opportunity that arrives at an item-specific Poisson rate λ_k. As in the theoretical framework in the main text, firms face a CES demand function.

[59] Allowing for both types of shocks introduces an additional (third) state variable which substantially increases computing and hence calibration and simulation time. The current version with two state variables takes about 60 hours on a 32-core machine to calibrate.

D.2 Calibration

We set the two parameters common across items to values common in the literature: mean reversion parameter $\eta = -\log(0.7)$ and elasticity of substitution $\theta = 7$. Setting $\eta = -\log(0.7)$ implies a mean reversion rate of 0.7 at monthly frequency. All item-specific parameters we calibrate for each item individually. Each item is characterized by the triple $\{\sigma_k, \lambda_k, \kappa_k\}$, as well as the distribution of inflation gaps $\hat{\pi}_{jk}$ and the distribution of product lifespans T_{jk}. The last two objects we read directly from our data. The former three we calibrate following the approach of Nakamura and Steinsson (2010). In particular, the arrival rate of free adjustment opportunities λ_k is set to the empirical (comparable) substitution rate, whereas σ_k and κ_k are jointly calibrated to match the frequency of costly adjustments and the average absolute size of costly adjustments. From the initial sample of 818 items we remove 16 items with outlying values of either the frequency of price adjustment, average adjustment size, substitution rate or average inflation gap, and calibrate the model to 802 items. The removed items correspond to a total 1.5% expenditure share in our sample. Figure 7 in the main text plots empirical and simulated targeted moments across all items in the calibrated model, together with the red 45-degree line.

D.3 Simulation and Estimation

For each calibrated item, we compute two measures of our sufficient statistics – one computed directly using the model's solution and one estimated using our empirical approach.

First, we use the stationary joint distribution of price gaps and idiosyncratic shocks to compute the 'true' approximated item-level productivity loss function:

$$\ln \frac{\Delta_k}{\Delta_k^f} = \frac{\theta}{2}\Big(Var_k(\ln z) + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu]\Big) + O(3)$$

and estimate this function to second order in inflation gap:

$$\ln \frac{\Delta_k}{\Delta_k^f} = \frac{\theta}{2}\left(F_k(0) + \underbrace{\frac{\partial F_k(0)}{\partial \hat{\pi}}}_{b_k} \mathbb{E}_k\left[\hat{\pi}\right] + \underbrace{\frac{1}{2}\frac{\partial^2 F_k(0)}{(\partial \hat{\pi})^2}}_{c_k} \mathbb{E}_k\left[\hat{\pi}^2\right]\right) + O(3)$$

where $F_k(\hat{\pi}) = Var_k(\ln z|\hat{\pi}) + 2\mathbb{E}_k[\ln z \ln \varepsilon^\mu|\hat{\pi}]$. Coefficients b_k and c_k are the 'true' sufficient statistics that are computed directly on the distribution of price gaps and shocks, which remain unobserved by an econometrician.

Second, we follow our empirical approach of estimating these sufficient statistics on simulated price series. We simulate 10000 products with inflation gaps drawn from the empirical item-specific distribution, for the number of periods drawn from the empirical

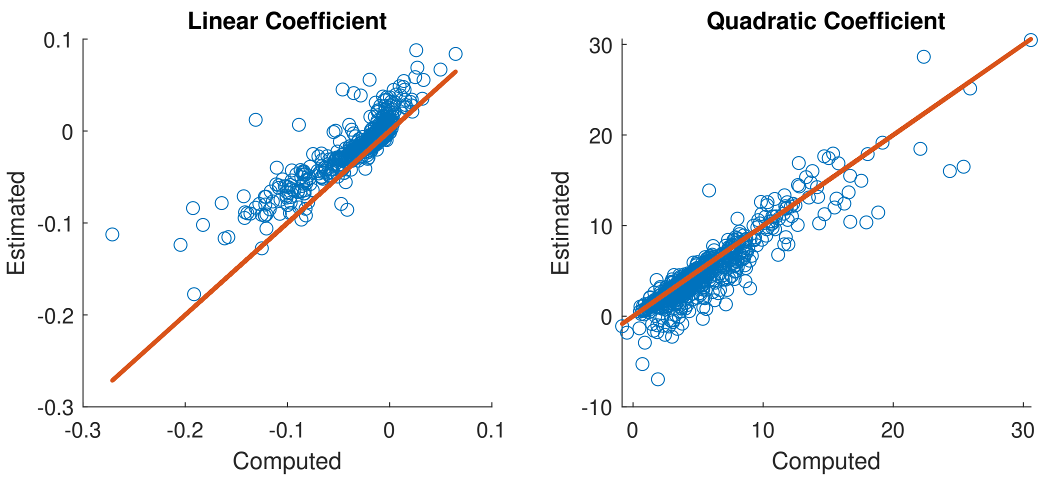

Figure 15: Computed vs Estimated Sufficient Statistics

item-specific distribution of product lengths. Our simulation procedure removes sampling uncertainty of products, but preserves the small-sample property of price time series on which product-specific inflation gaps are estimated, as well as the empirical distribution of inflation gaps on which we estimate sufficient statistics. We follow our two-step procedure and estimate product-specific inflation gaps and variances of residuals in the first-stage regression. We then run an item-level second-stage regression to obtain estimates of sufficient statistics $\widehat{b}_k$ and $\widehat{c}_k$, corresponding to coefficients in front of the linear and quadratic term, respectively.

We perform this exercise across all 802 items and provide the results in Figure 15. The figure shows the 'true' values of sufficient statistics b_k and c_k computed directly using the model's solution (x-axis) against our estimates $\widehat{b}_k$ and $\widehat{c}_k$ obtained on simulated price data (y-axis). The left panel corresponds to the linear term, and the right panel to the quadratic term. We add a 45-degree line in red. Both terms exhibit a good fit, with some attenuation bias due to measurement error in the first-stage regression because of the short-sample properties of our simulated data. Figure 16 plots the trends used to simulate products (on the x-axis) against the various percentiles of estimated trends distribution (on the y-axis). The distributions are centered around the 'true' value and exhibit some dispersion of estimated trends around it due to short samples on which we estimate the first stage. In fact, restricting the sample to products with at least 60 observations (in contrast to 30 in the baseline), reduces both the variability in trend estimates in the first stage and attenuation bias in the second stage, as shown on Figures 17 and 18.

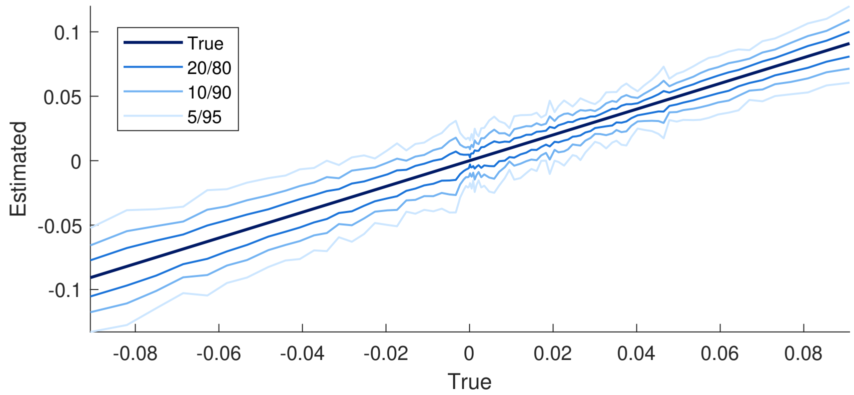

Figure 16: Distribution of Estimated Trends

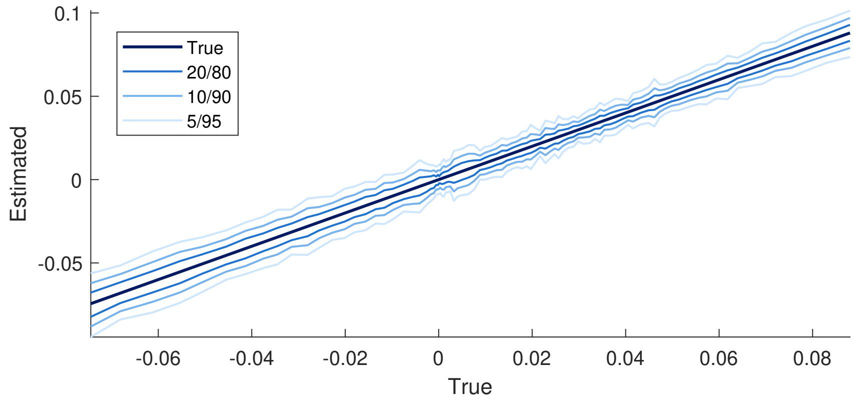

Figure 17: Distribution of Estimated Trends, at least 60 obs

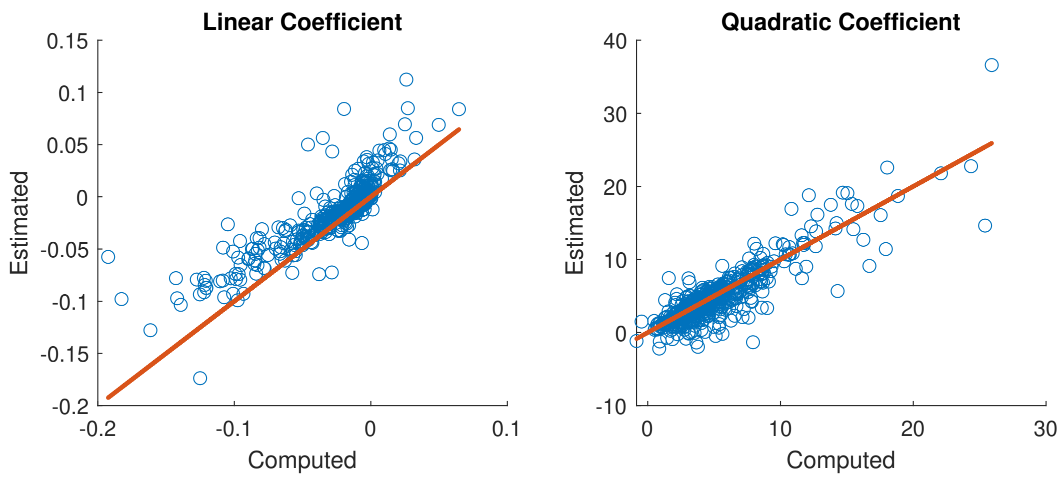

Figure 18: Computed vs Estimated Sufficient Statistics, at least 60 obs

D.4 Aggregate Losses

Finally, we compare the true (non-approximated to any order) aggregate loss defined in equation (109) as:[60]

$$\ln \frac{\Delta}{\Delta^f} = \sum_{k=1}^{K} \psi_k \ln \frac{\mu_k \Delta_k}{\mu_k^f \Delta_k^f} - \left(\ln \sum_{k=1}^{K} \psi_k \frac{1}{\mu_k^f} - \ln \sum_{k=1}^{K} \psi_k \frac{1}{\mu_k} \right)$$

against the one obtained using our approximation and estimation procedure:

$$\ln \frac{\Delta}{\Delta^f} = \frac{\theta}{2} \sum_{k=1}^{K} \psi_k \left(F_k(0) + \widehat{b}_k \mathbb{E}_k \left[\hat{\pi}\right] + \widehat{c}_k \mathbb{E}_k \left[\hat{\pi}^2\right] \right) + O\left(3\right)$$

for various levels of aggregate inflation rate. This exercise corresponds to our main result on figure 3. Figure 8 in the main text shows the true non-approximated aggregate loss in blue and our estimate in red (both normalized to zero at the mininum). Our approach recovers the true loss remarkably well despite targeting its second order approximation and small-sample properties of simulated data.

[60]See Appendix A for more details.