Productivity Shocks and Firm Real Business Cycles

Luca Gagliardone* Simone Lenzu† Joris Tielens§

October, 2026

Abstract

Do productivity shocks generate firm-level business cycles? We study the effects of transitory, idiosyncratic disturbances on firms' production and pricing decisions at business-cycle frequency. Using a quarterly administrative micro-dataset for the Belgian manufacturing sector, we identify firm-level productivity and demand shocks from orthogonal movements in technical efficiency (TFPQ) and revenue productivity (TFPR). These measures are recovered using a new methodology for production function estimation via panel error-correction models that allows for (i) nominal rigidities, (ii) adjustment costs on all inputs, (iii) endogenous dynamics of productivity (learning by doing, R&D), and (iv) variable markups. We document that such disturbances do not generate cyclical fluctuations in firms' production and pricing decisions, but have small and highly persistent effects. Using a quantitative model estimated with the microdata, we find that idiosyncratic wedges to the first-order conditions (e.g., firm-specific input prices) explain the lion's share of the movements in inputs and output prices.

Keywords: Real business cycle; microdata; productivity; firms; production function.

JEL: E32, L11, D24.

* Yale University. Email: luca.gagliardone@yale.edu; † Federal Reserve Bank of New York, University of Cambridge, and CEPR. Email: s.lenzu@jbs.cam.ac.uk; § National Bank of Belgium and KU Leuven. Email: joris.tielens@nbb.be. We thank Mohamad Adhami, Keshav Dogra, Joel Flynn, Mark Gertler, Zhen Huo, Ricardo Lagos, Geert Mesters, Giuseppe Moscarini, Georgios Nikolakoudis, Michael Peters, David Rivers, Aleh Tsyvinski, and Venky Venkateswaran for helpful comments and conversations. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the National Bank of Belgium, the Eurosystem, the Federal Reserve Bank of New York, the Federal Reserve System, or any other institution with which the authors are affiliated.

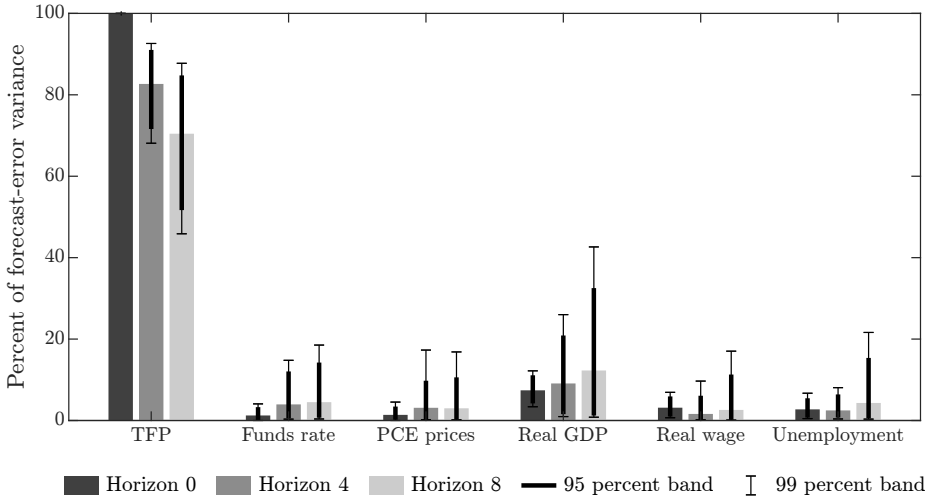
1 Introduction

The origins and propagation of aggregate business-cycle fluctuations have long been central questions of macroeconomics. The seminal contributions of Kydland and Prescott (1982; 1991) advanced the real business cycle (RBC) hypothesis, which attributes economic fluctuations primarily to total factor productivity (TFP) shocks. Subsequent research challenged this view through advances in time-series econometrics (Blanchard and Quah, 1989; Galí, 1999; Christiano et al., 2003; Francis and Ramey, 2005) and quantitative historical analyses (Christiano et al., 2005; Smets and Wouters, 2007; Justiniano et al., 2010). Today, the prevailing view is that shocks to measured aggregate productivity explain only a small share of business-cycle fluctuations. We replicate this finding in Figure 1 using a measure of utilization-adjusted total factor productivity (Fernald, 2014). The figure reports the forecast-error variance decomposition of a vector autoregression model estimated on US quarterly aggregate data over 1973–2019. Productivity shocks account for roughly one-tenth of the variance of GDP at one- and two-year horizons and less than 5 percent of that of prices, wages, and unemployment. The impulse responses reinforce the same conclusion (Basu et al., 2006): the effects on output, wages, and unemployment are modest and not statistically significant after a few quarters (see Figure A.1 in Appendix C).

The limited explanatory power of aggregate TFP in explaining aggregate fluctuations, however, does not preclude productivity shocks from driving business-cycle fluctuations within firms—a firm-level RBC mechanism. For one, prior research has shown that idiosyncratic efficiency shocks can induce substantial firm-level adjustment yet contribute little to aggregate fluctuations when they are weakly correlated across firms (Hopenhayn, 1992; Franco and Philippon, 2007). For another, aggregate measured TFP need not move one-for-one with technical efficiency because it also reflects the allocative effects of market distortions (Basu and Fernald, 2002; Hsieh and Klenow, 2009; Baqaee and Farhi, 2020). Finally, general-equilibrium movements in factor prices can induce offsetting responses across firms and make the effects of idiosyncratic productivity shocks differ from those of common technology shocks (Carvalho and Grassi, 2019; Khan and Thomas, 2008). Aggregate evidence alone therefore leaves the role of productivity shocks in firms' decisions at business-cycle frequency unresolved.

Obtaining credible evidence for a firm-level RBC mechanism is challenging for various reasons. Measuring technical efficiency requires high-frequency data not only on inputs and output but also on firms' prices and capacity utilization. As both variables may respond to productivity shocks, failing to account for them can contaminate measured TFP and confound the identification of the underlying disturbances. Moreover, production-function estimation methods —typically applied to low-frequency data— often rely on assumptions about input adjustment

Figure 1: Forecast-error variance of US aggregates explained by TFP shocks



Notes. Shares of forecast-error variance explained by a TFP shock at horizons of zero, four, and eight quarters. TFP shocks are identified recursively with TFP (Fernald, 2014) ordered first. The vector autoregression is estimated on quarterly US data over 1973Q1–2019Q4 and includes four lags and a constant. The variables are the log of utilization-adjusted TFP (Fernald, 2014); the proxy federal funds rate (Choi et al., 2022); the log PCE price index; log real GDP (Brave et al., 2019); the log real wage, measured as average hourly earnings of production and nonsupervisory employees deflated by the core PCE price index; the unemployment rate; and the excess bond premium (Gilchrist and Zakrajšek, 2012), whose variance decomposition is not shown. The impulse responses are reported in Figure A.1 in Appendix C.

and pricing that are difficult to support at business-cycle frequency. Finally, standard time-series econometric methods used to identify the dynamic effects of aggregate productivity shocks require long histories that are rarely available for individual firms.

To our knowledge, this paper provides the first direct empirical test of a firm-level RBC mechanism by identifying the dynamic effects of temporary, idiosyncratic productivity shocks on firms' pricing and production decisions at quarterly frequency. We assemble a micro-level panel that combines administrative and survey quarterly microdata covering the lion's share of manufacturing firms in Belgium over the period 1995-2023. These data allow us to observe production, output prices, inventories, inputs, investment, and capacity utilization at business-cycle frequency. We develop an econometric framework that exploits these long firm histories to recover productivity, allowing for nominal and real rigidities, variable markups, and endogenous productivity. Our identification approach embeds the long-run relationships underlying leading approaches to production-function estimation (Gandhi et al., 2020) in an error-correction model (Engle and Granger, 1987; Johansen, 1991) of firms' short-run dynamics.

Together, the data and framework allow us to identify and decouple transitory productivity shocks—innovations in technical efficiency (TFPQ)—from demand (sales) shocks—innovations in revenue productivity (TFPR) orthogonal to TFPQ. The resulting impulse responses illustrate the role of technology and demand as drivers of firms' cycles.

Our central finding is that transitory productivity shocks generate little cyclical adjustment in firms' inputs and output prices. Together with demand shocks, such shocks account for less than 5 percent of the variance decomposition of labor, materials, and capital at business-cycle horizons. Consistent with the aggregate evidence discussed, these findings provide little support for a firm-level RBC mechanism. Firms absorb much of the temporary increase in production through inventories, while the modest adjustments in inputs persist well beyond the initial shock.

A quantitative model with substantial input adjustment costs and low inventory holding costs rationalizes this pattern. Through the lens of the model, input fluctuations are driven primarily by firm-specific, time-varying wedges to the first-order conditions. These wedges reflect, among other disturbances, heterogenous access to input markets in the form of distorted and shadow prices. Their presence kills the propagation of productivity shocks to firms' policies —the RBC mechanism— and instead appear to be the main candidate for explaining production decisions and cycles within the firm.

Identifying productivity shocks at business-cycle frequency requires taking into account short-run nominal and real rigidities. These pose new identification challenges that standard production function estimation approaches are not designed to address. For example, at quarterly horizons, a residual interpreted as measurement error under flexible inputs and prices may instead capture persistent, endogenous adjustment (Fernald et al., 2026). Our methodological insight is to retain firms' frictionless optimality conditions as long-run cointegrating relationships and embed them in a firm-level vector error-correction model (VECM), a special case of a panel VAR in levels. Formally, we replace the assumption that productivity can be recovered from a single, often static optimality condition with the weaker assumption of dynamic invertibility, a standard assumption in VAR analysis: productivity shocks are recovered from the entire history of the whole system of firm-level pricing and production decisions. This representation allows us to identify technology without requiring any input or prices to adjust freely within the quarter. Additionally, the same insight allows the framework to accommodate variable markups and endogenous productivity dynamics, including those stemming from learning by doing and investment in R&D.

Section 2 develops the theoretical framework underlying our empirical strategy. Firms produce differentiated goods using a technology with constant elasticity of substitution, Hicks-neutral productivity, and labor, materials, capital, and utilization as inputs. Productivity follows a general process that can endogenously depend on the firm's other decisions. Firms face

a demand curve à la Kimball and compete monopolistically or oligopolistically, so that markups depend endogenously on relative prices. Pricing and cost minimization decisions are specified only for a long-run, frictionless benchmark. In this way, we allow for a broad class of short-run nominal rigidities and adjustment costs on all inputs.

Section 3 provides the identification results for the technology parameters that build on the economic microfoundations. Under the assumption of dynamic invertibility, the joint dynamics of firms' decisions admit a reduced-form vector error-correction representation that nests the control-function approach as a special, testable case. Under additional testable assumptions on the non-stationarity of prices, a long-run pricing relationship and a short-run output equation jointly identify the production-function parameters. In addition, the economic microfoundation and econometric assumptions deliver identified impulse responses to productivity shocks via movements in technical efficiency (TFPQ), and demand shocks via movements in revenue productivity (TFPR) that are orthogonal to TFPQ.

Sections 4 and 5 present the datasets used in the empirical analysis and produce empirical tests for the integration and cointegration of the variables in the system. Our data encompass almost the entire universe of firms in the Belgian manufacturing sector at quarterly frequency for the period 1995-2023. The descriptive statistics are comparable to those reported in previous studies. Relative to the datasets used in that literature, however, our panel offers two key advantages for studying firms' business-cycle dynamics: detailed measurement and long quarterly histories. First, observing firm-level prices and utilization allows us to account for their endogenous responses to demand shocks when measuring technical efficiency, while observing production separately from sales allows us to track inventory adjustment. Second, long firm histories provide the time dimension required for consistent estimation of firm-level time-series models and identification of the dynamic effects with the impulse response functions. Identification of technology rests on both nonstationary variation in the observables and stationary departures from their long-run relationships. We assess these properties using panel unit-root and cointegration tests, persistence measures, and direct tests of system rank.

Section 6 presents the estimates of the production function and dynamics of the system, and describes the recovered productivity measures. The data strongly reject the restriction on adjustment coefficients imposed by a standard control-function benchmark, underscoring the empirical importance of allowing short-run departures from frictionless optimality when working with quarterly data. The estimated production function has returns to scale statistically consistent with constant returns, with a high estimated elasticity on capital and a relatively low elasticity on materials. The estimate of the pricing equation implies long-run marginal-cost pass-through of approximately one-half due to variable markups. Finally, the recovered productivity measures

display substantial transitory movements at the quarterly frequency.

Section 7 traces firms' responses to identified productivity and demand shocks. Treating capital as predetermined within the quarter, we identify productivity shocks from innovations in technical efficiency (TFPQ), and demand shocks from innovations in revenue productivity (TFPR) orthogonal to TFPQ and capital. This distinction allows productivity shocks to affect revenue productivity on impact, while demand shocks can influence technical efficiency endogenously in subsequent quarters. We estimate small but highly persistent effects of both productivity and demand shocks on inputs and prices. The forecast-error variance decomposition confirms the limited explanatory power of both disturbances: together, they account for at most 2.7 percent of the forecast-error variance of inputs and utilization and 5.5 percent of that of prices over six years after the shock.

Section 8 develops a quantitative model to interpret these findings. The model combines input adjustment costs, inventories, nominal rigidities, strategic complementarities, and endogenous productivity dynamics. We discipline its technology parameters using our production-function estimates and estimate adjustment and inventory costs by matching the responses to the identified productivity shock. Substantial input adjustment costs rationalize the small but persistent responses of labor and materials, with implied adjustment half-lives of 6.6 and 5.5 years, respectively. Low inventory holding costs allow firms to absorb temporary increases in production without commensurate changes in sales, while nominal rigidities and strategic complementarities further dampen price adjustment. We then use Bayesian methods to estimate the processes for disturbances (wedges) to firms' pricing, labor, materials, and investment optimality conditions. The shocks are highly transitory but have persistent effects via the high estimated adjustment costs. The model's variance decomposition attributes almost all input variation to each input's own wedge, with limited effects across inputs.

Section 9 concludes. The empirical evidence and the model-based evidence provide little support for a firm-level RBC mechanism in which temporary productivity shocks drive cyclical fluctuations within the firm. These findings, however, do not challenge the role of persistent productivity improvements as an important driver of firms' longer-run dynamics and growth.

Relation to the literature. Our empirical analysis contributes to the study of sources and propagation of business-cycle fluctuations. To our knowledge, we are the first to bring this question to the microdata to test the RBC mechanism, separately identifying transitory productivity and demand innovations and tracing their effects on production, prices, inventories, and inputs. Our findings complement the aggregate evidence of Angeletos et al. (2020), who document a disconnect between TFP fluctuations and the main driver of the business cycle, and

Carlsson et al. (2021), who find modest employment responses to permanent technology shocks using annual Swedish manufacturing data. We show that, even within firms, large temporary gains in technical efficiency generate little of the cyclical input adjustment central to an RBC mechanism. Our quantitative analysis also relates to business-cycle accounting and the role of resource-allocation wedges (Chari et al., 2007; Restuccia and Rogerson, 2013). The estimated model distinguishes input-specific disturbances from the adjustment costs that govern their propagation and attributes most input variation to each input's own wedge. The joint responses of labor, materials, and capital also connect our analysis to the distinction between scale and input-mix factors in David and Venkateswaran (2019). We also complement work on adjustment costs and dynamic resource allocation in Asker et al. (2014) by showing how temporary disturbances to input conditions can leave persistent effects on factor use.

Our methodological contribution extends the production-function estimation approach to an environment with firms' adjustment dynamics at business-cycle horizons. Proxy-variable methods recover productivity by inverting firms' policy functions under restrictions on timing, monotonicity, and unobserved heterogeneity (Olley and Pakes, 1992; Levinsohn and Petrin, 2003; Wooldridge, 2009; Ackerberg et al., 2015). For gross-output production functions, Gandhi et al. (2020) exploit a flexible input's first-order condition to identify technology (Trunschke and Judd, 2024). Building on these insights, our framework also accommodates sluggish adjustment of all inputs, nominal rigidities, variable markups, and endogenous productivity. The key insight is showing how long-run pricing restrictions and short-run output dynamics can be combined to jointly identify technology. In this setting, departures from the long-run relationship capture persistent economic adjustment rather than measurement error alone, while the flexible-input/flexible-price benchmark provides testable restrictions.[1] Formally, by recasting the estimation of the production function in terms of a VAR, we replace the invertibility and monotonicity conditions of a single optimality condition with the weaker condition that the VAR is dynamically invertible.

Our framework connects production-function estimation to the literature on cointegrated VARs, which combines long-run economic relationships with short-run adjustment and structural analysis of shocks (Engle and Granger, 1987; Johansen, 1991; Watson, 1994). Within this framework, we supply the firm-level pricing and production restrictions that identify technology and the additional timing restrictions that distinguish productivity and demand innovations. As in King et al. (1991), economic theory supplies restrictions on long-run comovement within a joint dynamic system. Their analysis centers on permanent aggregate productivity disturbances

[1] See Fernald et al. (2026) for a discussion on how adjustment costs and differences between observed and allocative input prices affect production-based measurement.

associated with common stochastic trends, while Blanchard and Quah (1989) distinguish supply and demand disturbances through their permanent and transitory effects on aggregate output. Our focus is on recovering firm-level technical efficiency and tracing firms' responses to transitory productivity innovations. Long-run pricing relationships and short-run production dynamics jointly identify technology, while separate timing and impact restrictions distinguish productivity and demand innovations without requiring productivity shocks to have permanent effects. The panel implementation also connects to dynamic models of firm behavior (Arellano and Bond, 1991; Blundell and Bond, 2000), with long firm histories supporting estimation of the joint quarterly adjustment dynamics. We also build on evidence on cost pass-through (Nakamura and Zerom, 2010; Gagliardone et al., 2025) and the role of expectations in pricing (Ghassibe et al., 2025; Andrade et al., 2026; Gagliardone and Tielens, 2026). We incorporate these price-cost relationships into a joint system for production and input decisions, using them to identify technology.

Our results are relevant for theories that connect microeconomic disturbances to aggregate fluctuations through the firm-size distribution and firm networks (Gabaix, 2011; Acemoglu et al., 2012; Carvalho and Grassi, 2019; Baqaee and Farhi, 2019, 2020). The aggregate implications of these theories depend on how disturbances change firms' pricing and production decisions, which in turn affect factor markets and spill over to trading partners. We provide new empirical moments that can inform models of aggregation and amplification of idiosyncratic shocks. Our analysis also builds on work distinguishing physical productivity from revenue productivity and demand (Foster et al., 2008). In this literature, Lenzu et al. (2026) show how pricing responses affect the measurement of financing shocks' productivity effects and how inventory management helps firms absorb these disturbances. We study a complementary question: how productivity and demand innovations themselves propagate through firms' production, pricing, and input decisions at quarterly frequency. Finally, our findings connect to the literature on inventories and the propagation of shocks within firms. Inventories allow firms to separate production from sales and shift production toward periods of low marginal cost (Eichenbaum, 1989; Ramey and West, 1999), while financial conditions shape their cyclical adjustment (Gertler and Gilchrist, 1994; Kashyap et al., 1994).[2] We provide conditional firm-level evidence on these adjustment margins: following transitory efficiency gains, inventories absorb much of the increase in production, while sales and input use adjust only modestly.

[2]Granger and Lee (1989) provide an early application of error correction to production, sales, and inventory dynamics, studying multicointegration and asymmetric adjustment in industry-level data.

2 The general model

This section sets out the general economic framework that micro-founds the econometric framework in Section 3. For generality and clarity of exposition, we impose only a minimal set of modeling assumptions needed for the mapping to the econometric model. Unless stated explicitly, we impose no further restrictions on firms' behavior. A complete microfoundation is in Section 8.

2.1 Primitives and environment

Technology. Consider a CES production function with Hicks-neutral total factor productivity (TFPQ, $A_{i,t}$), labor ($L_{i,t}$), intermediates ($M_{i,t}$), capital ($K_{i,t}$), and capital utilization ($U_{i,t}$):

$$\begin{aligned} Q_{i,t} &= A_{i,t}\left[\alpha_K(U_{i,t}K_{i,t})^{\eta} + \alpha_L L_{i,t}^{\eta} + \alpha_M M_{i,t}^{\eta}\right]^{\nu/\eta}, \\ &\alpha_K, \alpha_L, \alpha_M > 0, \quad -1 < \eta < 1, \quad \nu > 0, \qquad \alpha_K + \alpha_L + \alpha_M = 1 \end{aligned} \tag{1}$$

where $\nu = 1$ gives constant returns to scale and a Cobb-Douglas production function is the special case with $\eta \to 0$. Physical production ($Q_{i,t}$), physical sales ($Y_{i,t}$), and the inventory stock ($S_{i,t}$) satisfy:

$$Q_{i,t} = Y_{i,t} + S_{i,t} - (1 - \delta_S)S_{i,t-1},$$

where $\delta_S \in [0, 1)$ denotes the depreciation rate of inventories. Capital depreciates at a rate $\delta(U)$ that may depend on utilization:

$$K_{i,t+1} = (1 - \delta(U_{i,t}))K_{i,t} + I_{i,t}, \qquad 0 \le \delta(U_{i,t}) < 1$$

where $I_{i,t}$ is real investment installed for use at $t+1$. We assume that TFPQ follows a causal, locally log-linear, and possibly endogenous process with any finite number of lags ($0 \le p < \infty$):

$$\ln(A_{i,t}) = \gamma_i^a + \gamma_t^a + \rho \ln(A_{i,t-1}) + \sum_{x \in \{Y,Q,P,U,K,L,M\}} \sum_{\tau=1}^{p} \psi_{\tau,x} \, \Delta \ln(x_{i,t-\tau}) + \varepsilon_{i,t}^a \tag{2}$$

where γ_i^a denotes the firm-specific long-run level of technical efficiency, γ_t^a denotes an industry-specific trend (with the industry subscript omitted), and $\varepsilon_{i,t}^a$ is an i.i.d. structural disturbance with variance $\sigma_a^2 > 0$. Under specification (2), TFPQ exhibits persistent but mean-reverting deviations from an industry trend plus a firm-specific, constant component.

Examples. To illustrate the flexibility of these TFPQ dynamics, consider a Cobb-Douglas production function with constant returns to scale ($\eta \to 0, \nu = 1$). The specification nests an AR(p) process:

$$\ln(A_{i,t}) = (1 - \sum_{\tau=1}^{p} \rho_\tau) \ln(A_i) + \sum_{\tau=1}^{p} \rho_\tau \ln(A_{i,t-\tau}) + \varepsilon_{i,t}^a$$

It also allows for arbitrary forms of learning by doing, such as:

$$\ln(A_{i,t}) = (1-\rho)\ln(A_i) + \rho\ln(A_{i,t-1}) + \psi_\ell \Delta\ln(L_{i,t-1}) + \psi_u \Delta\ln(U_{i,t-1}) + \varepsilon^a_{i,t},$$

Finally, it allows for endogenous firm-specific and aggregate productivity growth through TFPQ-enhancing capital investments, such as R&D or patent acquisition, with or without time-to-build:

$$\ln(A_{i,t}) = \ln(A_t) + \sum_{\tau=1}^{p} \psi_{\tau,K} \Delta\ln(K_{i,t-\tau}) + \varepsilon^a_{i,t}$$

The set of productivity determinants can also be extended to include, for example, intangible capital or demand proxies.

Consumer's demand. Industry consumption Y_t is a bundle of differentiated intermediate inputs $Y_{i,t}$, assembled into a final good using the Kimball (1995) aggregator:

$$\int_0^1 \Upsilon\left(\frac{Y_{i,t}e^{\varphi_{i,t}}}{Y_t}\right) di = 1 \tag{3}$$

where $\varphi_{i,t}$ is a mean-zero stationary demand shifter with i.i.d. innovation $\varepsilon^\varphi_{i,t}$ of variance $\sigma^2_\varphi \geq 0$, and $\Upsilon(\cdot)$ is locally three times differentiable, strictly increasing, and strictly concave, with $\Upsilon(1) = 1$.[3] Denote by $P_{i,t}$ the price of one unit of physical output $Y_{i,t}$. Taking industry demand Y_t as given, the household minimizes expenditure on differentiated goods subject to the aggregate constraint:

$$\min_{Y_{i,t}} \int_0^1 P_{i,t}Y_{i,t}di \qquad \text{s.t.} \quad \int_0^1 \Upsilon\left(\frac{Y_{i,t}e^{\varphi_{i,t}}}{Y_t}\right) di = 1.$$

Let P_t denote the industry price index consistent with the aggregator (3). The household's solution yields a downward-sloping demand curve for $Y_{i,t}$:

$$Y_{i,t} = \mathcal{D}(P_{i,t}, P_t, \varphi_{i,t})Y_t, \tag{4}$$

where $\mathcal{D}(P_{i,t}, P_t, \varphi_{i,t}) \equiv e^{-\varphi_{i,t}}(\Upsilon')^{-1}\left(P_{i,t}/(P_t e^{\varphi_{i,t}})\right)$ is strictly decreasing in relative prices. We denote by $\mu_{i,t}$ the log desired gross markup, which is a function of the residual elasticity of demand:

$$\mu_{i,t} = \ln\left(\frac{\epsilon_{i,t}}{\epsilon_{i,t}-1}\right), \qquad \epsilon_{i,t} \equiv -\frac{\partial \ln Y_{i,t}}{\partial \ln P_{i,t}} \tag{5}$$

We assume a continuum of industries, each with vanishing mass, and an aggregator with constant elasticity of substitution across industries. We omit the industry subscript: relative prices are measured within an industry, whereas deflators are measured at the economy level.

[3] In the control function approach (e.g., Gandhi et al., 2020), identification typically requires a minimum number of shocks. Our approach does not impose this requirement; in contrast, identification is generally stronger with fewer shocks. Accordingly, we allow for $\sigma^2_\varphi = 0$.

Product market competition. We allow for general forms of competition within an industry, including monopolistic competition, static oligopoly á la Bertrand or Cournot as in Atkeson and Burstein (2008), and dynamic oligopoly as in Wang and Werning (2022) and Gagliardone et al. (2025). With a continuum of industries, firms neither directly influence economy-wide prices nor internalize their effects on those prices.

2.2 Optimal policies

Production. We do not require a flexible input. As in Chari et al. (2007) and Restuccia and Rogerson (2013), we allow "wedges" in every input optimality condition, arising, for example, from adjustment costs, incomplete information, or departures from rationality. For identification, it suffices that these wedges are covariance stationary and converge over time to a firm-specific constant plus an industry-specific common factor. Section 5 tests this requirement.

We derive the frictionless benchmark to which firms' input choices converge over time. Taking input prices as given, the firm minimizes expenditure on labor and materials subject to the production function. Let W_t denote the industry wage and P_t^m the industry price of the intermediate good. The cost-minimization problem is:

$$TC_{i,t}^{\star} \equiv \min_{L_{i,t},\, M_{i,t}} W_t L_{i,t} + P_t^m M_{i,t}$$
$$\text{s.t.} \quad Q_{i,t} \geq \overline{Q}_{i,t}.$$

The function $TC_{i,t}^{\star}$ gives the total nominal variable cost of producing $\overline{Q}_{i,t}$ units of output without adjustment costs, given the state variables $A_{i,t}, K_{i,t}$. Let $MC_{i,t}^{\star}$ denote the Lagrange multiplier. The first-order conditions for labor and materials are:

$$W_t = MC_{i,t}^{\star} \frac{\partial Q_{i,t}}{\partial L_{i,t}}, \qquad P_t^m = MC_{i,t}^{\star} \frac{\partial Q_{i,t}}{\partial M_{i,t}}, \qquad MC_{i,t}^{\star} = \frac{\partial TC_{i,t}^{\star}}{\partial \overline{Q}_{i,t}} \tag{6}$$

Let $\theta_{i,t}^L$ and $\theta_{i,t}^M$ denote the output elasticities of labor and materials. Rearranging the first-order conditions gives:

$$MC_{i,t}^{\star} = \frac{1}{\theta_{i,t}^L + \theta_{i,t}^M} \frac{W_t L_{i,t} + P_t^m M_{i,t}}{Q_{i,t}}, \qquad \theta_{i,t}^L \equiv \frac{\partial Q_{i,t}}{\partial L_{i,t}} \frac{L_{i,t}}{Q_{i,t}}, \quad \theta_{i,t}^M \equiv \frac{\partial Q_{i,t}}{\partial M_{i,t}} \frac{M_{i,t}}{Q_{i,t}}. \tag{7}$$

In the short run, frictions may imply $MC_{i,t}^{\star} \neq MC_{i,t}$.

Pricing. Prices need not be flexible either. As with input choices, we introduce a frictionless benchmark and allow actual prices to deviate from it in the short run through a pricing wedge. Such deviations can arise from nominal rigidities in price adjustment (Taylor, 1979; Rotemberg, 1982; Calvo, 1983; Caballero and Engel, 2007; Golosov and Lucas Jr, 2007).

Again using a "$\star$" superscript for the frictionless price, the firm maximizes static real profits subject to demand and inventory dynamics, taking the aggregate price as given:

$$\begin{aligned} \max_{P_{i,t},\, Y_{i,t}} \quad & \frac{P_{i,t}}{P_t} Y_{i,t} - \frac{TC^{\star}_{i,t}}{P_t} \\ \text{s.t.} \quad & Y_{i,t} = \mathcal{D}(P_{i,t}, P_t, \varphi_{i,t}) Y_t \\ & Q_{i,t} = Y_{i,t} + S_{i,t} - (1 - \delta_S) S_{i,t-1} \end{aligned}$$

The first-order condition sets the flexible price at a markup over nominal marginal cost:

$$P^{\star}_{i,t} = e^{\mu_{i,t}} MC^{\star}_{i,t} \tag{8}$$

where $\mu_{i,t}$ is the log desired gross markup defined in Equation (5).

In the short run, a persistent "pricing wedge" may separate actual and flexible prices, $P_{i,t} \neq P^{\star}_{i,t}$ (Gagliardone et al., 2025, 2026; Gagliardone and Tielens, 2026). Under the nominal rigidities described above, the price chosen conditional on adjustment would be the present discounted value of current and future markups over marginal costs, with discounting that depends on the probability of future readjustment. The adjustment friction thus creates a persistent wedge that may be exogenous (as in Taylor, Rotemberg, or Calvo pricing) or endogenous (as in menu cost models). Consistent with this class of models, we assume that the wedge closes in the long run as the price converges to its flexible counterpart. Section 5 tests this requirement.

3 The econometric framework

We now state the additional assumptions needed to identify the model parameters. Lowercase letters denote variables in logs.

Assumption 1: Dynamic invertibility. The vector moving average process $\{\boldsymbol{x}_{i,t}\}$, $\boldsymbol{x}_{i,t} \equiv \left[q_{i,t}, y_{i,t}, p_{i,t}, \ell_{i,t}, m_{i,t}, k_{i,t}, u_{i,t}\right]^{\top}$ is invertible with respect to $\varepsilon^{a}_{i,t}$, i.e.:

$$\varepsilon^{a}_{i,t} \in \operatorname{span}\left(\boldsymbol{x}_{i,\tau},\ -\infty < \tau \leq t\right)$$

Assumption 1 is the standard invertibility condition for vector autoregression models: structural shocks can be recovered from the entire history of observables in the system. An econometrician observing $\{\boldsymbol{x}_{\tau}\}_{\tau<t}$ can therefore predict as well as an agent inside the model who observes the structural shock history $\{\varepsilon^{a}_{i,\tau}\}_{\tau<t}$. The requirement of dynamic invertibility replaces the invertibility requirement in the control-function approach, which is often assumed to be static, as it uses $\boldsymbol{x}_{i,t}$ to recover productivity shocks, and imposes a narrower information set, as it uses only a restriction on the variables in $\boldsymbol{x}_{i,t}$, implied by an optimality condition, as opposed to letting

the data select which variables to include.

Although standard in the VAR literature, invertibility is a substantive assumption. It fails when firms have information about future productivity growth that the observables do not capture: for example, a new invention that has not yet entered intangible capital or affected past decisions on investment, sales, or prices. Two observations are useful. First, when available, information on firms' expectations (Keiller et al., 2024) can be added to the observables to address invertibility concerns within the same identification strategy. Second, with one external instrument for each input, invertibility can be relaxed: these instruments address both invertibility and identification by allowing the production function parameters to be estimated directly from the production function identity (Equation 1). As the analysis below clarifies, however, lags of observables must enter as controls and therefore cannot also serve as instruments, complicating the search for valid instruments. Without external instruments, invertibility is necessary for identification.

Assumption 2: Existence of a balanced growth path. There exists a balanced growth path with input shares that are constant over time and homogeneous across firms.

We derive the identification result below from a first-order expansion of the system in levels around a balanced growth path. Assumption 2 ensures that the path exists and that elasticities along it are constant over time and homogeneous across firms.[4]

Assumption 3: Rank condition. The system of variables $\boldsymbol{x}_{i,t}$ has exactly two independent unit roots.

Assumption 3 specifies two independent unit roots, or two common stochastic trends, which we associate with capital and the nominal price level. Section 5 provides evidence supporting this assumption. For a system of $n = 7$ variables with $r = 2$ independent unit roots, the rank is $n-r = 5$: five linear combinations of the variables are stationary. Theory supplies three non-redundant restrictions: the production function (Equation 1), the first-order condition for relative inputs (from Equation 6), and the first-order condition for prices (Equation 8). The next assumption supplies the remaining two stationarity conditions.

Assumption 4: Stationary variables. Inventories and utilization are stationary:

$$q_{i,t} - y_{i,t} \sim I(0)$$

$$u_{i,t} \sim I(0)$$

[4]Heterogeneity across groups, such as industries, and over time can be accommodated with Bayesian methods or estimating parameters in cross-sectional subgroups and subsamples. We abstract from reporting these extensions and focus on the behavior of the median firm. Empirical explorations along these directions indicate that none of the results in Sections 7 and 8 change across subgroups ordered by size.

Assumption 4 determines the dynamics of the system along the remaining cointegrating directions: inventories converge in the long run to a firm-specific buffer. This statistical requirement can be motivated by the use of inventories to smooth shocks, with no time-varying uncertainty along the balanced growth path (a non-stochastic trend and constant variance of idiosyncratic shocks). Section 5 provides support for this assumption.

3.1 Identification of the technology parameters

The central methodological result identifies the production function coefficients in the general framework discussed. Let N denote the number of firms and T the number of periods in the sample. The consistency results apply when both N and T are large. Let θ^x denote the output elasticities of inputs $x \in \{L, M, K\}$ along the balanced growth path. Let $\Theta \equiv (\overline{WL})/(\overline{WL} + \overline{P^m M}) \in (0, 1)$ denote labor's share of total variable cost along that path. Finally, let $\Omega \equiv \Gamma/(1+\Gamma) \in [0, 1)$ denote the degree of strategic complementarities in pricing, where $\Gamma \geq 0$ is minus the derivative of the log desired gross markup (Equation 5) with respect to the log relative price.

Theorem 1 (VECM representation)**.** *Let $\boldsymbol{x}_{i,t} \equiv \left[q_{i,t}, y_{i,t}, p_{i,t}, \ell_{i,t}, m_{i,t}, k_{i,t}, u_{i,t}\right]^\top$ be the vector of endogenous observables. Under Assumptions 1-4, the system satisfies a reduced-form vector error correction representation (VECM) up to a first-order approximation:*

$$\Delta \boldsymbol{x}_{i,t} = \boldsymbol{\gamma_i} + \boldsymbol{\gamma_t} + \Lambda \delta^\top \boldsymbol{x}_{i,t-1} + B(\mathsf{L})^\top \Delta \boldsymbol{x}_{i,t-1} + \boldsymbol{\epsilon}_{i,t} \tag{9}$$

where $\boldsymbol{\gamma_i}, \boldsymbol{\gamma_t} \in \mathbb{R}^7$ are firm and industry-by-time fixed effects, respectively; $\Lambda \in \mathbb{R}^{7\times 5}$ is the loading matrix; $\delta \in \mathbb{R}^{7\times 5}$ is the cointegrating matrix; and $B(\mathsf{L})$ is a finite-order lag polynomial matrix. The vector of reduced-form innovations, $\boldsymbol{\epsilon}_{i,t}$, depends linearly on the structural shocks. Up to scaling, the system satisfies the following cointegrating relationships:

$$\delta^\top \boldsymbol{x} = \begin{bmatrix} (1-\Omega)(\Theta\ell + (1-\Theta)m - y) + (1-\Omega)\frac{\eta\theta^K}{\nu}(u + k - \Theta\ell - (1-\Theta)m) - p \\ \theta^K(u+k) + \theta^L\ell + \theta^M m - q \\ m - \ell \\ q - y \\ u \end{bmatrix}$$

The matrix of coefficients $\Lambda\delta^\top$ can be consistently estimated with ordinary least squares. Furthermore, an estimate of δ recovers all the parameters of the production function $\alpha_K, \alpha_L, \alpha_M, \nu$, and η.

Proof. See Appendix A.1. □

Theorem 1, the paper's main methodological contribution, fully characterizes the system's joint dynamics to first order and therefore the dynamic effects of TFPQ, discussed in the next subsection. Its key implication for identification is that two linear regressions, linked by cross-equation parameter restrictions, suffice to identify the production function in the presence of wedges and endogenous productivity. The following corollary states this result before we discuss the proof.

Corollary 1 (Identification of technology)**.** *The long-run pricing equation is:*

$$\begin{aligned} p_{i,t} = \gamma_i^p \; &+ (1-\Omega)\big(\Theta(w_t + \ell_{i,t}) + (1-\Theta)(p_t^m + m_{i,t}) - y_{i,t}\big) + \Omega p_{-i,t} \\ &+ (1-\Omega)\frac{\eta\theta^K}{\nu}(u_{i,t} + k_{i,t} - \Theta\ell_{i,t} - (1-\Theta)m_{i,t}) + \omega_{i,t}^p \end{aligned} \tag{10}$$

The ordinary least-squares estimator of Ω, $\Theta = \theta^L/\big(\theta^L + \theta^M\big)$, *and* $\eta\theta^K/\nu$ *is super-consistent. The short-run output equation is:*

$$\begin{aligned} \Delta q_{i,t} = \gamma_i^q + \gamma_t^q - \Lambda_{1,1}\omega_{i,t-1}^p + \Lambda_{1,2}\big(\theta^K(u_{i,t-1} + k_{i,t-1}) + \theta^L\ell_{i,t-1} + \theta^M m_{i,t-1} - q_{i,t-1}\big) \\ + \Lambda_{1,3}(m_{i,t-1} - \ell_{i,t-1}) + \Lambda_{1,4}(q_{i,t-1} - y_{i,t-1}) + \Lambda_{1,5}u_{i,t-1} + \sum_{x\in\boldsymbol{x}}\sum_{\tau=1}^{p} b_{\tau,x}\,\Delta x_{i,t-\tau} + \epsilon_{i,t}^q \end{aligned} \tag{11}$$

The ordinary least-squares estimator is consistent. The two-equation system (10-11), subject to nonlinear parameter restrictions, consistently recovers the production function parameters (α_K, α_L, α_M, ν, *and* η*) through:*

$$\frac{\alpha_K}{\alpha_L} = \frac{\theta^K}{\theta^L}\left(\frac{\mathbb{E}(L_{i,t})}{\mathbb{E}(u_{i,t}K_{i,t})}\right)^{\eta},$$

$$\frac{\alpha_M}{\alpha_L} = \frac{\theta^M}{\theta^L}\left(\frac{\mathbb{E}(L_{i,t})}{\mathbb{E}(M_{i,t})}\right)^{\eta},$$

$$\alpha_K + \alpha_L + \alpha_M = 1$$

$$\nu = \theta^K + \theta^L + \theta^M$$

where $\mathbb{E}(\cdot)$ *denotes the unconditional average across both time and firms.*

Proof. See Appendix A.2. □

The long-run pricing equation (10) is the price-cost cointegrating relationship. Up to a first-order approximation, it generalizes the first-stage input share equation discussed in Gandhi et al. (2020). Without strategic complementarities ($\Omega = 0$), it can be rearranged as a regression of revenues ($p_{i,t} + y_{i,t}$) on input expenditures. In the Cobb-Douglas limit ($\eta \to 0$), the coefficients on wage bill ($w_t + \ell_{i,t}$) and intermediate expenditures ($p_t^m + m_{i,t}$) sum to (minus) the coefficient on sales ($y_{i,t}$), which identifies the degree of strategic complementarities jointly with competitors' prices

($p_{-i,t}$). As both expenditures and competitors' prices are non-stationary, the estimates of Ω and Θ are super-consistent (see Section 5).[5] The remaining coefficient ($\eta\theta^K/\nu$) is super-consistently identified by capital, which has an independent stochastic trend (Section 5).

The short-run output equation (11) describes the dynamics of output in response to temporary disturbances that perturb the system away from its balanced growth path. The loadings Λ capture how rapidly output mean-reverts towards the long run when the disturbances originate from each wedge in the first-order conditions. A large coefficient $\Lambda_{1,1}$ would indicate that departures from flexible prices would trigger immediate adjustments in production, whereas a small $\Lambda_{1,3}$ implies that any wedge in inputs (e.g., firm-specific input prices) does not affect production decisions. $\Lambda_{1,2}$ captures mean-reversion in productivity, $\Lambda_{1,4}$ the effect of inventories in production, $\Lambda_{1,5}$ controls for utilization. Importantly, the elasticities of output to the lagged levels of inputs (θ^X) are identified from the short-run dynamics from the ratios of coefficients on the lagged levels (input divided by output); hence, they are invariant to the estimates of the loadings —and thus invariant to scaling. Their estimate is consistent, but not super-consistent, because real output is a stationary variable and thus the production function is not, fundamentally, a cointegrating relationship. The latter is an alternative way of interpreting the classical problem in production function estimation, a.k.a. the transmission bias (Marschak and Andrews Jr, 1944).

We stress the importance of controlling for lags in the short-run dynamics ($b_{\tau,x} \neq 0$). In the model, lags control for the endogenous evolution of productivity. In the empirical applications, lags also partially remove residual persistence in the reduced-form innovations (e.g., MA(1) shocks) and permit constructing impulse responses that have comparable theoretical properties to linear projections (Plagborg-Møller and Wolf, 2021; Montiel Olea et al., 2026). Finally, the control-function approach of Gandhi et al. (2020) assumes that the pricing wedge recovers a transitory component of productivity that affects output symmetrically to the persistent component of productivity; it is nested as a special case under the restriction $\Lambda_{1,1} = 1 - \Lambda_{1,2}$. We test this restriction in Section 6.

[5] In error-correction models, long-run cointegrating relationships can typically be estimated by ordinary least squares despite misspecification of short-run dynamics, persistent errors, and correlation between residuals and regressors. Super-consistency is the reason: along non-stationary directions, OLS is consistent at rate $T\sqrt{N}$ rather than the usual $\sqrt{T \cdot N}$. Under maintained Assumption 3, the wage bill, intermediate expenditure, and competitors' price index carry the nominal stochastic trend, so the pricing-equation coefficients lie along non-stationary directions. See Gagliardone et al. (2025) and Gagliardone and Tielens (2026) for approaches that estimate the structural form directly without relying on super-consistency.

Discussion of the proof of Theorem 1. We log-linearize the model around the balanced growth path (Assumption 2). The production function (Equation 1) becomes:

$$q_{i,t} = a_{i,t} + \theta^K(u_{i,t} + k_{i,t}) + \theta^L \ell_{i,t} + \theta^M m_{i,t}. \tag{12}$$

where $\theta^x \equiv \partial q_{i,t}/\partial x_{i,t}$, $x \in \{K, L, M\}$, are the output elasticities of capital, labor, and intermediates along the balanced growth path, and $\nu = \theta^K + \theta^L + \theta^M$ gives returns to scale. Constant, homogeneous input shares along the path imply that θ^x has no firm or time subscript.[6] Combining this expression with productivity dynamics (Equation 2) yields the structural error-correction equation for output:

$$\begin{aligned}\Delta q_{i,t} = \widetilde{\gamma}_i^q + \widetilde{\gamma}_t^q + (1-\rho)\Big(\theta^K(u_{i,t-1} + k_{i,t-1}) + \theta^L \ell_{i,t-1} + \theta^M m_{i,t-1} - q_{i,t-1}\Big) \\ + \sum_{x \in \boldsymbol{x}} \sum_{\tau=0}^{p} \widetilde{\psi}_{\tau,x}\, \Delta x_{i,t-\tau} + \varepsilon_{i,t}^a\end{aligned} \tag{13}$$

where $\widetilde{\gamma}_i^q \equiv (1-\rho)\gamma_i^a$, $\widetilde{\gamma}_t^q \equiv \gamma_t^a - \rho\gamma_{t-1}^a$, $\widetilde{\psi}_{0,x} \equiv \theta^x$ for $x \in \{K, L, M\}$, $\widetilde{\psi}_{0,U} \equiv \theta^K$, and $\widetilde{\psi}_{\tau,x} \equiv \psi_{\tau,x}$ otherwise. Note that the output growth rate depends on contemporaneous inputs.

We next derive the structural representation of the system. Under Kimball demand and the forms of competition discussed, the log-linearized markup (Equation 5) decreases with the relative price (Amiti et al., 2019; Gagliardone et al., 2025):

$$\mu_{i,t} - \mu_i = -\Gamma(p_{i,t} - p_{-i,t} - \varphi_{i,t}), \qquad \Gamma \geq 0. \tag{14}$$

where μ_i is the firm's markup along the balanced growth path (homogeneous when sales shares are homogeneous), and Γ is the markup elasticity with respect to relative prices, which depends on the residual demand elasticity. The competitors' price index, $p_{-i,t}$, coincides with the industry price index ($p_{-i,t} = p_t$) when firms have zero mass within their industry. Substituting Equation (14) into the flexible-price first-order condition in logs (Equation 8) and rearranging gives:

$$p_{i,t}^\star = (1-\Omega)(\mu_i + mc_{i,t}^\star) + \Omega(p_{-i,t} + \varphi_{i,t}) \tag{15}$$

The coefficient $\Omega \equiv \Gamma/(1+\Gamma) \in [0, 1)$ measures strategic complementarities: markup adjustment dampens the pass-through of costs to prices. Log-linearizing marginal cost in Equation (7) gives:

$$\begin{aligned}mc_{i,t}^\star = \gamma_i^{mc} + \Theta(w_t + \ell_{i,t}) + (1-\Theta)(p_t^m + m_{i,t}) - q_{i,t} \\ + \frac{\eta\theta^K}{\nu}(u_{i,t} + k_{i,t} - \Theta\ell_{i,t} - (1-\Theta)m_{i,t})\end{aligned} \tag{16}$$

[6] Nonlinearities, heterogeneity, and time variation in parameters can all be theoretically handled within our econometric framework. Nonlinearities would require adding polynomials of $\boldsymbol{x}_{i,t}$ to the system. Heterogeneity would require estimating parameters across subgroups. Time variation can be accommodated by recursive estimation. Bayesian methods can also be used to discipline the estimates.

where $\Theta \equiv (\overline{WL})/(\overline{WL} + \overline{P^m M}) = \theta^L/(\theta^L + \theta^M)$ is labor's share of total variable cost along the balanced growth path. Under CES ($\eta \neq 0$), the capital elasticity enters because the labor and intermediate elasticities ($\theta^L_{i,t}, \theta^M_{i,t}$) depend on capital. In the Cobb-Douglas limit ($\eta \to 0$), marginal cost reduces to a linear combination of the wage bill and intermediate costs, minus production.

Substituting Equation (16) into Equation (15) and collecting the stationary terms $\Omega\varphi_{i,t} - (1-\Omega)(q_{i,t} - y_{i,t})$ in $\omega^p_{i,t}$ yields the long-run relation in Equation (10), up to that stationary wedge:

$$\begin{aligned} p^\star_{i,t} = \gamma^p_i + \gamma^p_t + (1-\Omega)(\Theta\ell + (1-\Theta)m_{i,t} - y_{i,t}) \\ + (1-\Omega)\frac{\eta\theta^K}{\nu}(u_{i,t} + k_{i,t} - \Theta\ell_{i,t} - (1-\Theta)m_{i,t}) \end{aligned}$$

The firm fixed effect $\gamma^p_i \equiv (1-\Omega)(\mu_i + \gamma^{mc}_i)$ depends on the short-run output elasticity of flexible inputs and the steady-state markup. The time fixed effect $\gamma^p_t \equiv (1-\Omega)(\Theta w_t + (1-\Theta)p^m_t) + \Omega p_{-i,t}$ combines aggregate input prices and the competitors' price index. The pricing wedge $p^\star_{i,t} - p_{i,t} \sim I(0)$ is stationary under the assumed forms of nominal rigidity, providing the first cointegrating direction. The second is the Solow residual, stationary by Equation (2). Taking the ratio of the flexible first-order conditions in Equation (6) and then taking logs gives the third, $(1-\eta)(m_{i,t} - \ell_{i,t}) = 0$. Assumption 4 supplies the last two. These five non-redundant directions exhaust the system's rank (seven variables and two unit roots) and span the space orthogonal to the unit-root directions. The system therefore admits the structural VECM representation:

$$A_0 \Delta \boldsymbol{x}_{i,t} = \widetilde{\boldsymbol{\gamma}}_i + \widetilde{\boldsymbol{\gamma}}_t + \widetilde{\Lambda}\delta^\top \boldsymbol{x}_{i,t-1} + \widetilde{B(\mathsf{L})}^\top \Delta \boldsymbol{x}_{i,t-1} + \boldsymbol{\varepsilon}_{i,t} \tag{17}$$

where $A_0 \in \mathbb{R}^{7\times 7}$ is the structural impact matrix, δ is the same cointegrating vector of Theorem 1, and $\boldsymbol{\varepsilon}_{i,t} = \left[\varepsilon^a_{i,t}, \varepsilon^\varphi_{i,t}\right]$ is the vector of structural disturbances. Note that the output equation does not depend on the pricing residual in the structural VECM: $\widetilde{\Lambda}_{1,j} = 0$ for all $j \neq 2$.

Dependence on the pricing wedge—the first-stage residual in the control function approach—arises when we invert A_0 under Assumption 1 to obtain the reduced-form representation from Theorem 1:

$$\Delta \boldsymbol{x}_{i,t} = \boldsymbol{\gamma}_i + \boldsymbol{\gamma}_t + \Lambda\delta^\top \boldsymbol{x}_{i,t-1} + B(\mathsf{L})^\top \Delta \boldsymbol{x}_{i,t-1} + \boldsymbol{\epsilon}_{i,t}$$

Because A_0 cannot be lower triangular (as $\theta^x > 0$), the loading $\Lambda_{1,j} \equiv (A_0^{-1}\widetilde{\Lambda})_{1,j} \neq 0$ is generally nonzero for each j. The reduced-form output equation therefore depends on the other cointegrating residuals, including the pricing wedge. Inverting the structural matrix A_0 changes the loading matrix from $\widetilde{\Lambda}$ to Λ, so the structural loadings cannot be recovered from the reduced form. It leaves δ unchanged, however. The matrix $\Lambda\delta^\top$ can therefore be recovered by OLS from the reduced-form specification, and the cointegrating vector δ can be consistently estimated from the short-run dynamics and super-consistent estimates of the pricing equation (10). The estimate

of δ recovers the elasticities θ^x, which can then be inverted to recover the structural parameters α_x, $x \in \{K, L, M\}$. This concludes the discussion of the proof.

3.2 Identification of the IRFs to productivity and demand shocks

The VECM representation in Theorem 1 characterizes the joint dynamics of the endogenous firm-level variables. By the definition of TFPQ, its first-order dynamics follow from the linear combination:

$$a_{i,t} = \begin{bmatrix} 1 & 0 & 0 & -\theta^L & -\theta^M & -\theta^K & -\theta^K \end{bmatrix} \boldsymbol{x}_{i,t}$$

The next proposition applies the same linear combination to the reduced-form innovations. As this transformation removes the contribution of shocks other than the structural productivity shock, it identifies its impulse response function from the reduced-form innovations.

Proposition 1 (IRF to a TFPQ shock)**.** *Denote by* $\boldsymbol{e}_a := \begin{bmatrix} 1 & 0 & 0 & -\theta^L & -\theta^M & -\theta^K & -\theta^K \end{bmatrix}^\top$ *the selection vector of productivity. The following linear combination of reduced-form innovations recovers the structural productivity shock up to a multiplicative constant:*

$$\boldsymbol{e}_a^\top \boldsymbol{\epsilon}_{i,t} = \varepsilon^a_{i,t}$$

Therefore, the IRF to a productivity shock is identified from a Cholesky ordering.

Proof. See Appendix A.3. □

Proposition 1 identifies the dynamic effects of productivity on firms' decisions. In the empirical analysis in Section 7, we also make the standard assumption that capital is predetermined within the period. This yields the partial Cholesky ordering $\widetilde{\boldsymbol{x}}_{i,t} = \begin{bmatrix} k_{i,t}, a_{i,t}, \dots \end{bmatrix}^\top$, which identifies the IRF of TFPQ. The identification follows from the model's structure and requires no instruments.

Section 7 also studies demand shocks, identified from movements in total factor productivity of revenue (TFPR) that TFPQ does not explain. We accordingly interpret the demand shock $\varepsilon^\varphi_{i,t}$ as a "sales" shock that changes revenues while holding productivity fixed. Define TFPR as

$$r_{i,t} := p_{i,t} + y_{i,t} - \theta^K(u_{i,t} + k_{i,t}) - \theta^L \ell_{i,t} - \theta^M m_{i,t}$$

We use the partial Cholesky ordering

$$\widetilde{\boldsymbol{x}}_{i,t} = \begin{bmatrix} k_{i,t}, a_{i,t}, r_{i,t}, \dots \end{bmatrix}^\top \longrightarrow \begin{bmatrix} \varepsilon^a_{i,t}, & \varepsilon^\varphi_{i,t} \end{bmatrix}$$

to recover the dynamic effects of productivity and demand shocks. The mapping from the original system $\boldsymbol{x}_{i,t}$ to $\widetilde{\boldsymbol{x}}_{i,t}$ is linear and invertible, so it preserves the estimated reduced-form dynamics.

4 Data and measurement

4.1 Data sources

We construct a firm-level dataset for Belgian manufacturing from 1999 to 2023 at business-cycle frequency. It combines five confidential datasets on sales, production, intermediate inputs, prices, inventories, and capacity utilization.

First, the Belgian PRODCOM survey records quarterly sales in nominal values and quantities, and production in nominal values, for each narrowly defined manufacturing product (8-digit product codes). We use this information to construct firm-level prices and production, and obtain similar data on foreign competitors from Belgian customs declarations. Second, firms' value-added tax (VAT) declarations to the fiscal authorities provide administrative data on quarterly purchases of intermediates (materials and services) and investment goods for the universe of firms. Third, social security declarations provide quarterly labor costs, measured by the wage bill. Finally, the quarterly Business Survey of the National Bank of Belgium (NBB-BS) records the share of production capacity used by a representative panel of Belgian manufacturers.

4.2 Measurement

We now map the model variables to the data, using i for firms and t for quarters. Appendix B details the data cleaning procedures.

Output prices and competitors' prices. Output prices are obtained from PRODCOM. Following Gagliardone et al. (2025), we construct a firm-level price index at the 4-digit NACE level, $P_{i,t} \equiv \exp(p_{i,t})$, using a Törnqvist index of product unit values (8-digit level). For each product j, its unit value $P_{i,j,t}$ equals sales of product j divided by the physical quantity sold. Let $\mathcal{P}_{i,t}$ denote the set of products in quarter t and $s_{j,i,t}$ the within-firm sales share of product j.[7] Weighting by the average sales share in quarters t and $t-1$, denoted by $\bar{s}_{j,i,t}$, gives the quarterly change in the firm's price:

$$\frac{P_{i,t}}{P_{i,t-1}} = \prod_{j \in \mathcal{P}_{i,t}} \left(\frac{P_{i,j,t}}{P_{i,j,t-1}} \right)^{\bar{s}_{i,j,t}}, \qquad \bar{s}_{i,j,t} = \frac{s_{i,j,t} + s_{i,j,t-1}}{2}.$$

We recover the firm price index level by chaining these quarterly changes.

In a similar vein, we develop a competitors' price index, $P_{-i,t} \equiv exp(p_{-i,t})$, by chaining a

[7] The survey does not require firms to distinguish production and sales by domestic versus international customers. We therefore recover domestic sales values and quantities by combining PRODCOM with international trade data on firms' product-level exports (quantities and sales). We map product codes across revisions to calculate consecutive-quarter price changes.

quarterly index of competitors' prices:

$$P_{-i,t}/P_{-i,t-1} = \prod_{k \in \mathcal{S}_i} \left(P_{k,t}/P_{k,t-1}\right)^{\bar{s}^{-i}_{k,t}}.$$

Here, $\bar{s}^{-i}_{k,t} := \frac{1}{2}\left(\frac{s^{-i}_{k,t}}{1-s^{-i}_{k,t}} + \frac{s^{-i}_{k,t-1}}{1-s^{-i}_{k,t-1}}\right)$ is a Törnqvist weight: the average of competitor k's revenue share in the industry of firm i at time t (net of firm i's revenues) and time $t-1$. The set of competitors to firm i, $\mathcal{S}_i$, includes both Belgian manufacturers in the industry and foreign manufacturers selling the same goods to Belgian customers.

Sales, production, and inventories. For firm i in quarter t, the sales quantity index $Y_{i,t} \equiv \exp(y_{i,t})$ equals the quarterly value of goods sold divided by the firm's price index. Similarly, the physical quantity produced by firm i in quarter t, $Q_{i,t} \equiv \exp(q_{i,t})$, equals the value of production in quarter t (reported in PRODCOM) deflated by the same price index. The difference $q_{i,t} - y_{i,t}$ then measures the change in the real stock of inventories.

Labor, capital, and intermediate inputs. Labor input, $L_{i,t} \equiv \exp(\ell_{i,t})$, is the wage bill in quarterly Social Security declarations deflated by a sectoral wage index. It measures labor cost at constant prices, not hours or headcount. Intermediate inputs, $M_{i,t} \equiv \exp(m_{i,t})$, are quarterly purchases of materials and services from VAT declarations, deflated by a sectoral intermediate-input price index. We construct the stock of fixed tangible assets, $K_{i,t} \equiv \exp(k_{i,t})$, by the perpetual inventory method following Lenzu et al. (2026), using quarterly investment flows from VAT declarations deflated by an investment price index. While we deflate labor, intermediates, and investment at 2-digit industry levels, every regression downstream also includes 4-digit industry-by-quarter fixed effects; hence, we effectively deflate at a narrower granularity level in regressions.

Utilization. Capital utilization, $U_{i,t}$, is the percentage of production capacity used by firm i in quarter t, as reported in the National Bank of Belgium Business Survey (NBB-BS). Like the US Census Quarterly Survey of Plant Capacity Utilization, the NBB-BS defines capacity as the maximum output feasible under a realistic work schedule, allowing for variation in working days within the month and assuming sufficient inputs to operate the installed capital.

4.3 Summary statistics

The full integrated dataset contains 6,011 firms, spans $T = 116$ quarters from 1995Q1 to 2023Q4, and comprises 248,515 firm-quarter observations. The firms operate in 177 manufacturing industries (4-digit NACE codes) across 18 2-digit NACE manufacturing divisions. We group these

industries into 10 manufacturing sectors using Eurostat's A*38 classification. Table 1 summarizes the full dataset: Panel A describes the panel structure, Panel B labor and intermediate expenditures relative to turnover, and Panel C capacity utilization, capital, and changes in inventories.

A few observations are in order. First, the median firm is observed for 56 quarters, and the median longest consecutive run is 48 quarters. This long time dimension is important for identification (Theorem 1 relies on large T). Second, the dataset spans the entire size distribution of firms; the median firm has a quarterly turnover of 1.7 million euros, but the average firm is five times larger. Third, the median intermediates-to-turnover ratio of 0.62 is consistent with revenue elasticities estimated for Belgian manufacturing industries (De Loecker et al., 2014; Lenzu et al., 2026). Intermediates account for an average of 78 percent of the variable-input bill, close to the 75 percent reported by Gagliardone et al. (2025). The median wage-bill-to-turnover ratio is 0.16. Third, capacity utilization averages 77 percent, with a median of 79 percent. These levels and the substantial cross-sectional variation closely match the U.S. evidence (Boehm and Pandalai-Nayar, 2022). Finally, significant variation in inventory drawdowns and additions suggest it to be an active lever to accommodate short-run mismatches between production and sales.

Table 1: Summary statistics

	Mean	5^{th} pctle	25^{th} pctle	Median	75^{th} pctle	95^{th} pctle
Panel A: Sample structure						
Number of quarters in sample	60.96	16.00	36.00	56.00	88.00	116.00
Longest run of consecutive quarters in sample	55.27	12.00	31.00	48.00	76.00	116.00
Panel B: Input expenditures and turnover						
Turnover (million euros)	9.65	0.24	0.75	1.70	4.63	29.93
Wage bill / turnover	0.18	0.03	0.09	0.16	0.25	0.42
Intermediates / turnover	0.63	0.30	0.47	0.62	0.77	1.00
(Wage bill + intermediates) / turnover	0.82	0.43	0.66	0.81	0.95	1.22
Wage bill / (wage bill + intermediates)	0.22	0.04	0.12	0.21	0.31	0.48
Panel C: Utilization, capital, and inventories						
Capacity utilization (percent)	77.06	52.92	71.05	78.91	84.86	96.00
Capital / annual turnover	0.43	0.07	0.19	0.34	0.57	1.12
Change in inventories / turnover	0.03	−0.67	−0.06	0.00	0.13	0.80

Notes. The unit of observation is a VAT identifier in a quarter, with 224,866 observations. A firm in this table is a VAT identifier. All statistics are unweighted. The number of quarters in the sample counts all observed quarters, which need not be consecutive.

5 Evidence on the model specification

The identification strategy in Corollary 1 relies on (i) the super-consistency of the estimator of the long-run pricing equation, (ii) stationarity of the cointegrating residuals, (iii) rank five of the system (Assumption 3). This section provides empirical validation for these requirements.

The argument has five steps. First, we show that the output price ($p_{i,t}$), wage bill ($w_t + \ell_{i,t}$), the competitors' price index ($p_{-i,t}$), and capital ($k_{i,t}$) are non-stationary. Second, we associate their independent stochastic trends with the nominal price level (a common unit root to prices and

nominal expenditures) and capital. Third, we show that the cointegrating residuals are stationary and provide formal tests for cointegration. Fourth, we provide formal tests for rank selection of the system. Fifth, we compare the unit roots of the error-correction model to those of the associated vector autoregression in levels. The evidence is consistent with the stated assumptions and supports the empirical specifications. Appendix E reports the tests.

Unit-root tests on the variables— Figure 2 reports the cross-firm distribution of sample autocorrelations in log levels net of firm-by-quarter-of-year means. Prices are highly persistent: after eight quarters, the median autocorrelation is 0.50 for own prices and 0.62 for competitors' prices; after twenty-four quarters, it remains 0.18 and 0.28, respectively. The separately estimated AR(1) coefficient of competitors' prices is 0.997. Neither IPS with firm intercepts nor Fisher–ADF rejects a unit root in either price series (Table A.1 in Appendix E.2).

Capital is also highly persistent. Its median autocorrelation is 0.70 after eight quarters, and its AR(1) coefficient is 0.988. Even the 25th percentile of its sample autocorrelations remains above the figure's upper white-noise reference bound through sixteen quarters. IPS, Fisher–ADF, and cross-sectionally augmented IPS (CIPS) fail to reject a unit root.

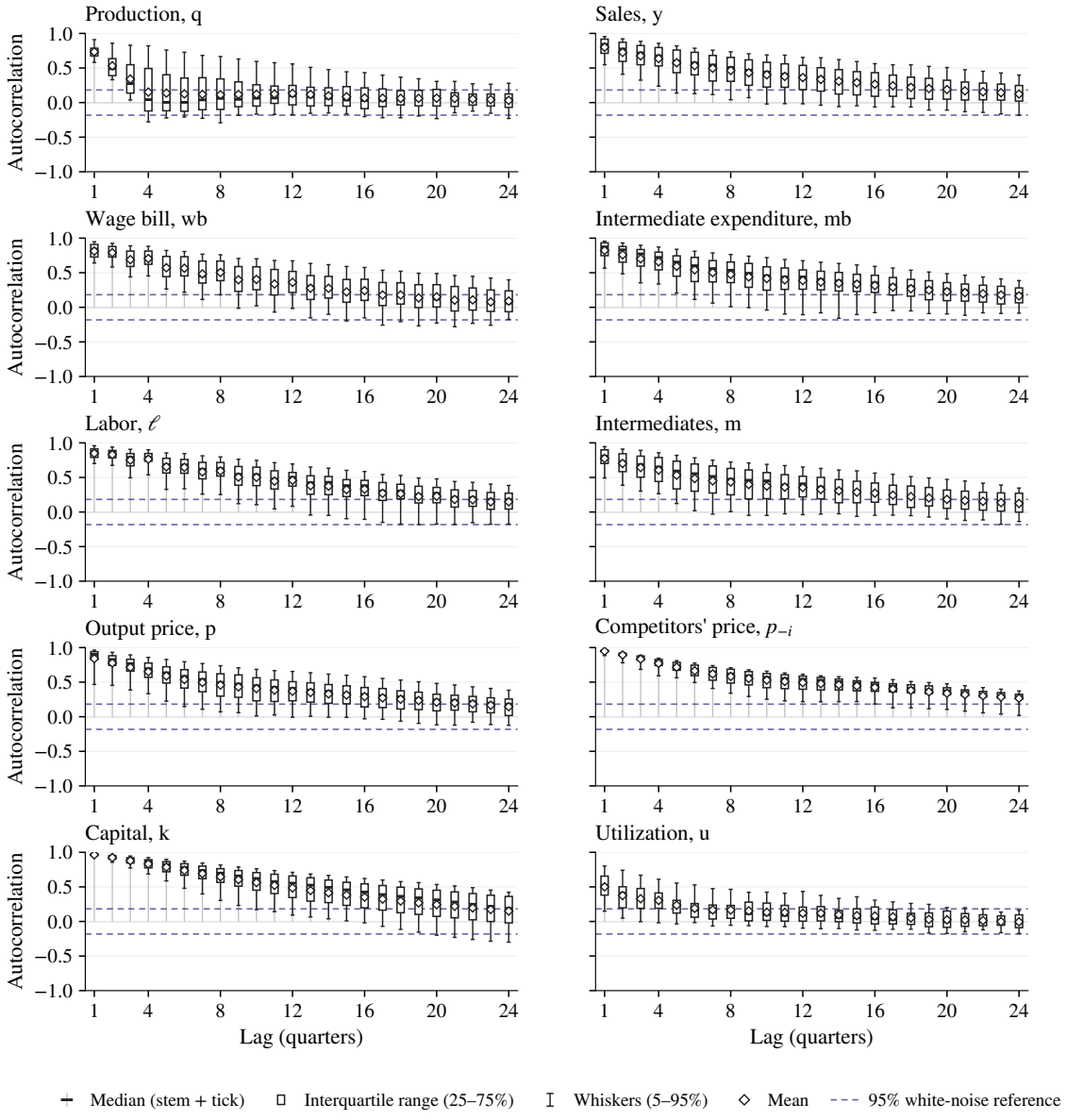
Nominal expenditures display substantial serial dependence: their eight-quarter median autocorrelations are 0.53 for the wage bill and 0.54 for intermediate expenditure. For the wage bill, the cross-sectionally augmented IPS (CIPS) test does not reject a unit root. This test allows for a common factor through cross-sectional averages, and its nonrejection supports treating the wage bill as nonstationary. For intermediate expenditure, all statistical tests reject at the 1 percent level. This indicates that, while theory is consistent with both expenditures being non-stationary, only the wage bill is detected to be non-stationary in the finite sample.

All tests reject the unit root for production, sales, real labor, real intermediates, and utilization at the 1 percent level; their eight-quarter median autocorrelations are 0.06, 0.51, 0.63, 0.48, and 0.16, respectively. Therefore, the real variables are stationary as predicted from theory. In first differences, all three tests reject a unit root in all ten series, using three lagged differences and firm intercepts. Hence, we exclude higher orders of integration.

These results are consistent with assigning stochastic trends to nominal prices and capital. The number of independent trends, and hence the system's rank, is assessed separately through the long-run relations and joint rank tests below.

Unit-root tests on the long-run relations— We test five theoretical long-run relations for stationarity: the pricing relation (pricing wedge, $-\omega^{p}_{i,t}$); the production function residual (productivity, $-a_{i,t}$); relative real inputs ($m_{i,t} - \ell_{i,t}$); inventories ($q_{i,t} - y_{i,t}$); and capacity utilization ($u_{i,t}$). Panel unit-root tests indicate mean reversion in these relations (Table A.2 in Appendix E.2).

Figure 2: Autocorrelation of the observed variables



Notes. Sample autocorrelations at lags of one to 24 quarters, computed separately for 101 firms observed in all 116 quarters from 1995Q1 to 2023Q4 with the nine variables and the competitors' price: production, sales, the wage bill, intermediate expenditure, labor, intermediates, the output price, the competitors' price, capital and utilization. The variables are log levels net of firm-by-quarter-of-year means. At each lag, the stem and horizontal tick mark the cross-firm median, the box spans the 25th–75th percentiles, the hollow diamond marks the mean, and the whiskers span the 5th–95th percentiles. Firms receive equal weight. The dashed lines are the approximate pointwise 95 percent white-noise reference bounds.

After industry-by-quarter means are removed, all tests reject a unit root in all five relations at the 5 percent level (the largest IPS, Fisher–ADF, and CIPS statistics across the five relations are −4.798, −4.981, and −2.113, respectively). The autocorrelation function plots also confirm strong mean reversion for the entire distribution (at twenty-four quarters, all five medians lie within 0.06 of zero, and the 75th and 95th percentiles are below 0.19 and 0.38, respectively; Figure A.2 in Appendix E.2). The estimated system's five error-correction terms approximate these relations and have half-lives of one to four quarters (Table A.5, Panel A; Section 6.2).

We also test for cointegration without imposing the parameter restrictions. All the Pedroni (1999) tests reject no cointegration at the 1 percent level in both raw log levels and log levels net of industry-time fixed effects (Table A.2, Panel C, in Appendix E.2).

Panel rank tests— We test for the rank of the system with the panel trace test of Larsson et al. (2001), for both the system with expenditures and the same with real inputs, for the variables in levels and in deviations from fixed effects. No test indicates a rank larger than five (Table A.3 in Appendix E.3). With real inputs, the test selects rank five in three of the four specifications. When using expenditures, the test selects rank five in two of these four specifications. Hence, the data provide strong support for a rank of five or less.

Diagnostics of the estimated system— The unrestricted VAR in levels also has near-unit roots where the restricted VECM model imposes unit roots: its largest root is 0.969, and the long-run matrix has two small singular values, 0.065 and 0.025. Together, the five stationary relations, approximated by the estimated error-correction terms, imply a rank of at least five, and no rank test selects more than five. We therefore maintain rank five throughout.

6 Estimates of the parameters

This section reports estimates of the framework in Section 2. Following Corollary 1, the estimation proceeds in two stages.

In the first stage, we estimate the cointegrating matrix δ from the two equations in Corollary 1: the long-run pricing equation (10) and the output equation (11). We estimate them jointly by nonlinear least squares, imposing the cross-equation restrictions, and also report unrestricted pricing estimates. The benchmark output equation includes eight lagged differences of ten variables ($p = 8$): the seven system variables, the implicit wage, the implicit intermediates price, and the competitors' price. In the second stage, we fix δ at its benchmark first-stage estimate and estimate each of the seven VECM equations (Equation 9) by ordinary least squares, with $p = 24$

lagged differences. We use the longer lag structure to produce impulse responses up to six years in Section 7, at the cost of a smaller sample: 80,921 firm-quarter observations on 2,526 firms. All regressions absorb firm-by-quarter-of-year fixed effects. Industry-by-quarter fixed effects enter every output equation, the pricing equations in columns (B)–(D), and all seven second-stage equations. Industries and standard-error clusters are defined at the four-digit level.

6.1 The production function

Panel A of Table 2 reports the pricing-equation estimates. Column (A) is the benchmark, estimated by nonlinear least squares under the parameter restrictions. Column (B) uses real labor and intermediates in the pricing equation and includes industry-by-quarter fixed effects. Theory supplies an additional restriction on the ratio of labor and material elasticities through long-run average input shares: $\Theta = \mathbb{E}(W_t L_{i,t})/(\mathbb{E}(W_t L_{i,t} + P_t^m W_{i,t}))$, where $\mathbb{E}(\cdot)$ denotes the unconditional average across time and firms. Column (C) adds this restriction to column (B), fixing $\Theta = .221$, the ratio of the sample means of wage bill to expenditure on intermediates. Column (D) estimates column (B)'s pricing equation on the same observations without parameter restrictions.

The benchmark estimate of strategic complementarities is $\Omega = .512$, with a standard error of .020. A firm therefore passes through 49 percent of a marginal-cost change to its price in the long run, holding competitors' prices fixed. This estimate is consistent with those obtained from the same dataset by Amiti et al. (2019), Gagliardone et al. (2025), and Gagliardone and Tielens (2026). These estimates show the importance of endogenous markups for recovering the pricing wedge correctly. The estimated degree of strategic complementarities is robust across all specifications. The baseline labor share in variable cost is $\Theta = .271$, with a standard error of .018. It exceeds the measured share of .221 imposed in column (C) but is below the estimate of .373 with industry-by-quarter fixed effects in column (B). These differences may reflect endogeneity of stationary regressors ($m - \ell$ is stationary in column B). This is addressed either by restricting the coefficients in column (C) or, under maintained Assumption 3, by relying on super-consistency of the expenditure coefficients in column (A). Finally, the benchmark estimate of the parameter combination $\eta\theta^K/\nu$ is .089, above .073 in column (B) and below .116 in column (C).

Panel B of Table 2 reports output elasticities, identified jointly by the output and pricing equations. In the benchmark (Column A), the elasticities of labor, intermediates, and capital services are $\theta^L = .195$, $\theta^M = .524$, and $\theta^K = .182$. Relative to the estimates obtained with the control-function approach of Gandhi et al. (2020) (Table A.4 in Appendix F.1), our labor elasticity is slightly below the estimate of .212, our capital elasticity is much higher than the estimate of .026, and our materials elasticity is below the estimate of .753. With comparable estimated returns

Table 2: Estimates of the long-run pricing equation and the production function

	(A) Benchmark	(B) Time FE	(C) $\Theta = .221$	(D) Unrestricted
	Panel A: long-run pricing equation			
$\ell_{i,t}$	.120	.163	.090	.150
	(.009)	(.012)	(.004)	(.013)
$m_{i,t}$	.324	.273	.318	.270
	(.015)	(.014)	(.013)	(.013)
$y_{i,t}$	-.488	-.470	-.462	-.472
	(.020)	(.020)	(.019)	(.020)
$k_{i,t}$	.043	.034	.054	.021
	(.007)	(.007)	(.008)	(.007)
$u_{i,t}$	.043	.034	.054	.036
	(.007)	(.007)	(.008)	(.008)
Ω	.512	.530	.538	.528
	(.020)	(.020)	(.019)	(.020)
Θ	.271	.373	.221	.358
	(.018)	(.021)		(.023)
$\eta\theta^K/\nu$	.089	.073	.116	
	(.013)	(.014)	(.015)	
Observations	147,244	147,308	147,308	147,308
Firm × quarter-of-year FE	✓	✓	✓	✓
Industry × quarter FE	—	✓	✓	✓
	Panel B: output equation			
θ^L	.195	.270	.162	
	(.021)	(.029)	(.016)	
θ^M	.524	.453	.573	
	(.057)	(.050)	(.057)	
θ^K	.182	.182	.170	
	(.046)	(.045)	(.045)	
	Implied structural parameters			
ν	.900	.905	.905	
	(.069)	(.069)	(.069)	
η	.438	.362	.618	
	(.116)	(.101)	(.167)	
Observations	147,244	147,308	147,308	147,308
Firm × quarter-of-year FE	✓	✓	✓	✓
Industry × quarter FE	✓	✓	✓	✓

Notes. Panel A reports long-run pricing coefficients and the structural parameters they imply. Panel B reports output elasticities and technology parameters. Columns (A)–(C) jointly estimate Equations (10) and (11) by nonlinear least squares, with eight lagged differences of $\boldsymbol{x}_{i,t}$; their output loadings appear in Panel A of Table 4. Column (D) estimates pricing alone by unrestricted ordinary least squares. All equations include firm-by-quarter-of-year effects. Industry-by-quarter effects enter pricing in (B)–(D) and every output equation (A)–(D). Column (A) uses 147,244 observations on 4,199 firms; columns (B)–(D) use 147,308 observations on 4,206 firms. Standard errors, in parentheses, are clustered across four-digit industries.

Table 3: Weights of the CES production function

	CES elasticities			*Input ratios*	
	(A) Benchmark	**(B)** Time FE	**(C)** $\Theta = .221$		
α_K	.137	.139	.109	$\overline{L}/(\overline{U}\,\overline{K})$	.108
	(.051)	(.050)	(.050)		
α_L	.388	.460	.413		
	(.069)	(.067)	(.089)		
α_M	.474	.402	.478	$\overline{L}/\overline{M}$	.165
	(.030)	(.030)	(.042)		

Notes. Columns (A)–(C) report the weights implied by Corollary 1 at the output elasticities and the CES parameter of the corresponding column of Table 2. The last column reports ratios of the inputs over the first-stage estimation sample of column (A). L is the deflated wage bill, M is deflated purchases of intermediates, and UK is capital services, the product of capacity utilization and the capital stock. Standard errors, in parentheses, are clustered across four-digit industries.

to scale, our approach thus shifts weight from labor and materials toward capital. The reason is that our approach does not require the flexible-input first-order condition to hold. The materials elasticity is therefore not directly tied to the ratio of intermediate expenditure to revenue, as it is when materials are assumed flexible in the control-function approach.

Differences in Θ across Panel A specifications affect the split between labor and material elasticities; differences in $\eta\theta^K/\nu$ affect the implied CES parameter η. Column (B) has a higher labor elasticity ($\theta^L = .270$) and a lower materials elasticity ($\theta^M = .453$) than the benchmark. Column (C) has a lower labor elasticity ($\theta^L = .162$) and a higher materials elasticity ($\theta^M = .573$). Both materials elasticities remain below the estimate of Gandhi et al. (2020). Returns to scale are similar across specifications: $\nu = .9$, with a standard error of .07. The 95 percent confidence interval, from .76 to 1.04, includes constant returns. The baseline CES parameter is $\eta = .438$, with a standard error of .116. It is statistically consistent with the other specifications: the estimate is lower in column (B) ($\eta = .36$) and higher in column (C) ($\eta = .62$), but the large standard errors imply no statistically significant differences. All specifications nevertheless imply that the inputs are gross substitutes. Using Corollary 1, we recover the CES weights α_K, α_L, and α_M from the output elasticities, the CES parameter η, and two ratios of average inputs. Table 3 reports those ratios and the implied weights. The benchmark weights are $\alpha_K = .137$, $\alpha_L = .388$, and $\alpha_M = .474$. Alternative specifications give comparable estimates that are not statistically different.

6.2 Loadings on the long-run relations

Table 4 reports the loadings Λ in Theorem 1 and Corollary 1. Each column corresponds to a long-run relation lagged one quarter: the pricing wedge, the output wedge $-a$, relative inputs $m-\ell$, inventories $q-y$, and utilization u. A loading measures the change in a variable's next-quarter growth rate associated with a one-log-point deviation of a relation from its long-run value. Panel A reports the output equation from the joint estimation in columns (A), (B), and (C) of Table 2. Panel B reports the seven VECM equations, with five error-correction terms expressed in the seven system variables at the benchmark estimates. As shares of each series' total variance, the within-industry-quarter variances are 1.2 percent for the implicit wage, 5.6 percent for the implicit intermediates price, and 2.5 percent for the competitors' price.

The output equation loads positively on the pricing wedge ($\Lambda_{1,1}$ between .105 and .132 across specifications). When costs are high relative to current prices, this is consistent with firms purchasing inputs to increase future production. The output-wedge loading is also positive ($\Lambda_{1,2}$ between .176 and .195): mean reversion implies that high productivity today anticipates a future decrease in output. The inventory-wedge loading ($\Lambda_{1,4}$ between $-.150$ and $-.213$) indicates that production adjusts to satisfy sales in the long run. The relative-input loading is not statistically significant in the benchmark or VECM, but is significant at the 5 percent level in the fixed-share specification. The utilization loading is not statistically significant. Finally, we test the control-function restriction $\Lambda_{1,1} = 1 - \Lambda_{1,2}$. The χ^2-statistic is around 250, depending on the specification, strongly rejecting the restriction.

The sales equation has loadings similar to those of output, consistent with their close comovement. The price equation loads positively on the pricing residual ($\Lambda_{3,1} = .111$): costs above current prices predict subsequent price growth. Its output-wedge loading is negative ($\Lambda_{3,2} = -.047$), as low productivity today predicts future productivity growth and hence lower prices. Relative-input deviations also matter ($\Lambda_{3,3} = -.016$): when materials are high relative to labor, the anticipated decline in material investment implies lower future prices. Both labor and materials load on the pricing wedge, and materials also load on the output wedge, with signs again explained by mean reversion. Both equations load significantly on the relative-input wedge. Capital loads on the pricing and output residuals and on utilization, but its loadings are small: none exceeds .02 in absolute value. This is consistent with capital carrying one stochastic trend.

The loadings govern adjustment toward the long-run relations. Appendix F.2 measures the persistence profile of each relation. The price and input wedges and utilization have half-lives of one quarter; the production-function relation and inventories have half-lives of four quarters.

Table 4: Loadings on the long-run relations

	$-\omega^p_{i,t-1}$	$-a_{i,t-1}$	$m_{i,t-1}-\ell_{i,t-1}$	$q_{i,t-1}-y_{i,t-1}$	$u_{i,t-1}$	$\Lambda_{1,1}=1-\Lambda_{1,2}$
	Panel A: output equation of the joint estimation, 8 lagged differences					$\chi^2(1)$
(A) $\Delta q_{i,t}$, benchmark	.108	.176	-.011	-.151	-.011	251.14
	(.028)	(.020)	(.011)	(.022)	(.037)	[<.001]
(B) $\Delta q_{i,t}$, time FE	.105	.176	.003	-.151	-.010	252.78
	(.028)	(.020)	(.011)	(.022)	(.037)	[<.001]
(C) $\Delta q_{i,t}$, $\Theta = .221$	.105	.177	-.023	-.150	-.010	249.67
	(.028)	(.020)	(.011)	(.022)	(.037)	[<.001]
	Panel B: vector error-correction model, 24 lagged differences					Within R^2
$\Delta q_{i,t}$	.132	.195	-.025	-.213	-.002	.398
	(.044)	(.027)	(.020)	(.033)	(.102)	
$\Delta y_{i,t}$	.149	.196	-.009	.217	-.038	.287
	(.033)	(.018)	(.011)	(.021)	(.027)	
$\Delta p_{i,t}$	.111	-.047	-.016	-.048	.017	.149
	(.018)	(.006)	(.006)	(.007)	(.013)	
$\Delta \ell_{i,t}$	-.016	-.001	.043	.005	-.011	.047
	(.006)	(.004)	(.004)	(.004)	(.019)	
$\Delta m_{i,t}$	-.064	-.034	-.137	-.016	.006	.241
	(.012)	(.008)	(.012)	(.009)	(.039)	
$\Delta k_{i,t}$	-.016	-.011	.004	-.013	.018	.101
	(.004)	(.003)	(.002)	(.002)	(.005)	
$\Delta u_{i,t}$	-.001	.004	.009	.006	-.462	.326
	(.006)	(.005)	(.004)	(.005)	(.034)	

Notes. Each row reports the loadings of one equation on the five long-run relations of Theorem 1, lagged one quarter, with the signs of Corollary 1: the pricing wedge $-\omega^p$, the production function $\theta^K(u+k) + \theta^L\ell + \theta^M m - q = -a$, relative inputs $m-\ell$, inventories $q-y$, and utilization u. Panel A reports the output equation of the joint estimation in columns (A) and (B) of Table 2, with eight lagged differences and 147,308 observations. Its $\chi^2(1)$ statistic tests the restriction $\Lambda_{1,1} = 1 - \Lambda_{1,2}$. Panel B reports the seven equations of the VECM (9), each estimated by ordinary least squares with 24 lagged differences, given the benchmark cointegrating matrix. Panel B uses 80,921 observations on 2,526 firms in 118 industries. All regressions absorb firm-by-quarter-of-year and industry-by-quarter fixed effects, and the within R^2 is net of them. Standard errors, in parentheses, are clustered by industry.

6.3 Productivity and prices

Table 5 describes firm-level productivity and price dynamics. Log TFPQ is the production-function residual evaluated at the estimated output elasticities. Log TFPR subtracts the same input index from revenue, $p_{i,t} + y_{i,t}$, as defined in Section 2. The first three columns of Table 5 describe

Table 5: Productivity and prices

	Standard deviation	Correlation with TFPR	Correlation with price	AR(1), quarterly	AR(1), annual
TFPQ, $a_{i,t}$	0.31	0.78	−0.46	0.55	0.70
TFPR, $r_{i,t}$	0.28	–	0.16	0.35	0.63
Output price, $p_{i,t}$	0.19	0.16	–	0.83	0.77

Notes. The first three columns refer to the four-quarter changes of TFPQ, TFPR, and the price. Each change is net of its industry-by-quarter mean. The quarterly AR(1) regresses the log level on its value in the previous quarter, with firm and industry-by-quarter fixed effects. The annual AR(1) regresses the calendar-year mean of the log level on its value in the previous year, with firm and industry-by-year fixed effects.

four-quarter changes net of industry-by-quarter fixed effects. The two AR(1) columns use log levels net of firm and industry-by-quarter fixed effects.

Three observations follow. First, four-quarter changes in TFPQ are the most dispersed, followed by TFPR and the price. On the same dataset, Lenzu et al. (2026) construct productivity using the approach of Gandhi et al. (2020) and find the same ranking. Second, TFPQ changes are positively correlated with TFPR changes and negatively correlated with price changes, with correlations of 0.78 and −0.46. For U.S. plants producing homogeneous goods, Haltiwanger et al. (2018) report correlations of 0.66 and −0.59 in levels. Third, productivity is persistent at annual frequency. The AR(1) coefficient of calendar-year means is 0.70 for TFPQ and 0.63 for TFPR; the quarterly coefficients are lower, at 0.55 and 0.35. Lower quarterly persistence indicates a transitory component, which could arise when production or input purchases do not occur every quarter. In Colombian and Chilean manufacturing, Gandhi et al. (2020) report an average current-lagged productivity correlation of 0.64 in levels, including the ex-post output shock. Using the same data, Trunschke and Judd (2024) estimate log-productivity AR(1) coefficients between 0.77 and 0.91 without firm fixed effects. Both studies measure output with deflated revenue, so their productivity measure is revenue-based, like our TFPR. Without firm fixed effects, our annual TFPR AR(1) coefficient is 0.96. With them, our coefficients measure persistence around each firm's mean.

7 The impulse responses to TFP shocks

The aggregate evidence discussed in the introduction leaves open whether productivity shocks drive cyclical fluctuations at the firm level. We address this question using the responses to

productivity and demand shocks implied by the reduced-form VECM estimated in Section 6. We first trace firms' responses to each shock, highlighting the contrast between large movements in production and small but persistent changes in input use. We then show, through the forecast-error variance decomposition, that neither shock accounts for much of firms' input and price variation.

We treat capital as predetermined within the quarter and order it first, followed by TFPQ, as in Proposition 1, and then TFPR. We interpret TFPR innovations orthogonal to TFPQ and capital as demand, or sales, shocks and assess this interpretation in Section 8. This ordering also allows technical productivity shocks (TFPQ) to affect measured TFPR on impact, while demand shocks can affect technical efficiency in subsequent quarters through endogenous adjustments such as learning by doing or investments in innovation that are triggered by the demand impulse.

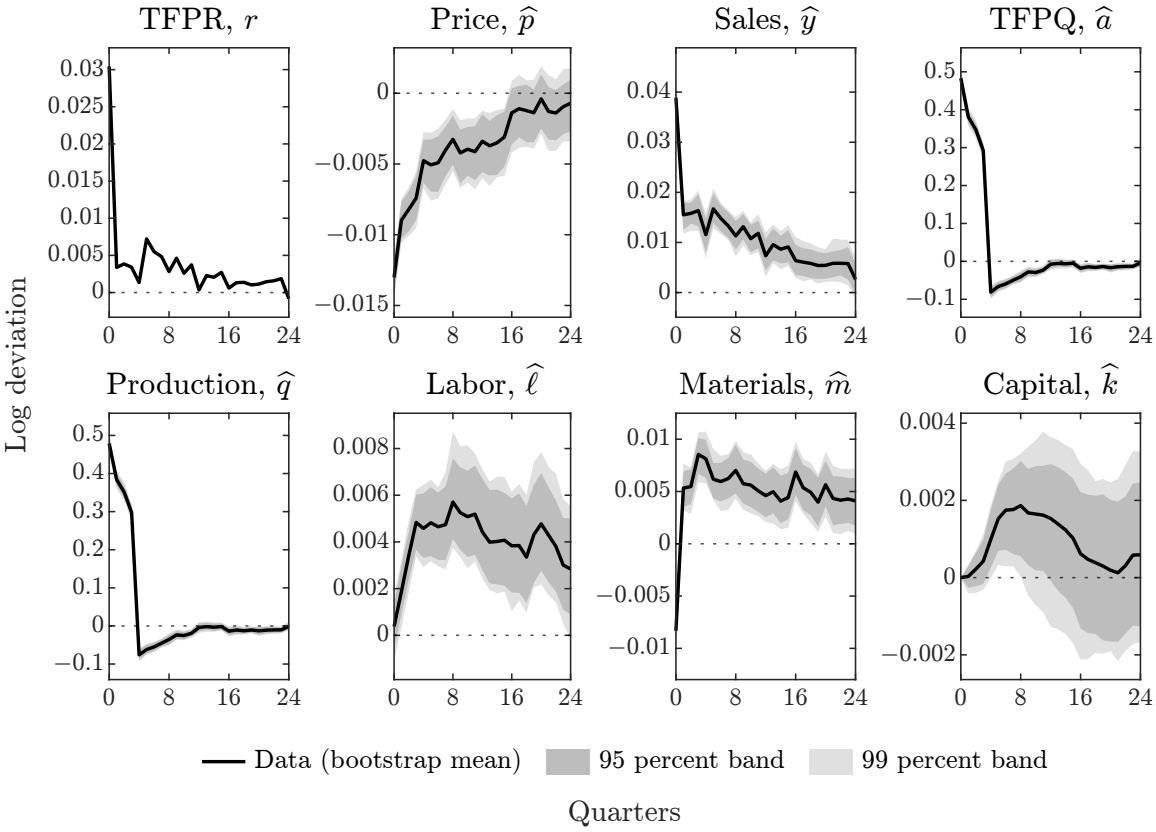
Impulse responses to a TFPQ shock. Figure 3 reports the responses to a one-standard-deviation productivity shock. TFPQ rises by about 50% on impact, then reverts quickly, turning negative after a year. The productivity shock triggers adjustment in pricing, sales, and inputs. Such adjustments are very modest—given the size of the initial productivity impulse—but highly persistent. Prices fall, as expected following a favorable productivity disturbance; however, their response is modest relative to the efficiency gain: the price declines by about 1.3 percent on impact, while sales rise by less than 4 percent. Input adjustment is also limited. Capital is unchanged on impact by the ordering, labor rises by about 4 basis points, and materials fall by about 83 basis points before turning positive after one quarter.

Because inputs move comparatively little, the productivity shock moves production almost one-for-one and follows much the same path over time. The difference between additional production and sales is explained by inventories. The pattern of inventories accords with the cost-optimization role of inventories discussed by, e.g., Ramey and West (1999): firms can concentrate production in periods of temporarily low costs while smoothing sales and avoiding temporary and large swings in output prices.

Also the paths of the variables reveal a marked difference in the dynamic effects of productivity shocks from those predicted by a standard RBC mechanism. Labor and materials remain elevated through the six-year horizon, outlasting the initial productivity gain. Similarly, prices remain significantly below their pre-shock level for about four years, and sales remain above for almost six years. Firms' pricing and sales responses therefore cannot be summarized by contemporaneous movements in technical efficiency.

Taken together, the responses contrast the magnitude of input adjustment with the persistence. Large temporary efficiency gains generate little immediate expansion in factor use, yet the small adjustments that occur persist for years. The modest labor response complements

Figure 3: Impulse responses to the TFPQ shock

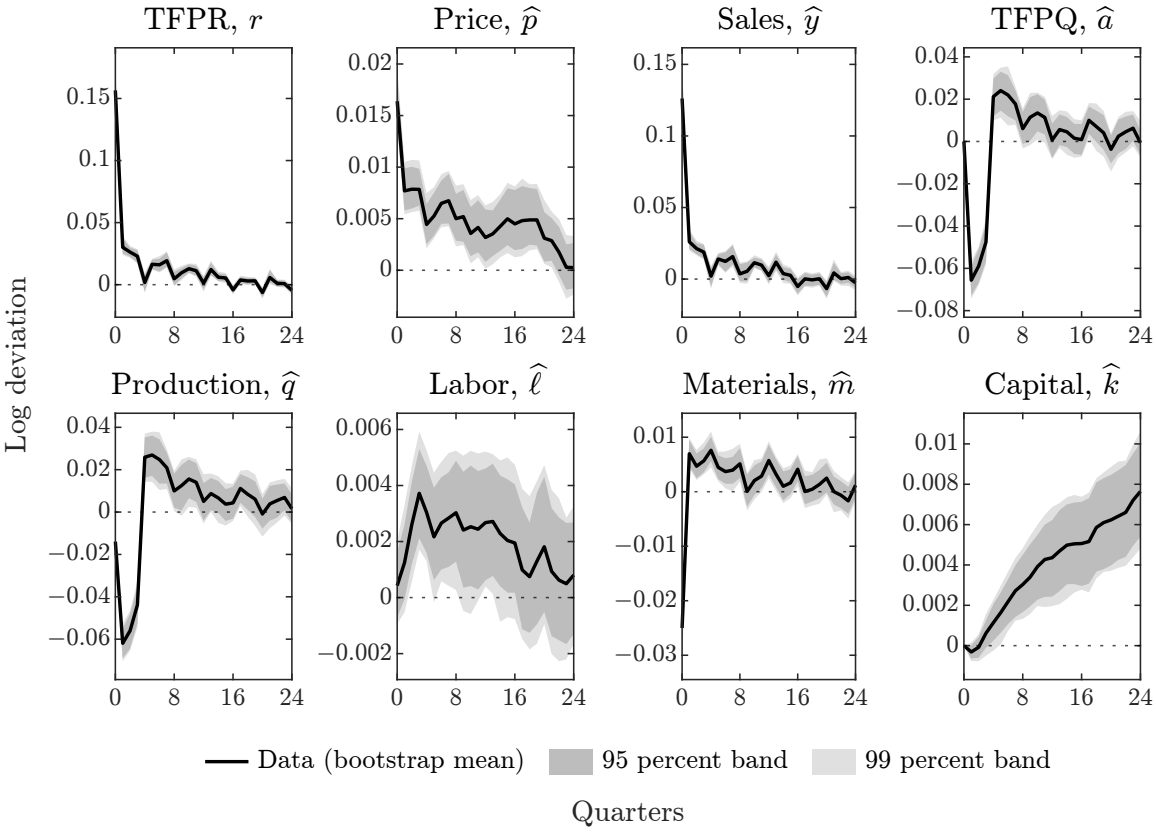


Notes. Responses to a one-standard-deviation TFPQ shock, in log deviations. The black line is the mean of 300 Mammen wild bootstrap replications of the estimated response. The dark and light gray areas are the 95 and 99 percent bootstrap bands; the TFPR panel reports only the bootstrap mean.

Carlsson et al. (2021), who find limited employment effects of permanent technology shocks in Swedish manufacturing. Our quarterly evidence documents limited adjustment across labor, materials, and capital following temporary efficiency gains. These responses are consistent with firms expecting the productivity gains to be temporary, and the persistence of the responses with slow adjustments of labor and materials. A related interpretation is that the short duration of the gain discourages substantial input expansion, while adjustment costs prolong the changes firms do undertake. Section 8 examines whether real and nominal frictions can account for these patterns.

Impulse responses to a TFPR shock. Figure 4 reports the responses to the orthogonal TFPR innovation. A TFPR increase of 15.7% is associated with a price increase of 1.6% and a sales increase of 12.6%. The joint rise in prices and sales is consistent with a demand disturbance. Sales initially rise much more than production, implying an inventory drawdown. TFPR reverts within

Figure 4: Impulse responses to the TFPR shock



Notes. Responses to a one-standard-deviation demand shock, in log deviations. The black line is the mean of 300 Mammen wild bootstrap replications of the estimated response. The dark and light gray areas are the 95 and 99 percent bootstrap bands.

approximately a year, whereas the price response persists for up to six years.

TFPQ is unchanged on impact by the identification assumption, but the demand shock has dynamic effects on TFPQ (and production) through its endogenous propagation. A short-run productivity bust (-6.6% after one quarter) is followed by a temporary productivity boom that peaks at 2.4% after five quarters, with no long-run effects on productivity. In the estimated system, the initial decline and reversal are driven almost entirely by sales (Table A.6 in Appendix F.3).[8] The decline and rebound in measured productivity resemble the productivity J-curve of Brynjolfsson et al. (2021), in which investment in intangibles initially depresses measured productivity before its benefits materialize. Here, the response follows a demand shock, and the productivity gains

[8]Because the estimates control for utilization, the productivity decline cannot be explained by changes in capital-use intensity. Measurement error in utilization and inventories of materials remains a concern, but at the estimated elasticities the productivity slump is large enough to be qualitatively robust.

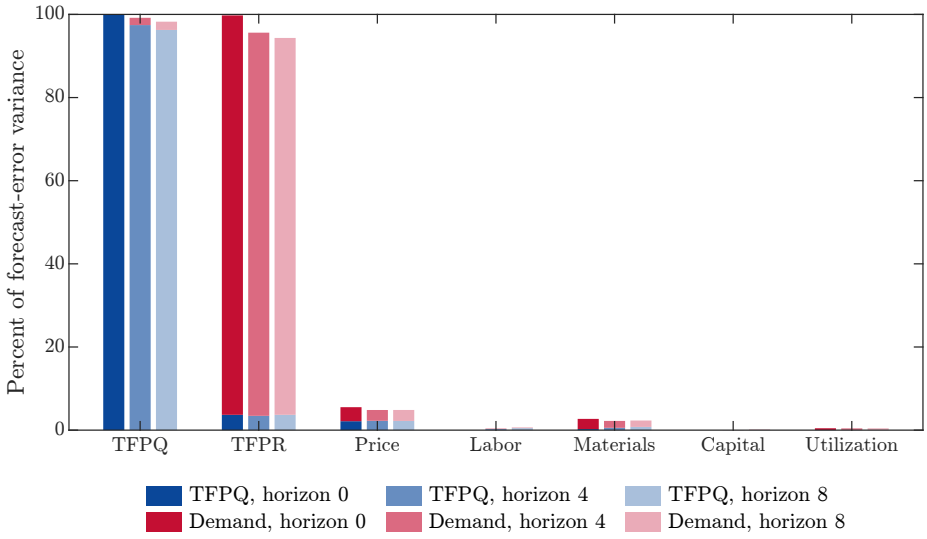
are temporary. One possible explanation is a reallocation of resources from production to sales activities —for example, from the plant to marketing— that raises demand while reducing measured technical efficiency. This interpretation draws on models in which firms devote resources to acquiring customers and expanding demand (as in, e.g., Gourio and Rudanko, 2014). As resources return to production to meet higher demand, the larger scale may generate efficiency gains, consistent with the productivity rebound.

As with the impulse response to the productivity shock, changes in inputs are small and persistent. Labor rises by less than 1% and remains elevated. Materials fall by about 2.5% on impact, then turn positive by less than 1% and remain elevated for much of the six-year horizon. Capital accumulates to about .8% after six years. Thus, on the one hand, both shocks induce large short-run movements in TFP relative to their input responses, while, on the other hand, the changes persist well beyond the initial disturbance.

Forecast-error variance decomposition. Figure 5 assesses the importance of the shocks for firm-level fluctuations through the forecast-error variance decomposition of the empirical VECM, on impact and after four and eight quarters. The productivity shock explains all of TFPQ's impact forecast-error variance, with its share declining only moderately thereafter. In the estimated system, firm productivity is therefore driven mostly by exogenous shocks, despite the endogenous response to demand documented in Figure 4. Demand shocks explain about 96% of TFPR's forecast-error variance on impact and about 91% after eight quarters.

Their contribution to input and price variation is much smaller. Together, TFPQ and demand shocks account for at most 2.7 percent of the forecast-error variance of labor, materials, capital, and utilization, and at most 5.5 percent of the price variance. This accords with the small input and price responses in Figures 3 and 4 and gives the aggregate evidence in Figure 1 a microeconomic counterpart. Productivity gains raise production substantially, but inventories absorb much of the increase while inputs change little. The input responses persist, but the two shocks explain little of firms' cyclical input and price variation. Conditional on industry-by-quarter effects, the evidence thus provides little support for a firm-level RBC mechanism in which temporary idiosyncratic productivity shocks drive cyclical input adjustment.

Figure 5: VECM forecast-error variance decomposition



Notes. Share of the empirical forecast-error variance of each variable explained by the productivity shock and demand shocks, in percent. The share at horizon h is that of the error in forecasting the variable $h + 1$ quarters ahead.

8 The quantitative model

The empirical responses pose a joint challenge: temporary productivity gains raise production substantially but induce little input adjustment, while prices and sales change modestly and inventories absorb much of the additional output. We develop a quantitative model within the framework of Section 2 to ask which frictions can account for this pattern and for the persistence of the resulting adjustments.

We first estimate input adjustment costs and inventory holding costs by matching responses to the identified productivity shock, and examine the model's implications for demand disturbances. We then estimate additional disturbances to firms' optimality conditions to assess the sources of input and price variation. The model is in partial equilibrium, matching the idiosyncratic variation isolated by the VECM's industry-by-quarter fixed effects.

Primitives and market structure. We retain the technology of Section 2: CES production with Hicks-neutral productivity (Equation 1). Kimball preferences give the downward-sloping demand curve in Equation (4). We allow monopolistic competition or oligopoly.

Productivity evolves endogenously as in Equation (2). Under rational expectations, this creates feedback with input and pricing decisions in the presence of adjustment costs. To recover

standard input optimality conditions, we assume that firms do not internalize this feedback and instead forecast productivity with a simple AR(1) model. We solve their decision rules under these perceived AR(1) dynamics, then hold the rules fixed while realized productivity follows the estimated endogenous law in Equation (2).

Adjustment costs. We introduce adjustment costs for investment, materials, and labor, together with a net inventory holding cost. Bars denote variables along the balanced growth path.

The net inventory holding cost is

$$C_S(S_{i,t}) = C_0 + \frac{\psi_S}{2}\frac{\overline{MC}}{\bar{S}}\left(S_{i,t} - \bar{s}\,\overline{Y}\right)^2, \qquad \bar{s} = \frac{\bar{S}}{\overline{Y}}\left[1 + \frac{1-\beta(1-\delta_S)}{\psi_S}\right], \tag{18}$$

where $\overline{MC}$ is real marginal cost, δ_S is inventory depreciation, and $\psi_S > 0$ governs the normalized curvature of the holding cost. The constant C_0 is an additive normalization. The coefficient $\bar{s}$ locates the cost minimum and is determined by the inventory Euler equation.

For $X \in \{I, M, L\}$, define

$$C_X\left(\frac{X_{i,t}}{X_{i,t-1}}\right) = \frac{\psi_X}{2}\left(\frac{X_{i,t}}{X_{i,t-1}} - 1\right)^2, \qquad \psi_X \geq 0. \tag{19}$$

These functions satisfy $C_X(1) = C_X'(1) = 0$ and $C_X''(1) = \psi_X$. Adjustment expenditure is C_X multiplied by current expenditure on the corresponding flow.

We include a variable capital utilization cost as a real resource cost $(P_t^k/P_t)u(U_{i,t})K_{i,t}$, where $u(\overline{U}) = 0$, $u'(\overline{U}) > 0$, $u''(\overline{U}) > 0$, and assume that capital depreciates at rate $0 < \delta_K < 1$.

Firm problem. Real dividends are

$$\begin{aligned} D_{i,t} = {} & \frac{P_{i,t}}{P_t}Y_{i,t} - \frac{W_t}{P_t}\left[1 + C_L\left(\frac{L_{i,t}}{L_{i,t-1}}\right)\right]L_{i,t} - \frac{P_t^m}{P_t}\left[1 + C_M\left(\frac{M_{i,t}}{M_{i,t-1}}\right)\right]M_{i,t} \\ & - \frac{P_t^k}{P_t}\left\{\left[1 + C_I\left(\frac{I_{i,t}}{I_{i,t-1}}\right)\right]I_{i,t} + u(U_{i,t})K_{i,t}\right\} - C_S(S_{i,t}). \end{aligned} \tag{20}$$

Firms face Calvo nominal rigidity. Let $\mathbb{I}_{i,t}$ indicate a reset opportunity, and let $\theta \equiv \Pr\{1 - \mathbb{I}_{i,t}\} \in [0, 1)$ be the probability that a firm cannot adjust its price in a period, independently across periods and of the shocks. The firm observes the opportunity before choosing its date-t allocation.

Denote the reset price by $P^o_{i,t}$. The firm solves

$$\max_{\{Q_{i,t},P_{i,t},P^o_{i,t},U_{i,t},L_{i,t},M_{i,t},I_{i,t},S_{i,t},Y_{i,t},K_{i,t+1}\}_{t\geq 0}} \mathbb{E}_0 \sum_{t=0}^{\infty} \Lambda_{0,t} D_{i,t}, \tag{21}$$

$$\text{subject to} \quad Q_{i,t} = A_{i,t}\left[\alpha_K (U_{i,t}K_{i,t})^\eta + \alpha_L L_{i,t}^\eta + \alpha_M M_{i,t}^\eta\right]^{\nu/\eta}, \qquad (MC_{i,t})$$

$$Y_{i,t} = \mathcal{D}(P_{i,t}, P_t, \varphi_{i,t}) Y_t, \qquad (\lambda^Y_{i,t})$$

$$S_{i,t} = (1-\delta_S) S_{i,t-1} + Q_{i,t} - Y_{i,t}, \qquad (\lambda^S_{i,t})$$

$$K_{i,t+1} = (1-\delta_K) K_{i,t} + I_{i,t}, \qquad (\lambda^K_{i,t})$$

$$P_{i,t} = \mathbb{I}_{i,t} P^o_{i,t} + (1-\mathbb{I}_{i,t}) P_{i,t-1}. \qquad (\lambda^P_{i,t})$$

Here $\Lambda_{0,t} = \beta^t$ discounts real dividends for the risk-neutral firm. The reset-price condition also weights by the probability of no adjustment, through $(\beta\theta)^h$ at horizon h. Current capital $K_{i,t}$ is predetermined, and the expectation $\mathbb{E}_t$ uses the firm's perceived AR(1) productivity process. Lagrange multipliers appear beside the constraints. Appendix B.1 provides the first-order conditions. Appendix B.2 discusses the log-linearized model and shows the derivations.

8.1 Estimation by IRFs matching

We discipline the log-linearized model using the estimated technology parameters and firms' responses to the identified productivity shock. The input and price responses discipline the costs of adjusting labor, materials, and investment and the curvature of inventory holding costs. We then examine the remaining responses to productivity and demand disturbances, which provide additional evidence on the model's mechanisms and limitations.

Calibration. The model is at quarterly frequency. We calibrate most parameters using the estimates in Section 6.1 and values from the literature. Five remaining parameters are estimated: adjustment costs for labor, materials, and investment, the inventory holding cost, and utilization-cost curvature. We fit the first four to the impulse responses to an identified productivity shock (Section 7); utilization curvature is estimated separately from the model, as described below. Table 6 reports the calibration.

First, we use the benchmark technology estimates from Section 6.1: output elasticities θ^L, θ^M, and θ^K, returns to scale ν, the CES parameter η, and strategic complementarities Ω.

Second, the quarterly capital depreciation rate is $\delta_K = .025$, as in Christiano et al. (2005) and Smets and Wouters (2007). We set demand elasticity to $\bar{\epsilon} = 4$, implying a gross markup of $\bar{\epsilon}/(\bar{\epsilon}-1) = 1.33$. The probability of no price reset is $\theta = .70$, close to the estimate of .711 in Gagliardone et al. (2025). The discount factor takes the standard value $\beta = .99$. Inventory

Table 6: Parameters of the quantitative model

Parameter	Description	Value	Source
	Technology		
θ^L	Output elasticity of labor	.195	Table 2
θ^M	Output elasticity of materials	.524	Table 2
θ^K	Output elasticity of capital services	.182	Table 2
η	CES parameter	.438	Table 2
Ω	Strategic complementarity	.512	Table 2
	Capital and utilization		
ξ_u	Curvature of the utilization cost	9.26	Estimated, see text
δ_K	Capital depreciation	.025	Christiano et al. 2005, Smets and Wouters 2007
	Demand and pricing		
$\bar{\epsilon}$	Demand elasticity	4.00	Standard
θ	Probability that the price is not reset	.700	Gagliardone et al. 2025
β	Discount factor	.990	Standard
	Inventories		
$\bar{S}/\overline{Y}$	Inventory-to-sales ratio	.520	Lenzu et al. 2026
δ_S	Inventory depreciation	.026	Lenzu et al. 2026
	Shocks		
ρ_a	Perceived persistence of TFPQ	.699	Estimated, AR(1) of log TFPQ
σ_a	Standard deviation of ε^a	.483	Impact response of TFPQ
ρ_φ	Persistence of the demand shifter	.481	Responses of sales and price to TFPR
σ_φ	Standard deviation of ε^φ	.064	Impact responses of sales and price
	Adjustment costs, estimated by matching IRFs		
$\psi_S/C_S'(\bar{S})$	Inventory holding cost	.008	Elasticity of the cost of inventories
$C_L(1.01)$	Labor adjustment cost	2.94	Cost of a 1% change in % of wage bill
$C_M(1.01)$	Materials adjustment cost	2.14	Cost of a 1% change in % of materials bill
$C_I(1.01)$	Investment adjustment cost	.034	Cost of a 1% change in % of investment expenditure

Notes. The technology parameters are the benchmark estimates in column (A) of Table 2. The curvature ξ_u is estimated from the first-order conditions for utilization and labor, as described in the text. The inventory parameters convert the annual values of Lenzu et al. (2026) to a quarterly frequency. The parameters σ_a, ρ_φ, and σ_φ are set from the estimated impulse responses, as described in the text. The adjustment costs are reported on the scales defined in the text. For $X \in \{L, M, I\}$, the value is the cost of raising X by 1 percent in one quarter, in percent of that quarter's expenditure on X. For inventories, the value is the elasticity of the marginal benefit of holding inventories with respect to the stock.

parameters come from the annual model of Lenzu et al. (2026). Their inventory-to-sales ratio of .13 becomes $\overline{S}/\overline{Y} = 4 \times .13 = .52$ at quarterly frequency, and their annual inventory depreciation rate of .10 becomes $\delta_S = 1 - (1 - .10)^{1/4} = .026$ per quarter.

Third, we estimate utilization-cost curvature by combining the long-run first-order conditions for utilization and labor:

$$\widehat{u}_{i,t} = \frac{\eta - 1}{1 + \xi_u - \eta} (\widehat{k}_{i,t} - \widehat{\ell}_{i,t})$$

We regress $u_{i,t}$ on $k_{i,t} - \ell_{i,t}$ with firm and industry-by-quarter fixed effects. At $\eta = .438$, the estimate implies $\xi_u = 9.3$ (standard error 2.5). Estimated U.S. business-cycle models yield comparable values: Primiceri et al. (2006) estimates $\xi_u = 7.28$, and Justiniano et al. (2010) obtains $\xi_u = 5.30$.

Fourth, we estimate the shock processes from the data. Perceived TFPQ persistence is $\rho_a = .699$, the estimated quarterly AR(1) coefficient of log TFPQ with firm and industry-time fixed effects. We normalize the innovation standard deviation to $\sigma_a = .483$ to match TFPQ's impact response to its own shock. The realized productivity process is calibrated at the estimates of the VECM in Table A.6. Although TFPR identifies the demand shock, its standard deviation and persistence are set from empirical sales and price responses. Let $\widehat{y}_h$ and $\widehat{p}_h$ denote the log-sales and log-price responses at horizon h, in quarters after the shock. Given the calibrated demand elasticity $\bar{\epsilon}$, Equation (A.27) implies the response $\widehat{\varphi}_h$ of the demand shifter φ at horizon h:

$$\widehat{\varphi}_h = \frac{\widehat{y}_h + \bar{\epsilon}\,\widehat{p}_h}{\bar{\epsilon} - 1}.$$

The impact value $\widehat{\varphi}_0$ gives the demand-shock standard deviation, $\sigma_\varphi = .064$. Regressing $\widehat{\varphi}_h$ on its lag by OLS through the origin over quarters 0 to 16 gives persistence $\rho_\varphi = .481$. Apart from these two parameters, the responses are untargeted.

IRF matching and parameter estimates. We fit the four adjustment-cost parameters, $\psi \equiv (\psi_S, \psi_L, \psi_M, \psi_I)^\top$, to the labor, materials, capital, and price responses over quarters 0-16 in Figure 6. The procedure follows Gagliardone and Gertler (2026). The other responses to the TFPQ shock are untargeted. The responses to the TFPR shock are also untargeted apart from the calibration of its standard deviation and persistence described above.

The estimates are $\psi_S = 2.79 \times 10^{-4}$, $\psi_L = 588$, $\psi_M = 427$, and $\psi_I = 6.76$. The last panel of Table 6 translates them into economically interpretable scales. For labor, materials, and investment, the table reports the adjustment cost of changing the flow by Δ, given by $C_X(1+\Delta)$. A 1 percent increase entails an adjustment cost equal to 3 percent of the wage bill, 2.1 percent of the intermediate bill, and .03 percent of investment expenditure, respectively. The substantial costs of changing labor and materials argue against treating them as flexible inputs (Levinsohn and Petrin,

2003). The estimate $\psi_I = 6.76$ is comparable in scale to the literature.[9] An equivalent benchmark with a constant probability of contract adjustment each quarter implies half-lives of 6.6 years for labor contracts and 5.5 years for materials contracts.[10] These long adjustment horizons rationalize why inputs respond little to temporary productivity gains.

For inventories, the table reports the rate at which the marginal cost of holding inventories falls as the stock rises:

$$\left.\frac{d \ln \left[C_S'(S)\right]}{d \ln S}\right|_{S=\bar{S}} = \frac{\psi_S}{C_S'(\bar{S})} = \frac{\psi_S}{1-\beta(1-\delta_S)},$$

The marginal benefit of holding inventories changes little with the stock: a 1 percent stock increase lowers it by .008 percent.

The estimates combine substantial costs of adjusting inputs with a small inventory carrying cost. Costly adjustment limits the response of labor, materials, and investment to temporary productivity gains, so firms initially produce more with nearly the same input bundle. A high inventory carrying cost would encourage them to sell the additional production by cutting prices. Yet the empirical price response is small, even allowing for nominal rigidity and the estimated strategic complementarities. The IRFs matching therefore selects a low inventory carrying cost, allowing firms to absorb much of the additional production in inventories while sales rise only modestly. Together, these frictions separate large fluctuations in technical efficiency and production from the cyclical input adjustment central to a firm-level RBC mechanism.

Fit to the TFPQ responses. Figure 6 compares the model with the four targeted empirical responses. The model reproduces the signs and magnitudes of the responses correctly. Given the narrow confidence bands and small effects, the model misses at some horizons. Nevertheless, price, capital, and materials responses lie within the 95 percent bands for most targeted quarters (13, 16, and 14 of the 17 quarters, respectively).

Figure 7 reports the remaining, untargeted responses to the TFPQ shock. Two features of the fit stand out. First, the model reproduces the reversal of TFPQ and production after one year,

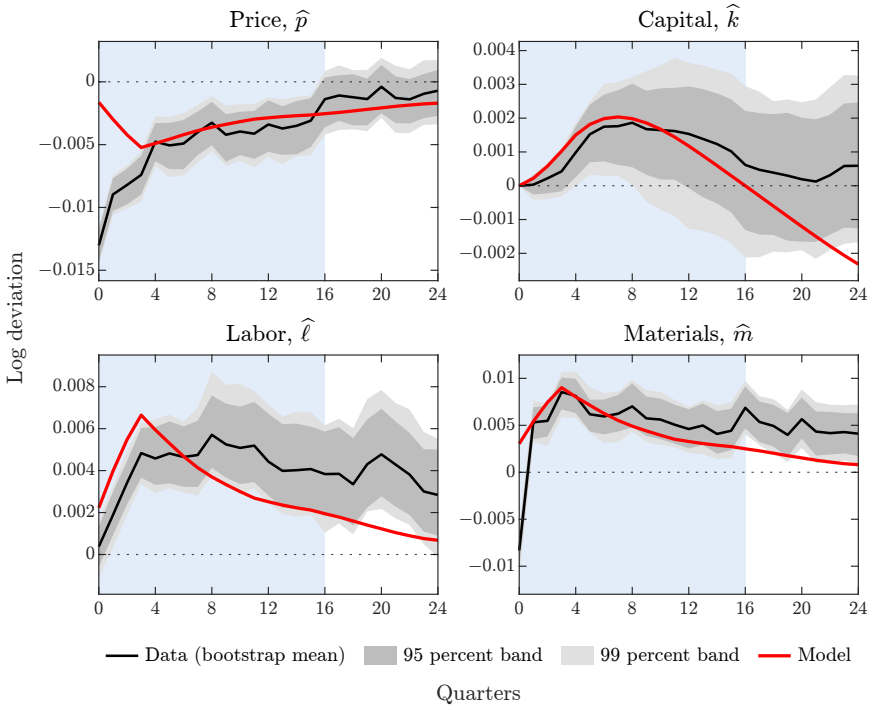
[9] Christiano et al. (2005) estimate 2.48; the posterior median is 2.85 in Justiniano et al. (2010), and the posterior mean is 5.74 in Smets and Wouters (2007). These studies estimate aggregate U.S. models, whereas ψ_I here is estimated from firm-level responses.

[10] For $X \in \{L, M\}$, let $\rho_X \in (0,1)$ denote the quarterly probability that a contract remains unchanged. The linearized input-adjustment condition gives

$$\frac{(1-\rho_X)(1-\beta\rho_X)}{\rho_X} = \frac{1-\eta}{\psi_X}.$$

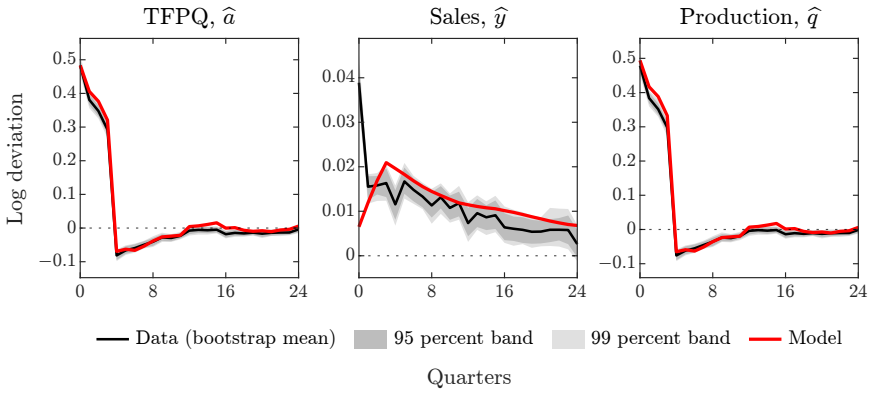
Using the calibration gives $\rho_L = .97$ and $\rho_M = .968$. The half-life, measured in years, is the time by which half of the initial contracts have adjusted, which gives $\log(1/2)/(4\log\rho_X) = 6.6$ and 5.5 years, respectively.

Figure 6: Targeted impulse responses to the TFPQ shock



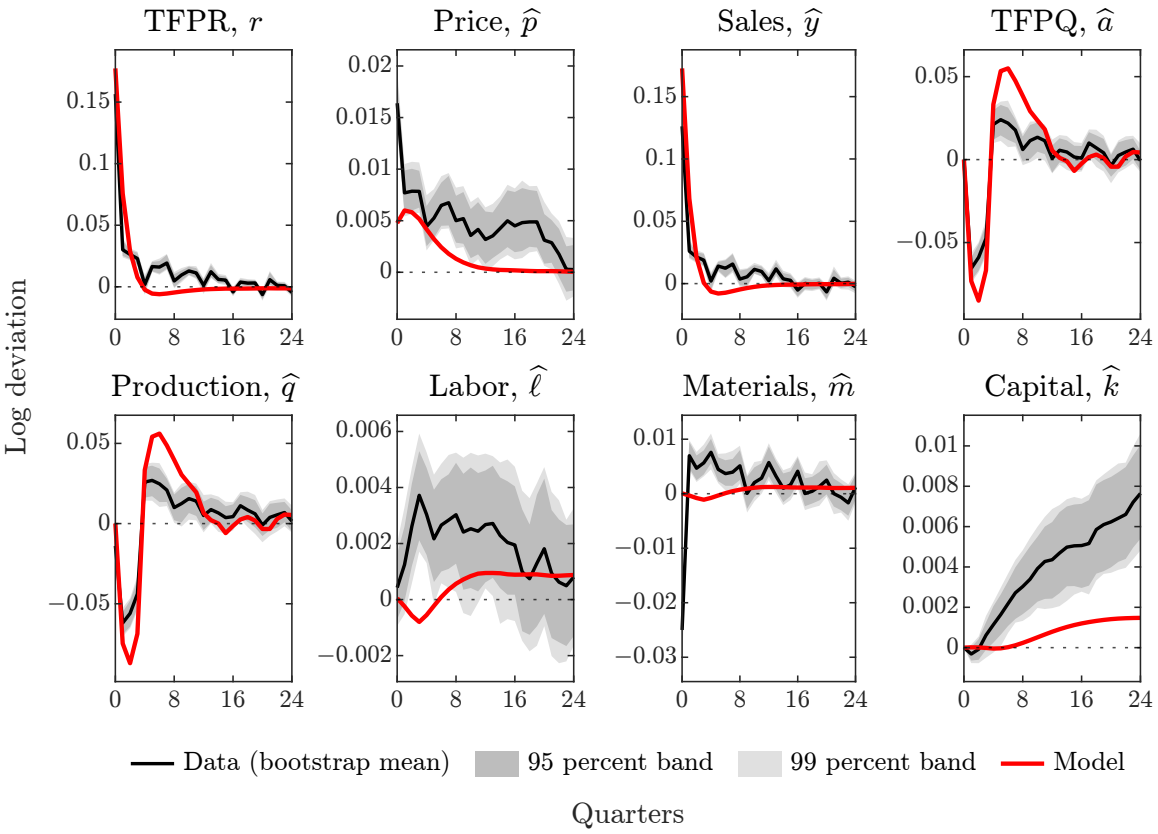
Notes. Responses to a one-standard-deviation TFPQ shock, in log deviations. The black line is the mean of 300 Mammen wild bootstrap replications of the estimated response. The dark and light gray areas are the 95 and 99 percent bootstrap bands. The red line is the response of the model. The shaded area marks quarters 0 to 16, which enter the criterion.

Figure 7: Untargeted impulse responses to the TFPQ shock



Notes. Responses to a one-standard-deviation TFPQ shock, in log deviations. The black line is the mean of 300 Mammen wild bootstrap replications of the estimated response. The dark and light gray areas are the 95 and 99 percent bootstrap bands. The red line is the response of the model.

Figure 8: Untargeted impulse responses to the demand shock



Notes. Responses to a one-standard-deviation demand shock, in log deviations. The black line is the mean of 300 Mammen wild bootstrap replications of the estimated response. The dark and light gray areas are the 95 and 99 percent bootstrap bands. The red line is the response of the model.

with both responses inside the bands from quarter 4 to quarter 11. Both turn positive again in the model at quarters 12 to 15. Second, the model reproduces the sign and magnitude of the response of sales. The model thus links limited input adjustment to adjustment costs, and the small price and sales responses to nominal rigidities, strategic complementarities, and low inventory costs.

Fit to the TFPR responses. The model matches the response to the TFPR shock via the demand disturbance. As the IRFs are untargeted, the fit provides a direct validation of the model mechanism. Figure 8 compares the empirical responses with the model evaluated at the same adjustment-cost estimates and the calibrated demand-shock process.

Three features characterize the fit. First, the model reproduces the joint rise in price and sales on impact. Its price response is less than one-third of the empirical response (0.48 against 1.6 percent); sales rise by 17 percent in the model and 13 percent in the data. Second, the

model reproduces the productivity and production bust and rebound through its calibration to the estimated endogenous productivity dynamics. Third, the responses of inputs are consistent with the empirical responses, as they are largely muted. The order of magnitude is consistent with the data, though the model lies outside of the confidence bands for most quarters for capital (0.15 percent in the model against 0.76 percent in the data after six years). The model also implies a temporary decline in labor and materials due to the overstated decline in productivity. Overall, given that the IRFs are not targeted, the model reproduces remarkably well the effects of demand.

8.2 The role of wedges to the first-order conditions

We established, in the previous section, that the model correctly reproduces the impulse responses to TFP shocks, as driven by supply and demand disturbances. Building from the classical exercises in Chari et al. (2007) and Restuccia and Rogerson (2013), we now show the importance of wedges to the first-order conditions at the firm level for explaining business-cycle fluctuations.

We estimate the processes for the wedges with Bayesian methods applied to the Kalman smoother on the historical panel of firms. As in the model, we demean the data with respect to firm and industry-time fixed effects; the details of the implementation are discussed in Appendix G. Each wedge's prior median is chosen so that the wedge alone reproduces the observed variance of the corresponding growth rate, with persistence parameters at their prior medians. Table 7 reports priors and posterior estimates for the wedge processes; given the amount of firm-level data, priors hardly play any role in the estimation. Two findings stand out. First, labor and materials wedges are close to serially uncorrelated: the posterior modes of ρ_ℓ and ρ_m are .024 and .014, with 90 percent posterior intervals of $[.009, .047]$ and $[.005, .034]$. Second, the investment wedge is negatively auto-correlated, with a posterior mode of ρ_ι of $-.211$ and a 90 percent posterior interval of $[-.222, -.200]$. Overall, the estimated persistence of the shocks is small, pointing to highly transitory disturbances and persistent internal propagation via the high adjustment costs. Table A.7 in Appendix G.3 reports the full set of estimates, including measurement error and correlations.

Using the estimated processes at the posterior mode, we construct impulse responses to the wedges. Figure 9 reports model responses to one-standard-deviation labor, materials, and investment wedge shocks. The figure shows point responses over 24 quarters. Labor and materials respond most strongly to their own wedges (8.2 and 10.3 percent on impact, respectively), with small responses of the other input. Both shocks raise production on impact. The effects on the highly transitory shocks to the inputs is highly persistent due to the estimated adjustment costs; nevertheless, the effects on prices and sales are an order of magnitude smaller than the effects on production. The investment wedge raises capital gradually. None of the shocks moves TFPQ on

Table 7: Priors and posterior estimates of the wedge processes

	Prior			Posterior			
Parameter	Distribution	Mean	SD	Mode	Mean	5%	95%
ρ_ℓ	Beta	.500	.200	.024	.026	.009	.047
σ_ℓ	Inverse gamma	.066	-	.0822	.0819	.0786	.0848
ρ_m	Beta	.500	.200	.0136	.0178	.0050	.0345
σ_m	Inverse gamma	.148	-	.1046	.1037	.1002	.1065
ρ_ι	Beta	.500	.300	$-.211$	$-.211$	$-.222$	$-.200$
σ_ι	Inverse gamma	.418	-	1.830	1.830	1.796	1.864
σ_p	Inverse gamma	.159	-	.0420	.0420	.0417	.0423

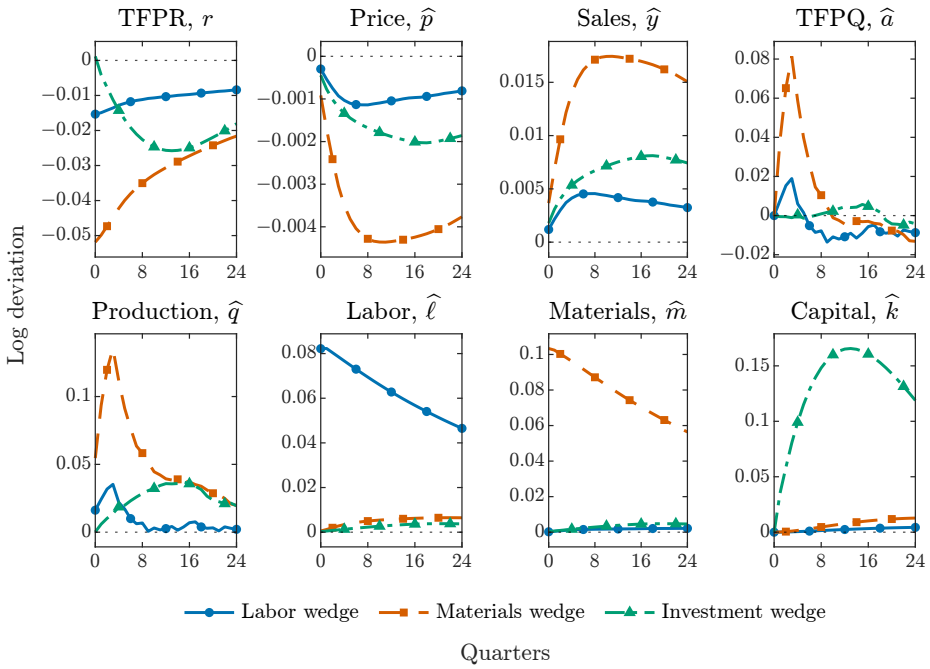
Notes. Estimates of parameters of the wedge processes. The full estimates are in Table A.7 of Appendix G.3.

impact; TFPQ responses arise through endogenous productivity dynamics. TFPR declines because sales adjust more slowly than production. The responses of TFPQ and TFPR indicate a substantial difference in the two measures, reflected in their empirical correlation with inputs and production.

Variance decomposition of the quantitative model. We assess the contributions of the shocks to cyclical fluctuations by performing the same exercise that motivated the firm-level exploration from aggregate data in Figure 1: a forecast-error variance decomposition of the contributions of the estimated idiosyncratic disturbances. Figure 10 illustrates the results at the posterior mode among TFPQ, demand, and the four wedges, on impact and after 4, 8, and 24 quarters. Appendix G.4 describes the computation. Three findings stand out. First, each input is driven by its own wedge. Labor's wedge explains more than 98 percent of its forecast-error variance at every horizon; materials' wedge explains more than 99 percent of its variance. After impact, the investment wedge explains more than 99 percent of capital's forecast-error variance. TFPQ and demand together explain at most .6 percent of labor and materials variance. Second, TFPQ drives production and utilization, explaining 98.9 and 99.8 percent on impact and 81 and 78 percent after 24 quarters, respectively. Third, the price wedge drives prices and, together with demand, sales. It explains 98.6 percent of price forecast-error variance on impact and 77.6 percent after 24 quarters, when TFPQ and demand together explain 10.7 percent. Demand and the price wedge each account for about half of sales forecast-error variance on impact; after 24 quarters, their shares are 34.4 and 52.9 percent.

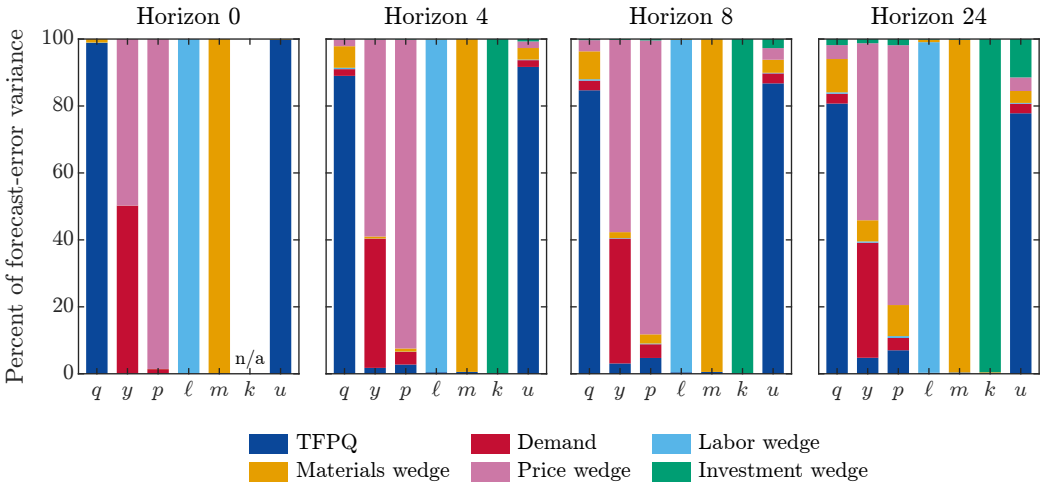
Overall, the variance decomposition provides a clear picture. There is no single shock that explains the bulk of cyclical fluctuations within the firm at business-cycle frequency. The wedges to the first-order conditions explain most of the production and pricing decisions at business-cycle

Figure 9: Impulse responses to the wedges



Notes. Responses to one-standard-deviation innovations that ease the labor, materials, and investment wedges, in log deviations. The blue solid line with circles, the vermillion dashed line with squares, and the green dash-dotted line with triangles are the responses to the labor, materials, and investment wedge shocks, respectively. The shock processes are evaluated at the baseline posterior mode.

Figure 10: Forecast-error variance decomposition of the estimated model



Notes. Share of the forecast-error variance of each log level explained by each shock, in percent, at the posterior mode of Table A.7. The share at horizon h is that of the error in forecasting the variable $h + 1$ quarters ahead. Capital is predetermined and has no forecast error on impact (n/a).

frequency. These firm-specific wedges reflect heterogenous access to input markets in the form of distorted and shadow prices. The presence of such wedges kills the propagation of productivity shocks to firms' policies —the RBC mechanism— and instead appear to be the main candidate for explaining production decisions and cycles within the firm.

9 Conclusions

We use firm-level panel data to identify the dynamic effects of idiosyncratic shocks to firm pricing and production decisions. The impulse responses to these shocks are identified using a new methodology for estimating production functions via firm-level panel vector autoregression models that permits short-run adjustment frictions, endogenous productivity dynamics, and variable markups. Our findings reveal that firms respond little to transitory productivity and demand disturbances, but the effects persist well beyond the shocks. This evidence does not support a propagation mechanism within the firm driven by total factor productivity, the prediction of a firm real business cycle model.

The ability of transitory, idiosyncratic disturbances to generate fluctuations at the firm level should not be of narrow interest. Our findings speak to the nature of the primitive shocks underlying modern theories that aim at micro-founding aggregate fluctuations from firm-level shocks (Gabaix, 2011; Acemoglu et al., 2012; Carvalho and Grassi, 2019; Baqaee and Farhi, 2019, 2020). Our analysis suggests that such aggregation and amplification mechanisms are more likely to originate from disturbances other than productivity. In particular, using our quantitative framework, we highlight the importance of firm-specific, time-varying wedges to the first-order conditions, which explain the lion's share of fluctuations of inputs and prices.

As these wedges may capture, among other disturbances, firm-specific input prices, heterogenous access to input markets and shadow prices (due to, e.g., financial frictions) appear to be likely candidates for firm-level fluctuations. Combining this insight with the theoretical literature on propagation of micro-disturbances, our evidence is compatible with changes over time in the (mis)allocation of resources (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2013) playing an important role for aggregate fluctuations. Therefore, our findings point to heterogeneity in production networks and supply-chain access (Bernard et al., 2019), bank-specific relations (Chodorow-Reich, 2014), and labor-market access (Davis et al., 2013) as interesting venues for research towards understanding business-cycle fluctuations at both the micro and macro levels.

References

Daron Acemoglu, Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016, 2012.

Daniel A Ackerberg, Kevin Caves, and Garth Frazer. Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451, 2015.

Mary Amiti, Oleg Itskhoki, and Jozef Konings. International shocks, variable markups, and domestic prices. *The Review of economic studies*, 86(6):2356–2402, 2019.

Philippe Andrade, Alexander M Dietrich, John Leer, Raphael S Schoenle, Jenny Tang, and Egon Zakrajšek. The future in today's prices: Evidence from a survey of us firms. *Journal of Monetary Economics*, page 103975, 2026.

George-Marios Angeletos, Fabrice Collard, and Harris Dellas. Business-cycle anatomy. *American Economic Review*, 110(10):3030–3070, 2020. doi: 10.1257/aer.20181174.

Manuel Arellano and Stephen Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2): 277–297, 1991.

John Asker, Allan Collard-Wexler, and Jan De Loecker. Dynamic inputs and resource (mis) allocation. *Journal of Political Economy*, 122(5):1013–1063, 2014.

Andrew Atkeson and Ariel Burstein. Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, 98(5):1998–2031, 2008.

David Rezza Baqaee and Emmanuel Farhi. The macroeconomic impact of microeconomic shocks: Beyond hulten's theorem. *Econometrica*, 87(4):1155–1203, 2019.

David Rezza Baqaee and Emmanuel Farhi. Productivity and misallocation in general equilibrium. *The quarterly journal of economics*, 135(1):105–163, 2020.

Susanto Basu and John G. Fernald. Aggregate productivity and aggregate technology. *European Economic Review*, 46(6):963–991, 2002.

Susanto Basu, John G. Fernald, and Miles S. Kimball. Are technology improvements contractionary? *American Economic Review*, 96(5):1418–1448, 2006.

Andrew B Bernard, Andreas Moxnes, and Yukiko U Saito. Production networks, geography, and firm performance. *Journal of political economy*, 127(2):639–688, 2019.

Olivier Jean Blanchard and Danny Quah. The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, 79(4):655–673, 1989.

Richard Blundell and Stephen Bond. Gmm estimation with persistent panel data: an application to production functions. *Econometric reviews*, 19(3):321–340, 2000.

Christoph E Boehm and Nitya Pandalai-Nayar. Convex supply curves. *American Economic Review*, 112(12):3941–3969, 2022.

Scott A. Brave, R. Andrew Butters, and David Kelley. A new "big data" index of U.S. economic activity. *Federal Reserve Bank of Chicago Economic Perspectives*, 43(1):1–30, 2019. doi: 10.21033/ep-2019-1.

Erik Brynjolfsson, Daniel Rock, and Chad Syverson. The productivity j-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, 13(1): 333–372, 2021.

Ricardo J Caballero and Eduardo MRA Engel. Price stickiness in ss models: New interpretations of old results. *Journal of monetary economics*, 54:100–121, 2007.

Guillermo A Calvo. Staggered prices in a utility-maximizing framework. *Journal of monetary Economics*, 12(3):383–398, 1983.

Mikael Carlsson, Julián Messina, and Oskar Nordström Skans. Firm-level shocks and labour flows. *The Economic Journal*, 131(634):598–623, 2021.

Vasco M Carvalho and Basile Grassi. Large firm dynamics and the business cycle. *American Economic Review*, 109(4):1375–1425, 2019.

Varadarajan V Chari, Patrick J Kehoe, and Ellen R McGrattan. Business cycle accounting. *Econometrica*, 75(3):781–836, 2007.

Gabriel Chodorow-Reich. The employment effects of credit market disruptions: Firm-level evidence from the 2008–9 financial crisis. *The Quarterly Journal of Economics*, 129(1):1–59, 2014.

Jason Choi, Taeyoung Doh, Andrew Foerster, and Zinnia Martinez. Monetary policy stance is tighter than federal funds rate. Economic Letter 2022-30, Federal Reserve Bank of San Francisco, 2022.

Lawrence Christiano, Martin S Eichenbaum, and Robert J Vigfusson. What happens after a technology shock?, 2003.

Lawrence J. Christiano, Martin Eichenbaum, and Charles L. Evans. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45, 2005. doi: 10.1086/426038.

Joel M David and Venky Venkateswaran. The sources of capital misallocation. *American Economic Review*, 109(7):2531–2567, 2019.

Steven J Davis, R Jason Faberman, and John C Haltiwanger. The establishment-level behavior of vacancies and hiring. *The Quarterly Journal of Economics*, 128(2):581–622, 2013.

Jan De Loecker, Catherine Fuss, and Johannes Van Biesebroeck. International competition and firm performance: Evidence from belgium. Technical report, NBB Working Paper, 2014.

Martin Eichenbaum. Some empirical evidence on the production level and production cost smoothing models of inventory investment. *American Economic Review*, 79(4):853–864, 1989.

Robert F Engle and Clive WJ Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276, 1987.

John Fernald, Amit Gandhi, Dimitrije Ruzic, and James Traina. Micro and macro perspectives on production-based markups. *Journal of Economic Literature*, 2026. Forthcoming.

John G. Fernald. A quarterly, utilization-adjusted series on total factor productivity. Working Paper 2012-19, Federal Reserve Bank of San Francisco, 2014.

Lucia Foster, John Haltiwanger, and Chad Syverson. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425, 2008. doi: 10.1257/aer.98.1.394.

Neville Francis and Valerie A Ramey. Is the technology-driven real business cycle hypothesis dead? shocks and aggregate fluctuations revisited. *Journal of Monetary Economics*, 52(8):1379–1399, 2005.

Francesco Franco and Thomas Philippon. Firms and aggregate dynamics. *The Review of Economics and Statistics*, 89(4):587–600, 2007.

Xavier Gabaix. The granular origins of aggregate fluctuations. *Econometrica*, 79(3):733–772, 2011.

Luca Gagliardone and Mark Gertler. Oil prices, monetary policy, and inflation surges. *American Economic Journal: Macroeconomics*, 18(4):468–501, 2026.

Luca Gagliardone and Joris Tielens. Dynamic pricing under information frictions: Evidence from firm-level subjective expectations. 2026.

Luca Gagliardone, Mark Gertler, Simone Lenzu, and Joris Tielens. Anatomy of the phillips curve: micro evidence and macro implications. *American Economic Review*, 115(11):3941–3974, 2025.

Luca Gagliardone, Mark Gertler, Simone Lenzu, and Joris Tielens. Micro and macro cost-price dynamics in normal times and during inflation surges. Technical report, Centre for Economic Policy Research, 2026.

Jordi Galí. Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? *American Economic Review*, 89(1):249–271, 1999. doi: 10.1257/aer.89.1.249.

Amit Gandhi, Salvador Navarro, and David A Rivers. On the identification of gross output production functions. *Journal of Political Economy*, 128(8):2973–3016, 2020.

Mark Gertler and Simon Gilchrist. Monetary policy, business cycles, and the behavior of small manufacturing firms. *The Quarterly Journal of Economics*, 109(2):309–340, 1994. doi: 10.2307/2118465.

Mishel Ghassibe, Boromeus Wanengkirtyo, and Ivan Yotzov. Intertemporal pass-through. 2025.

Simon Gilchrist and Egon Zakrajšek. Credit spreads and business cycle fluctuations. *American Economic Review*, 102(4):1692–1720, 2012. doi: 10.1257/aer.102.4.1692.

Mikhail Golosov and Robert E Lucas Jr. Menu costs and phillips curves. *Journal of Political Economy*, 115(2):171–199, 2007.

François Gourio and Leena Rudanko. Customer capital. *The Review of Economic Studies*, 81(3): 1102–1136, 2014. doi: 10.1093/restud/rdu007.

Clive W. J. Granger and Tae-Hwy Lee. Investigation of production, sales and inventory relationships using multicointegration and non-symmetric error correction models. *Journal of Applied Econometrics*, 4(S1):S145–S159, 1989. doi: 10.1002/jae.3950040508.

John Haltiwanger, Robert Kulick, and Chad Syverson. Misallocation measures: The distortion that ate the residual. Technical report, National Bureau of Economic Research, 2018.

Hugo A. Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60(5): 1127–1150, 1992.

Chang-Tai Hsieh and Peter J. Klenow. Misallocation and manufacturing TFP in china and india. *The Quarterly Journal of Economics*, 124(4):1403–1448, 2009.

Søren Johansen. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica: journal of the Econometric Society*, pages 1551–1580, 1991.

Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti. Investment shocks and business cycles. *Journal of Monetary Economics*, 57(2):132–145, 2010. doi: 10.1016/j.jmoneco.2009.12.008.

Anil K. Kashyap, Owen A. Lamont, and Jeremy C. Stein. Credit conditions and the cyclical behavior of inventories. *The Quarterly Journal of Economics*, 109(3):565–592, 1994. doi: 10.2307/2118414.

Agnes Norris Keiller, Áureo de Paula, and John Van Reenen. Production function estimation using subjective expectations data. Technical report, National Bureau of Economic Research, 2024.

Aubhik Khan and Julia K. Thomas. Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica*, 76(2):395–436, 2008.

Miles S Kimball. The quantitative analytics of the basic neomonetarist model, 1995.

Robert G. King, Charles I. Plosser, James H. Stock, and Mark W. Watson. Stochastic trends and economic fluctuations. *American Economic Review*, 81(4):819–840, 1991.

Finn E. Kydland and Edward C. Prescott. Time to build and aggregate fluctuations. *Econometrica*, 50(6):1345–1370, 1982. doi: 10.2307/1913386.

Finn E. Kydland and Edward C. Prescott. The econometrics of the general equilibrium approach to business cycles. *Scandinavian Journal of Economics*, 93(2):161–178, 1991. doi: 10.2307/3440324.

Rolf Larsson, Johan Lyhagen, and Mickael Löthgren. Likelihood-based cointegration tests in heterogeneous panels. *The Econometrics Journal*, 4(1):109–142, 2001.

Simone Lenzu, David A. Rivers, Joris Tielens, and Shi Hu. Financial shocks, productivity, and prices. *Review of Economic Studies*, 2026. Forthcoming.

James Levinsohn and Amil Petrin. Estimating production functions using inputs to control for unobservables. *The review of economic studies*, 70(2):317–341, 2003.

Jacob Marschak and William H Andrews Jr. Random simultaneous equations and the theory of production. *Econometrica, Journal of the Econometric Society*, pages 143–205, 1944.

José Luis Montiel Olea, Mikkel Plagborg-Møller, Eric Qian, and Christian K Wolf. Double robustness of local projections and some unpleasant varithmetic. *Econometrica*, 94(4):1313–1343, 2026.

Emi Nakamura and Dawit Zerom. Accounting for incomplete pass-through. *The review of economic studies*, 77(3):1192–1230, 2010.

Steven Olley and Ariel Pakes. The dynamics of productivity in the telecommunications equipment industry, 1992.

Peter Pedroni. Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, 61(S1):653–670, 1999.

Mikkel Plagborg-Møller and Christian K Wolf. Local projections and vars estimate the same impulse responses. *Econometrica*, 89(2):955–980, 2021.

Giorgio E. Primiceri, Ernst Schaumburg, and Andrea Tambalotti. Intertemporal disturbances. NBER Working Paper 12243, National Bureau of Economic Research, 2006.

Valerie A. Ramey and Kenneth D. West. Inventories. In John B. Taylor and Michael Woodford, editors, *Handbook of Macroeconomics*, volume 1B, chapter 13, pages 863–923. Elsevier, 1999. doi: 10.1016/S1574-0048(99)10021-1.

Diego Restuccia and Richard Rogerson. Misallocation and productivity, 2013.

Julio J Rotemberg. Sticky prices in the united states. *Journal of political economy*, 90(6):1187–1211, 1982.

Frank Smets and Rafael Wouters. Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, 97(3):586–606, 2007. doi: 10.1257/aer.97.3.586.

John B Taylor. Staggered wage setting in a macro model. *The American Economic Review*, 69(2): 108–113, 1979.

Markus Trunschke and Kenneth L Judd. Estimating gross output production functions. *NBER Working Paper*, (w33205), 2024.

Olivier Wang and Iván Werning. Dynamic oligopoly and price stickiness. *American Economic Review*, 112(8):2815–2849, 2022.

Mark W Watson. Vector autoregressions and cointegration. *Handbook of econometrics*, 4: 2843–2915, 1994.

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

Productivity Shocks and Firm Real Business Cycles

L. Gagliardone S. Lenzu J. Tielens

Appendix

A Proofs for Section 2

A.1 Proof of Theorem 1

Proof. We follow the theorem discussion, retaining aggregate prices until the final fixed-effects step. Before time fixed effects are included, the observable vector is

$$[q_{i,t}, y_{i,t}, p_{i,t}, \ell_{i,t}, m_{i,t}, k_{i,t}, u_{i,t}, w_t + \ell_{i,t}, p_t^m + m_{i,t}]^\top.$$

We thus retain both real inputs and nominal expenditures. The pricing restriction also conditions on the competitors' price index $p_{-i,t}$. We derive five independent economic equilibrium restrictions on these observables. Bars denote the balanced growth path and hats log deviations from it: $\widehat{\ell}_{i,t} = \ell_{i,t} - \bar{\ell}_{i,t}$, and similarly for prices and other quantities. For the log markup, $\widehat{\mu}_{i,t} = \mu_{i,t} - \bar{\mu}$. All expansions are first order.

Production and productivity. The expansion of Equation (1) is

$$\widehat{q}_{i,t} = \widehat{a}_{i,t} + \theta^K(\widehat{u}_{i,t} + \widehat{k}_{i,t}) + \theta^L\widehat{\ell}_{i,t} + \theta^M\widehat{m}_{i,t}, \qquad \nu = \theta^K + \theta^L + \theta^M. \tag{A.1}$$

The elasticities are evaluated along the path and are constant by Assumption 2. Substituting this identity into Equation (2) and subtracting the equation evaluated along the reference path gives the structural output equation

$$\begin{aligned}\Delta\widehat{q}_{i,t} = (1-\rho)\left[\theta^K(\widehat{u}_{i,t-1} + \widehat{k}_{i,t-1}) + \theta^L\widehat{\ell}_{i,t-1} + \theta^M\widehat{m}_{i,t-1} - \widehat{q}_{i,t-1}\right] \\ + \theta^K(\Delta\widehat{u}_{i,t} + \Delta\widehat{k}_{i,t}) + \theta^L\Delta\widehat{\ell}_{i,t} + \theta^M\Delta\widehat{m}_{i,t} + \sum_{x\in\boldsymbol{x}}\sum_{\tau=1}^{p}\psi_{\tau,x}\Delta\widehat{x}_{i,t-\tau} + \varepsilon^a_{i,t}.\end{aligned} \tag{A.2}$$

Marginal cost and relative inputs. Taking logs of Equation (7) gives

$$mc^\star_{i,t} = \ln(W_tL_{i,t} + P_t^mM_{i,t}) - \ln(\theta^L_{i,t} + \theta^M_{i,t}) - q_{i,t}.$$

The input first-order conditions imply, at the reference allocation,

$$\Theta = \frac{\overline{W}\,\overline{L}}{\overline{W}\,\overline{L} + \overline{P}^m\overline{M}} = \frac{\theta^L}{\theta^L + \theta^M}.$$

Consequently, the first cost term has the expansion

$$\ln(W_t L_{i,t} + P_t^m M_{i,t}) = \ln(\overline{W}\,\overline{L} + \overline{P}^m \overline{M}) + \Theta(\widehat{w}_t + \widehat{\ell}_{i,t}) + (1-\Theta)(\widehat{p}_t^m + \widehat{m}_{i,t}).$$

For the second term, the CES elasticities imply the exact identity

$$\ln(\theta_{i,t}^L + \theta_{i,t}^M) = \ln \nu + \ln(\alpha_L L_{i,t}^\eta + \alpha_M M_{i,t}^\eta) - \ln\left[\alpha_K (U_{i,t}K_{i,t})^\eta + \alpha_L L_{i,t}^\eta + \alpha_M M_{i,t}^\eta\right].$$

The last two terms expand, respectively, as

$$\ln(\alpha_L \overline{L}^\eta + \alpha_M \overline{M}^\eta) + \eta\Theta\widehat{\ell}_{i,t} + \eta(1-\Theta)\widehat{m}_{i,t},$$
$$\ln\left[\alpha_K(\overline{U}\,\overline{K})^\eta + \alpha_L \overline{L}^\eta + \alpha_M \overline{M}^\eta\right] + \frac{\eta}{\nu}\left[\theta^K(\widehat{u}_{i,t} + \widehat{k}_{i,t}) + \theta^L \widehat{\ell}_{i,t} + \theta^M \widehat{m}_{i,t}\right].$$

Using $\theta^L/\nu = \Theta(1-\theta^K/\nu)$ and $\theta^M/\nu = (1-\Theta)(1-\theta^K/\nu)$ therefore gives

$$\ln(\theta_{i,t}^L + \theta_{i,t}^M) = \ln(\theta^L + \theta^M) + \frac{\eta\theta^K}{\nu}\left[\Theta\widehat{\ell}_{i,t} + (1-\Theta)\widehat{m}_{i,t} - \widehat{u}_{i,t} - \widehat{k}_{i,t}\right].$$

Combining these expressions yields

$$\begin{aligned} mc_{i,t}^\star = \ln\left(\frac{\overline{W}\,\overline{L} + \overline{P}^m \overline{M}}{(\theta^L + \theta^M)\overline{Q}}\right) + \Theta(\widehat{w}_t + \widehat{\ell}_{i,t}) + (1-\Theta)(\widehat{p}_t^m + \widehat{m}_{i,t}) \\ + \frac{\eta\theta^K}{\nu}\left[\widehat{u}_{i,t} + \widehat{k}_{i,t} - \Theta\widehat{\ell}_{i,t} - (1-\Theta)\widehat{m}_{i,t}\right] - \widehat{q}_{i,t}. \end{aligned} \tag{A.3}$$

Dividing the two input first-order conditions also gives

$$\frac{W_t}{P_t^m} = \frac{\alpha_L}{\alpha_M}\left(\frac{L_{i,t}}{M_{i,t}}\right)^{\eta-1}, \qquad (1-\eta)(m_{i,t} - \ell_{i,t}) = w_t - p_t^m - \ln(\alpha_L/\alpha_M).$$

Hence the relative-input equilibrium error is, up to its stationary wedge,

$$\widehat{m}_{i,t} - \widehat{\ell}_{i,t} - \frac{\widehat{w}_t - \widehat{p}_t^m}{1-\eta} = \omega_{i,t}^m \sim I(0). \tag{A.4}$$

Markups under Kimball demand and oligopoly. We evaluate the markup expansion in Gagliardone et al. (2025) along the balanced growth path with equal sales shares. Since Equation (5) defines the log gross markup, its first-order expansion is

$$\widehat{\mu}_{i,t} = -\frac{\widehat{\epsilon}_{i,t}}{\bar{\epsilon} - 1}, \qquad \widehat{\epsilon}_{i,t} = \ln \epsilon_{i,t} - \ln \bar{\epsilon}. \tag{A.5}$$

Under Kimball demand, Equation (4) implies

$$\Upsilon'\left(\frac{e^{\varphi_{i,t}} Y_{i,t}}{Y_t}\right) = e^{p_{i,t} - p_t - \varphi_{i,t}}.$$

Only the relative price enters this identity. At the symmetric reference allocation,

$$\bar{\epsilon} = -\frac{\Upsilon'(1)}{\Upsilon''(1)}, \qquad \widehat{y}_{i,t} - \widehat{y}_t + \varphi_{i,t} = -\bar{\epsilon}(\widehat{p}_{i,t} - \widehat{p}_t - \varphi_{i,t}).$$

The first-order change in log elasticity equals the change in the log demand argument times the log

derivative of $-\Upsilon'(x)/(x\Upsilon''(x))$. Substituting into Equation (A.5) and using $p_{-i,t} = p_t$ for atomistic firms gives

$$\begin{aligned} \widehat{\mu}_{i,t} &= -\Gamma(\widehat{p}_{i,t} - \widehat{p}_{-i,t} - \varphi_{i,t}), \\ \Gamma &= -\frac{\bar{\epsilon}}{\bar{\epsilon}-1} \left.\frac{\partial \ln[-\Upsilon'(x)/(x\Upsilon''(x))]}{\partial \ln x}\right|_{x=1}. \end{aligned} \tag{A.6}$$

For nested-CES oligopoly, let $\gamma > \sigma > 1$ denote the within-industry and across-industry substitution elasticities, and let $\bar{s}$ denote the expenditure share under symmetry. To use the competitors' price convention of Equation (14), first note that

$$\widehat{p}_t = \bar{s}(\widehat{p}_{i,t} - \varphi_{i,t}) + (1-\bar{s})\widehat{p}_{-i,t}.$$

With the demand shifter expressed in price units, the share expansion is therefore

$$\widehat{s}_{i,t} = -(\gamma-1)(1-\bar{s})(\widehat{p}_{i,t} - \widehat{p}_{-i,t} - \varphi_{i,t}), \qquad \widehat{s}_{i,t} = \ln s_{i,t} - \ln \bar{s}.$$

For Bertrand competition, $\epsilon_{i,t} = \gamma - (\gamma-\sigma)s_{i,t}$ and $\bar{\epsilon} = \gamma - (\gamma-\sigma)\bar{s}$, so

$$\widehat{\epsilon}_{i,t} = -\frac{(\gamma-\sigma)\bar{s}}{\bar{\epsilon}}\widehat{s}_{i,t}, \qquad \Gamma = \frac{(\gamma-\sigma)(\gamma-1)\bar{s}(1-\bar{s})}{\bar{\epsilon}(\bar{\epsilon}-1)}.$$

For Cournot competition, $1/\epsilon_{i,t} = (1-s_{i,t})/\gamma + s_{i,t}/\sigma$, with $1/\bar{\epsilon} = (1-\bar{s})/\gamma + \bar{s}/\sigma$. Thus

$$\widehat{\epsilon}_{i,t} = -\bar{\epsilon}\bar{s}(1/\sigma - 1/\gamma)\widehat{s}_{i,t}, \qquad \Gamma = \frac{(\gamma-1)\bar{s}(1-\bar{s})(1/\sigma - 1/\gamma)}{1-1/\bar{\epsilon}}.$$

Both expressions give Equation (A.6), with Γ defined throughout relative to competitors' prices.

For dynamic oligopoly, let $\chi_{i,t}$ be the response of competitors' prices to the firm's reset price. Around the Nash reference $\bar{\chi} = 0$, the residual elasticity

$$\epsilon_{i,t} = \gamma - (\gamma-\sigma)[s_{i,t} + (1-s_{i,t})\chi_{i,t}]$$

implies

$$\widehat{\mu}_{i,t} = -\Gamma(\widehat{p}_{i,t} - \widehat{p}_{-i,t} - \varphi_{i,t}) + \frac{(\gamma-\sigma)(1-\bar{s})}{\bar{\epsilon}(\bar{\epsilon}-1)}\widehat{\chi}_{i,t},$$

where Γ is the Bertrand expression above and $\widehat{\chi}_{i,t} = \chi_{i,t} - \bar{\chi}$ is a level deviation. A stationary reaction-slope term is included in the pricing wedge.

Pricing and the equilibrium errors. With the log-markup convention of Equation (5), the frictionless pricing condition is $p^\star_{i,t} = \mu_{i,t} + mc^\star_{i,t}$, or $P^\star_{i,t} = e^{\mu_{i,t}} MC^\star_{i,t}$ in levels. Using the competitors' price convention of Equation (A.6), evaluating the markup at the frictionless price, and setting $\Omega = \Gamma/(1+\Gamma)$ gives

$$\widehat{p}^\star_{i,t} = (1-\Omega)\widehat{mc}^\star_{i,t} + \Omega\widehat{p}_{-i,t} + \Omega\varphi_{i,t}.$$

The reference price is $\bar{p} = \bar{\mu} + \overline{mc}^\star$, with reference marginal cost given by the logarithm in Equation (A.3). Substituting that equation, using stationarity of $q_{i,t} - y_{i,t}$, and including stationary pricing

and input wedges gives

$$\begin{aligned}\widehat{p}_{i,t} = (1-\Omega)\big[\Theta(\widehat{w}_t + \widehat{\ell}_{i,t}) + (1-\Theta)(\widehat{p}^{\,m}_t + \widehat{m}_{i,t}) - \widehat{y}_{i,t}\big] \\ + (1-\Omega)\frac{\eta\theta^K}{\nu}\big[\widehat{u}_{i,t} + \widehat{k}_{i,t} - \Theta\widehat{\ell}_{i,t} - (1-\Theta)\widehat{m}_{i,t}\big] + \Omega\widehat{p}_{-i,t} + \omega^p_{i,t}.\end{aligned} \tag{A.7}$$

Here $\omega^p_{i,t}$ is the stationary pricing residual of Equation (10), including the demand disturbance and, where relevant, the stationary reaction-slope term.

Equations (A.7), (A.1), and (A.4) give the pricing residual $\omega^p_{i,t}$, productivity $\widehat{a}_{i,t}$, and the relative-input residual $\omega^m_{i,t}$. Assumption 4 gives the remaining stationary errors $\widehat{q}_{i,t} - \widehat{y}_{i,t}$ and $\widehat{u}_{i,t}$.

Reduced form and fixed effects. Substitute $\widehat{x} = x - \bar{x}$. Collect reference constants into firm effects and common reference paths, aggregate prices, and their lags and changes into industry-by-time effects. The common price component of Equation (A.7) is

$$(1-\Omega)[\Theta\widehat{w}_t + (1-\Theta)\widehat{p}^{\,m}_t] + \Omega\widehat{p}_{-i,t},$$

where the last term is absorbed insofar as the competitors' index is common within the industry. The nine observables reduce to the seven distinct variables $\boldsymbol{x}_{i,t}$ after removing time fixed effects.

The five equilibrium errors, with the signs used in the theorem, are

$$\boldsymbol{\omega}_{i,t} = [-\omega^p_{i,t}, -\widehat{a}_{i,t}, \omega^m_{i,t}, \widehat{q}_{i,t} - \widehat{y}_{i,t}, \widehat{u}_{i,t}]^\top.$$

After collecting the common components, they equal $\delta^\top \boldsymbol{x}_{i,t}$ up to the fixed effects. The five rows of $\delta^\top$ are independent: the minor formed by columns (p, q, m, y, u) has determinant -1. Thus the economic restriction matrix has rank five. The structural representation is Equation (17). For nonsingular A_0, premultiplication by A_0^{-1} gives Equation (9), with

$$\Lambda = A_0^{-1}\widetilde{\Lambda}, \qquad B(\mathsf{L})^\top = A_0^{-1}\widetilde{B(\mathsf{L})}^\top, \qquad \boldsymbol{\epsilon}_{i,t} = A_0^{-1}\boldsymbol{\varepsilon}_{i,t}.$$

The structural right-hand side contains seven equation disturbances, possibly generated by fewer independent shocks. Invertibility makes the reduced-form innovations orthogonal to the observed past. Inversion changes adjustment loadings but preserves the equilibrium-error coefficient vectors. Under the usual finite-moment and regressor-rank conditions, with the time dimension increasing so that fixed-effects bias vanishes, this orthogonality establishes consistency of reduced-form least squares. Recovering individual technology parameters additionally requires the identification and estimation conditions in the proof of Corollary 1. □

A.2 Proof of Corollary 1

Proof. We first distinguish estimation of the long-run pricing relation from estimation of short-run output dynamics, then show how their coefficients jointly recover the technology parameters.

Long-run relation and super-consistency. Equation (A.7), expressed in levels and retaining the competitors' price, gives Equation (10). Its residual $\omega^p_{i,t}$ is stationary. The first-stage pricing regression retains nominal expenditures and competitors' prices explicitly and includes firm fixed effects, without industry-by-time fixed effects. To describe its estimation, let $\boldsymbol{z}_{i,t}$ collect a nonredundant basis of its regressors after removing firm fixed effects, and let $\boldsymbol{b}^p$ collect their coefficients. The least-squares normal equations give

$$\widehat{\boldsymbol{b}}^p - \boldsymbol{b}^p = \left(\sum_{i,t} \boldsymbol{z}_{i,t} \boldsymbol{z}_{i,t}^\top \right)^{-1} \sum_{i,t} \boldsymbol{z}_{i,t} \omega^p_{i,t}.$$

In an identified nonstationary direction, the time-series orders of the denominator and numerator are T^2 and T, respectively, under the usual functional limit and nonsingularity conditions. The corresponding coefficient is therefore super-consistent, with error $O_p(T^{-1})$ for fixed N, even when the regressor innovations and the stationary residual are correlated.

Short-run dynamics and consistency. The first row of Equation (9), using the sign convention $\delta_1^\top \boldsymbol{x}_{i,t-1} = -\omega^p_{i,t-1}$ up to fixed effects, gives Equation (11). All its regressors are dated $t-1$ or earlier. Invertibility and the innovation property imply that $\epsilon^q_{i,t}$ is orthogonal to these regressors. Consequently, with a correctly specified lag structure, finite moments, sufficient regressor variation, and $T \to \infty$ so that the dynamic fixed-effects bias vanishes, least squares consistently estimates the reduced-form output coefficients. The coefficients on stationary equilibrium errors and lagged differences have the ordinary time-series rate, rather than the super-consistency rate of an identified cointegrating coefficient.

If the pricing relation is estimated consistently under the conditions just described, its fitted residual can replace the true pricing residual. In particular, super-consistent estimation in nonstationary directions and consistent estimation in stationary directions imply, under the corresponding moment bounds,

$$\frac{1}{NT} \sum_{i,t} (\widehat{\omega}^p_{i,t} - \omega^p_{i,t})^2 = o_p(1).$$

This permits consistent estimation of the output equation with the generated residual.

Identification from the two equations. Let b^p_x denote a pricing coefficient and b^q_x the coefficient on the lagged level of x in the output equation. Use the lagged pricing residual in place of lagged price, alongside the six other lagged levels. A separate lagged-price regressor would duplicate information already contained in these regressors.

Before removing common input prices, the nominal-expenditure coefficients satisfy

$$b^p_{w+\ell} = (1-\Omega)\Theta, \qquad b^p_{p^m+m} = (1-\Omega)(1-\Theta), \qquad b^p_y = -(1-\Omega), \qquad b^p_k = (1-\Omega)\frac{\eta\theta^K}{\nu}.$$

Hence

$$\Theta = \frac{b^p_{w+\ell}}{b^p_{w+\ell} + b^p_{p^m+m}}, \qquad \Omega = 1 + b^p_y, \qquad \frac{\eta\theta^K}{\nu} = -\frac{b^p_k}{b^p_y}.$$

With common input prices absorbed by time effects, the labor and materials coefficients instead sum the nominal-expenditure and real-input coefficients:

$$b^p_\ell = (1-\Omega)\left(1 - \frac{\eta\theta^K}{\nu}\right)\Theta, \qquad b^p_m = (1-\Omega)\left(1 - \frac{\eta\theta^K}{\nu}\right)(1-\Theta).$$

Their ratio gives the same split, $\Theta = b^p_\ell/(b^p_\ell + b^p_m)$, provided that denominator is nonzero.

Expanding the other lagged levels in the output equation gives

$$\begin{aligned} b^q_q &= -\Lambda_{1,2} + \Lambda_{1,4}, & b^q_y &= -\Lambda_{1,4}, \\ b^q_\ell &= \Lambda_{1,2}\theta^L - \Lambda_{1,3}, & b^q_m &= \Lambda_{1,2}\theta^M + \Lambda_{1,3}, \\ b^q_k &= \Lambda_{1,2}\theta^K. \end{aligned}$$

Adding the first pair recovers $\Lambda_{1,2} = -(b^q_q + b^q_y)$; adding the second pair recovers $\Lambda_{1,2}(\theta^L + \theta^M)$. Combining these sums with the pricing split $\Theta = \theta^L/(\theta^L + \theta^M)$ gives

$$\begin{aligned} \theta^L &= -\Theta\frac{b^q_\ell + b^q_m}{b^q_q + b^q_y}, & \theta^M &= -(1-\Theta)\frac{b^q_\ell + b^q_m}{b^q_q + b^q_y}, \\ \theta^K &= -\frac{b^q_k}{b^q_q + b^q_y}, & \Lambda_{1,3} &= \Theta b^q_m - (1-\Theta)b^q_\ell. \end{aligned} \tag{A.8}$$

This recovery requires $\Lambda_{1,2} \neq 0$. The remaining technology parameters follow from

$$\nu = \theta^L + \theta^M + \theta^K, \qquad \eta = -\frac{b^p_k}{b^p_y}\frac{\nu}{\theta^K},$$

with $\theta^K > 0$, $\theta^L > 0$, $\theta^M > 0$, and $\Omega < 1$.

Finally, evaluating the CES elasticities at the same reference allocation used for the linearization gives

$$\frac{\alpha_K}{\alpha_L} = \frac{\theta^K}{\theta^L}\left(\frac{\bar{L}}{\overline{UK}}\right)^\eta, \qquad \frac{\alpha_M}{\alpha_L} = \frac{\theta^M}{\theta^L}\left(\frac{\bar{L}}{\bar{M}}\right)^\eta.$$

Bars denote reference input levels. Under the corollary's average-based reference, $\bar{L} = \mathbb{E}(L_{i,t})$, $\bar{M} = \mathbb{E}(M_{i,t})$, and $\overline{UK} = \mathbb{E}(U_{i,t}K_{i,t})$; capital services use utilization in levels. Normalizing the three weights to sum to one uniquely recovers $\alpha_K, \alpha_L, \alpha_M$. The inverse mapping is continuous under the stated nondegeneracy conditions. Consistent estimates of both equations and of the reference input ratios therefore consistently recover the technology parameters by the continuous mapping theorem. □

A.3 Proof of Proposition 1

Proof. Start from the structural representation (17), written as

$$A_0 \Delta \boldsymbol{x}_{i,t} = \boldsymbol{h}_{i,t-1} + \boldsymbol{\varepsilon}_{i,t},$$

where $\boldsymbol{h}_{i,t-1}$ collects the fixed effects and lagged terms, and $\boldsymbol{\varepsilon}_{i,t} \in \mathbb{R}^7$ is the vector of disturbances entering the seven structural equations, possibly driven by fewer than seven independent shocks. By Equation (13), its first component is $\varepsilon^a_{i,t}$.

The first row of A_0 is exactly the productivity selection vector. Thus, letting $\boldsymbol{e}_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^7$ denote the selection vector for the first entry,

$$\boldsymbol{e}_1^\top A_0 = \begin{bmatrix} 1 & 0 & 0 & -\theta^L & -\theta^M & -\theta^K & -\theta^K \end{bmatrix} = \boldsymbol{e}_a^\top.$$

Since A_0 is nonsingular, the reduced-form innovation in Equation (9) is $\boldsymbol{\epsilon}_{i,t} = A_0^{-1}\boldsymbol{\varepsilon}_{i,t}$. Consequently,

$$\begin{aligned} \boldsymbol{e}_a^\top \boldsymbol{\epsilon}_{i,t} &= \boldsymbol{e}_a^\top A_0^{-1} \boldsymbol{\varepsilon}_{i,t} \\ &= \boldsymbol{e}_1^\top A_0 A_0^{-1} \boldsymbol{\varepsilon}_{i,t} \\ &= \boldsymbol{e}_1^\top \boldsymbol{\varepsilon}_{i,t} = \varepsilon^a_{i,t}. \end{aligned} \tag{A.9}$$

Hence the linear combination eliminates every other structural disturbance and recovers the productivity shock exactly under the maintained normalization.

Replacing $q_{i,t}$ with $a_{i,t} = \boldsymbol{e}_a^\top \boldsymbol{x}_{i,t}$ transforms the system invertibly. When $a_{i,t}$ is ordered first, its reduced-form innovation is the structural productivity shock. Under the maintained orthogonality to other structural shocks, the first Cholesky innovation is the standardized productivity shock. Propagating its associated impact vector through the reduced-form dynamics identifies the impulse responses. □

B Derivations for quantitative model (Section 8)

B.1 Optimality conditions

Define the physical marginal products by

$$MPL_{i,t} \equiv \frac{\partial Q_{i,t}}{\partial L_{i,t}}, \qquad MPM_{i,t} \equiv \frac{\partial Q_{i,t}}{\partial M_{i,t}}, \qquad MPK_{i,t} \equiv \frac{\partial Q_{i,t}}{\partial K_{i,t}}.$$

Their log deviations are denoted by $\widehat{mpl}_{i,t}$, $\widehat{mpm}_{i,t}$, and $\widehat{mpk}_{i,t}$. The conditions for production and sales give

$$\lambda^S_{i,t} = MC_{i,t}, \qquad \lambda^Y_{i,t} = \frac{P_{i,t}}{P_t} - MC_{i,t}. \tag{A.10}$$

Production and stored output have the same shadow value; the sales multiplier is the relative price less marginal cost.

Inputs and utilization. The labor and materials conditions are

$$\begin{aligned} MC_{i,t}MPL_{i,t} = \frac{W_t}{P_t}\left[1 + C_L\left(\frac{L_{i,t}}{L_{i,t-1}}\right) + \frac{L_{i,t}}{L_{i,t-1}}C_L'\left(\frac{L_{i,t}}{L_{i,t-1}}\right)\right] \\ - \beta\mathbb{E}_t\left[\frac{W_{t+1}}{P_{t+1}}\left(\frac{L_{i,t+1}}{L_{i,t}}\right)^2 C_L'\left(\frac{L_{i,t+1}}{L_{i,t}}\right)\right], \end{aligned} \tag{A.11}$$

The marginal value of labor equals the wage plus current marginal adjustment expenditure, less the discounted expected reduction in next period's adjustment expenditure.

$$\begin{aligned} MC_{i,t}MPM_{i,t} = \frac{P_t^m}{P_t}\left[1 + C_M\left(\frac{M_{i,t}}{M_{i,t-1}}\right) + \frac{M_{i,t}}{M_{i,t-1}}C_M'\left(\frac{M_{i,t}}{M_{i,t-1}}\right)\right] \\ - \beta\mathbb{E}_t\left[\frac{P_{t+1}^m}{P_{t+1}}\left(\frac{M_{i,t+1}}{M_{i,t}}\right)^2 C_M'\left(\frac{M_{i,t+1}}{M_{i,t}}\right)\right]. \end{aligned} \tag{A.12}$$

The materials condition has the same interpretation. The future growth ratios are squared because adjustment expenditure is scaled by the input used in that future period.

The CES identity $U_{i,t}(\partial Q_{i,t}/\partial U_{i,t}) = K_{i,t}MPK_{i,t}$ implies that the utilization condition is

$$MC_{i,t}MPK_{i,t} = \frac{P_t^k}{P_t}U_{i,t}u'(U_{i,t}). \tag{A.13}$$

The left-hand side values capital's marginal product; the right-hand side is marginal utilization resource cost. Constant depreciation removes any direct utilization effect from the capital constraint.

Investment and capital. Using $\lambda_{i,t}^K = (P_t^k/P_t)Q_{i,t}^K$, the investment condition is

$$\frac{P_t^k}{P_t}Q_{i,t}^K = \frac{P_t^k}{P_t}\left[1 + C_I\left(\frac{I_{i,t}}{I_{i,t-1}}\right) + \frac{I_{i,t}}{I_{i,t-1}}C_I'\left(\frac{I_{i,t}}{I_{i,t-1}}\right)\right] - \beta\mathbb{E}_t\left[\frac{P_{t+1}^k}{P_{t+1}}\left(\frac{I_{i,t+1}}{I_{i,t}}\right)^2 C_I'\left(\frac{I_{i,t+1}}{I_{i,t}}\right)\right]. \tag{A.14}$$

The value of installed capital equals its acquisition price plus net marginal adjustment expenditure. The capital Euler equation is

$$\frac{P_t^k}{P_t}Q_{i,t}^K = \beta\mathbb{E}_t\left[MC_{i,t+1}MPK_{i,t+1} - \frac{P_{t+1}^k}{P_{t+1}}u(U_{i,t+1}) + (1-\delta_K)\frac{P_{t+1}^k}{P_{t+1}}Q_{i,t+1}^K\right]. \tag{A.15}$$

An additional unit of capital delivers its marginal product net of utilization expenditure, together with the value of the undepreciated unit next period.

Inventories. The inventory Euler equation is

$$MC_{i,t} = -\frac{\psi_S \overline{MC}}{\bar{S}}(S_{i,t} - \bar{s}\,\overline{Y}) + \beta(1-\delta_S)\mathbb{E}_t MC_{i,t+1}. \tag{A.16}$$

The current marginal value equals the discounted value of the surviving inventory unit net of its marginal holding cost.

Pricing. Let $t+h \mid t$ denote the path on which the price set at t remains unchanged through $t+h$. The reset-price condition is

$$\mathbb{E}_t \sum_{h=0}^{\infty} (\beta\theta)^h Y_{i,t+h|t}(\epsilon_{i,t+h|t}-1)\left[\frac{P^o_{i,t}}{P_{t+h}} - \frac{\epsilon_{i,t+h|t}}{\epsilon_{i,t+h|t}-1} MC_{i,t+h|t}\right] = 0. \tag{A.17}$$

The reset price balances deviations from the desired markup over marginal cost, weighted by sales, demand elasticity, discounting, and price survival. Dynamic oligopoly additionally requires the effects of current prices on rivals' future actions to be specified.

B.2 Log-linearization of the quantitative model

B.2.1 Log-linearized model

We express the log-linearized model in deviations from industry-time and firm fixed effects. These effects absorb common aggregate prices and the constants from the first-order expansion around the balanced growth path. Hats denote deviations from the fixed effects. Derivations are provided in the next subsection.

Production and inputs. The production function, in deviations from the fixed effects, is

$$\widehat{q}_{i,t} = \widehat{a}_{i,t} + \theta^K(\widehat{u}_{i,t} + \widehat{k}_{i,t}) + \theta^L \widehat{\ell}_{i,t} + \theta^M \widehat{m}_{i,t}. \tag{A.18}$$

Inputs are weighted by their output elasticities, with $\theta^K + \theta^L + \theta^M = \nu$. The log-linearized first-order condition for labor is

$$\widehat{mc}_{i,t} + \widehat{mpl}_{i,t} = \psi_L\left[\Delta\widehat{\ell}_{i,t} - \beta\mathbb{E}_t\Delta\widehat{\ell}_{i,t+1} + w^{\ell}_{i,t}\right]. \tag{A.19}$$

The left-hand side is labor's marginal value, with the common wage absorbed by fixed effects. The right-hand side is the current adjustment cost net of the discounted effect of today's employment on next period's adjustment cost. The firm-specific wedge $w^{\ell}_{i,t}$ is a persistent shock between the marginal product of labor and marginal cost. The corresponding condition for materials is

$$\widehat{mc}_{i,t} + \widehat{mpm}_{i,t} = \psi_M\left[\Delta\widehat{m}_{i,t} - \beta\mathbb{E}_t\Delta\widehat{m}_{i,t+1} + w^{m}_{i,t}\right]. \tag{A.20}$$

The marginal value of materials equals the analogous adjustment-cost wedge, with the common materials price absorbed by fixed effects. The firm-specific wedge $w^{m}_{i,t}$ separates the marginal

product of materials from marginal cost. The marginal products are

$$\widehat{mpl}_{i,t} \equiv (\widehat{q}_{i,t} - \widehat{\ell}_{i,t}) + \frac{\eta}{\nu}(\widehat{a}_{i,t} - \widehat{q}_{i,t} + \nu\widehat{\ell}_{i,t}),$$
$$\widehat{mpm}_{i,t} \equiv (\widehat{q}_{i,t} - \widehat{m}_{i,t}) + \frac{\eta}{\nu}(\widehat{a}_{i,t} - \widehat{q}_{i,t} + \nu\widehat{m}_{i,t}).$$

In each expression, the first term is the log deviation of output per unit of input; the second captures the change in its CES output elasticity. The utilization first-order condition is

$$\widehat{mc}_{i,t} + \widehat{mpk}_{i,t} = (1 + \xi_u)\widehat{u}_{i,t}, \qquad \xi_u \equiv \frac{\overline{U}u''(\overline{U})}{u'(\overline{U})}. \tag{A.21}$$

The marginal value of capital equals marginal utilization expenditure, whose log response to utilization is $1 + \xi_u$. The physical marginal product of capital is

$$\widehat{mpk}_{i,t} \equiv (\widehat{q}_{i,t} - \widehat{k}_{i,t}) + \frac{\eta}{\nu}\Big(\widehat{a}_{i,t} - \widehat{q}_{i,t} + \nu(\widehat{u}_{i,t} + \widehat{k}_{i,t})\Big).$$

The first term is the log deviation of output per unit of installed capital; the second captures how capital services affect the capital output elasticity. Define Tobin's Q as

$$Q^K_{i,t} \equiv \frac{\lambda^K_{i,t}}{P^k_t / P_t}, \qquad \widehat{q}^K_{i,t} \equiv \ln(Q^K_{i,t}/\overline{Q}^K), \qquad \overline{Q}^K = 1.$$

Here $\lambda^K_{i,t}$ is the real shadow value of installed capital, so $Q^K_{i,t}$ expresses that value relative to its acquisition price. The first-order condition for investment is

$$\widehat{q}^K_{i,t} = \psi_I \left[\Delta\widehat{\imath}_{i,t} - \beta\mathbb{E}_t\Delta\widehat{\imath}_{i,t+1} + w^I_{i,t}\right]. \tag{A.22}$$

Tobin's Q deviates from one by the current marginal investment adjustment cost net of its discounted effect on future adjustment costs. Here $\widehat{\imath}_{i,t}$ is log investment in deviations from fixed effects, and $w^I_{i,t}$ is a firm-specific investment shock. The capital Euler equation is

$$\widehat{q}^K_{i,t} = \beta(1-\delta_K)\mathbb{E}_t\widehat{q}^K_{i,t+1} + (1-\beta(1-\delta_K))\mathbb{E}_t\left[\widehat{mc}_{i,t+1} + \widehat{mpk}_{i,t+1} - \widehat{u}_{i,t+1}\right]. \tag{A.23}$$

The first term is the discounted continuation value of capital. The second is its expected marginal product, valued at marginal cost, net of utilization expenditure. Capital evolves according to

$$\widehat{k}_{i,t+1} = (1-\delta_K)\widehat{k}_{i,t} + \delta_K\widehat{\imath}_{i,t}. \tag{A.24}$$

Next period's installed capital combines the surviving capital stock with current investment; utilization does not change the depreciation rate.

Inventories. Inventory accumulation gives

$$\left(1 + \delta_S \cdot \frac{\overline{S}}{\overline{Y}}\right)\widehat{q}_{i,t} - \widehat{y}_{i,t} = \frac{\overline{S}}{\overline{Y}}\left[\widehat{s}_{i,t} - (1-\delta_S)\widehat{s}_{i,t-1}\right]. \tag{A.25}$$

Production in excess of sales adds to inventories, after depreciation. The coefficients convert log

changes using the balanced-growth inventory-to-sales ratio.

The inventory Euler equation is

$$\widehat{mc}_{i,t} = -\psi_S \widehat{s}_{i,t} + \beta(1-\delta_S)\mathbb{E}_t \widehat{mc}_{i,t+1}. \tag{A.26}$$

The current shadow value of output equals the discounted value of a surviving inventory unit, net of its marginal holding cost. The change in that cost is governed by $\psi_S \widehat{s}_{i,t}$.

Pricing. Denote by $\bar{\epsilon} > 1$ the demand elasticity along the balanced growth path. Demand is

$$\widehat{y}_{i,t} = -\bar{\epsilon}\widehat{p}_{i,t} + (\bar{\epsilon}-1)\varphi_{i,t}. \tag{A.27}$$

Sales fall with the firm's price and rise with its demand shifter; the coefficient on $\varphi_{i,t}$ reflects the Kimball demand normalization. The reset price is

$$\widehat{p}^{\,o}_{i,t} = (1-\beta\theta)\Big((1-\Omega)\widehat{mc}_{i,t} + \Omega\varphi_{i,t}\Big) + \beta\theta\mathbb{E}_t\widehat{p}^{\,o}_{i,t+1}. \tag{A.28}$$

The current pricing target combines marginal cost with the demand-induced markup component. The continuation is discounted by β and the probability θ that the price survives. The price law is

$$\widehat{p}_{i,t} = (1-\theta)\widehat{p}^{\,o}_{i,t} + \theta\widehat{p}_{i,t-1} + \varepsilon^p_{i,t}. \tag{A.29}$$

where $\varepsilon^p_{i,t}$ is a mean-zero i.i.d. shock.

Shocks. The perceived shock dynamics are all modeled as AR(1) processes:

$$\begin{aligned}
\widehat{a}_{i,t} &= \rho_a \widehat{a}_{i,t-1} + \varepsilon^a_{i,t}, && |\rho_a| < 1 \\
\varphi_{i,t} &= \rho_\varphi \varphi_{i,t-1} + \varepsilon^\varphi_{i,t}, && |\rho_\varphi| < 1, \\
w^\ell_{i,t} &= \rho_\ell w^\ell_{i,t-1} + \varepsilon^\ell_{i,t}, && |\rho_\ell| < 1, \\
w^m_{i,t} &= \rho_m w^m_{i,t-1} + \varepsilon^m_{i,t}, && |\rho_m| < 1, \\
w^\iota_{i,t} &= \rho_\iota w^\iota_{i,t-1} + \varepsilon^\iota_{i,t}, && |\rho_\iota| < 1,
\end{aligned} \tag{A.30}$$

The persistence parameters govern firms' forecasts, and the ε terms are innovations. Realized productivity follows Equation (2), while decision rules remain those computed under the perceived processes. All innovations are assumed normally distributed.

Observables. The observables are the VECM variables other than utilization: output, sales, labor, materials, capital, and price. All are demeaned with respect to firm-quarter and industry-time fixed effects. We allow five measurement errors, each i.i.d. over time. They enter the log levels of production, labor, materials, and capital separately, and the log levels of price and sales through a shared error. Their contributions to growth rates are moving average processes of order one, MA(1). The shared error reflects the construction of the sales index (Section 4.2).

B.2.2 Derivations

Production and inputs. Differentiating the CES production function gives Equation (A.18). Its physical marginal products satisfy

$$MPL_{i,t} = \nu\alpha_L A_{i,t}^{\eta/\nu} Q_{i,t}^{1-\eta/\nu} L_{i,t}^{\eta-1},$$
$$MPM_{i,t} = \nu\alpha_M A_{i,t}^{\eta/\nu} Q_{i,t}^{1-\eta/\nu} M_{i,t}^{\eta-1},$$
$$MPK_{i,t} = \nu\alpha_K A_{i,t}^{\eta/\nu} Q_{i,t}^{1-\eta/\nu} U_{i,t}^{\eta} K_{i,t}^{\eta-1}.$$

Taking logs and removing fixed effects gives the three marginal-product expressions in the main text. None of these physical marginal products includes $MC_{i,t}$.

For $X \in \{L, M, I\}$, the quadratic cost and its derivative satisfy

$$C_X\left(\frac{X_{i,t}}{X_{i,t-1}}\right) = 0 \quad \text{to first order,} \qquad C_X'\left(\frac{X_{i,t}}{X_{i,t-1}}\right) = \psi_X \Delta\widehat{x}_{i,t}.$$

Here $\widehat{x}$ denotes $\widehat{\ell}$, $\widehat{m}$, or $\widehat{\imath}$. Since the growth-ratio factors multiply a derivative that is zero along the balanced growth path, their deviations contribute only at second order. Thus the current marginal adjustment term is $\psi_X \Delta\widehat{x}_{i,t}$ and the future term is $\psi_X \mathbb{E}_t \Delta\widehat{x}_{i,t+1}$. Substituting these into Equations (A.11) and (A.12) and absorbing common input prices gives Equations (A.19) and (A.20).

Utilization. Taking logs of Equation (A.13) and absorbing the common acquisition price gives

$$\widehat{mc}_{i,t} + \widehat{mpk}_{i,t} = \widehat{u}_{i,t} + \frac{\overline{U}u''(\overline{U})}{u'(\overline{U})}\widehat{u}_{i,t} = (1 + \xi_u)\widehat{u}_{i,t}.$$

Investment and capital. Along the balanced growth path, the investment condition gives $\overline{Q}^K = 1$. Applying the adjustment-cost expansion above to Equation (A.14) and dividing by the acquisition price gives Equation (A.22).

For the capital Euler equation, use $u(\overline{U}) = 0$ and the utilization condition to obtain the balanced-growth identity

$$\frac{\overline{MC}\,\overline{MPK}}{\overline{P}^k/\overline{P}} = \overline{U}u'(\overline{U}) = \frac{1}{\beta} - (1 - \delta_K).$$

Consequently, the first-order change in the utilization cost is

$$u(U_{i,t}) = u'(\overline{U})\overline{U}\widehat{u}_{i,t} = \left[\frac{1}{\beta} - (1 - \delta_K)\right]\widehat{u}_{i,t}.$$

Expanding Equation (A.15), dividing by the balanced-growth acquisition price, and absorbing

common price terms then gives

$$\widehat{q}^K_{i,t} = \beta(1-\delta_K)\mathbb{E}_t\widehat{q}^K_{i,t+1} + \beta\frac{\overline{MC}\,\overline{MPK}}{\overline{P}^k/\overline{P}}\mathbb{E}_t[\widehat{mc}_{i,t+1} + \widehat{mpk}_{i,t+1}] - \beta\overline{U}u'(\overline{U})\mathbb{E}_t\widehat{u}_{i,t+1}.$$

The identity above makes the last two coefficients equal to $1-\beta(1-\delta_K)$, yielding Equation (A.23).

Finally, $\bar{I} = \delta_K\overline{K}$. Expanding capital accumulation gives

$$\overline{K}\widehat{k}_{i,t+1} = (1-\delta_K)\overline{K}\widehat{k}_{i,t} + \bar{I}\widehat{\imath}_{i,t}.$$

Dividing by $\overline{K}$ yields Equation (A.24).

Inventories. Along the balanced growth path, $\overline{Q} = \overline{Y} + \delta_S\bar{S}$. Expanding inventory accumulation gives

$$\overline{Q}\widehat{q}_{i,t} - \overline{Y}\widehat{y}_{i,t} = \bar{S}[\widehat{s}_{i,t} - (1-\delta_S)\widehat{s}_{i,t-1}].$$

Dividing by $\overline{Y}$ gives Equation (A.25). For the inventory Euler equation, the holding-cost curvature is $C_S''(\bar{S}) = \psi_S\overline{MC}/\bar{S}$. Subtracting the balanced-growth condition gives

$$\overline{MC}\,\widehat{mc}_{i,t} = -\psi_S\overline{MC}\,\widehat{s}_{i,t} + \beta(1-\delta_S)\overline{MC}\,\mathbb{E}_t\widehat{mc}_{i,t+1}.$$

Dividing by $\overline{MC}$ yields Equation (A.26).

Pricing. Taking logs of inverse Kimball demand and removing fixed effects gives

$$-\frac{\widehat{y}_{i,t} + \varphi_{i,t}}{\bar{\epsilon}} = \widehat{p}_{i,t} - \varphi_{i,t},$$

which yields Equation (A.27). Along the balanced growth path, the bracket in Equation (A.17) is zero. Changes in the sales and elasticity weights therefore do not enter its first-order expansion. Using the desired-markup slope $-\Omega/(1-\Omega)$ gives

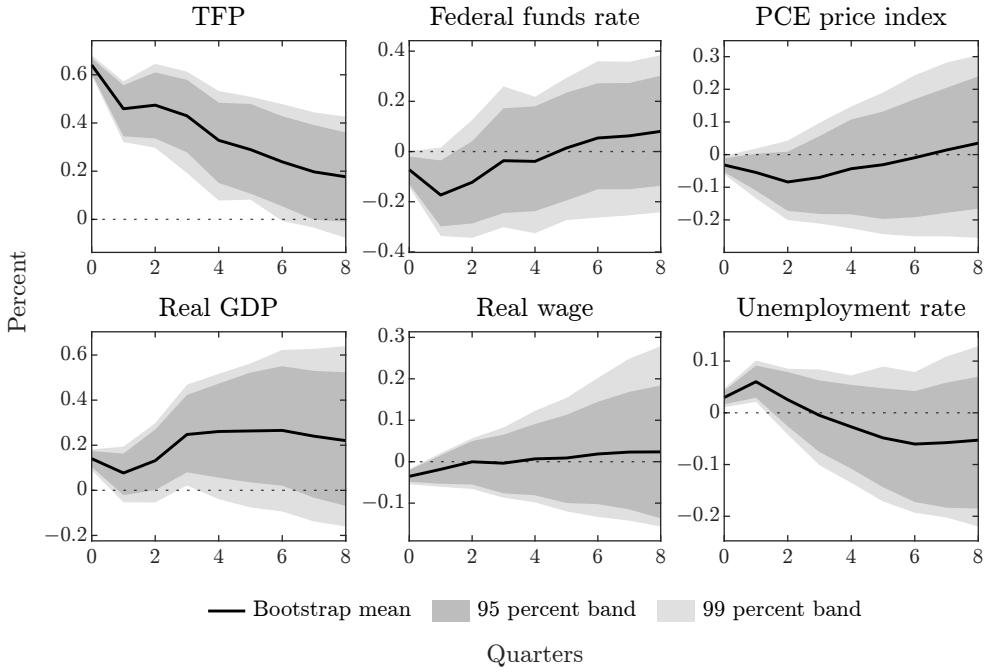
$$\mathbb{E}_t\sum_{h=0}^{\infty}(\beta\theta)^h\left[\widehat{p}^o_{i,t} - (1-\Omega)\widehat{mc}_{i,t+h|t} - \Omega\varphi_{i,t+h}\right] = 0.$$

For the scalar pricing block, we maintain the approximation of replacing marginal cost along the unchanged-price path with the marginal-cost process in the main text. Summing the geometric weights then gives

$$\widehat{p}^o_{i,t} = (1-\beta\theta)\sum_{h=0}^{\infty}(\beta\theta)^h\mathbb{E}_t\left[(1-\Omega)\widehat{mc}_{i,t+h} + \Omega\varphi_{i,t+h}\right].$$

Separating the current term and applying iterated expectations gives Equation (A.28).

Figure A.1: Responses of US aggregates to a TFP shock



Notes. The figure reports impulse responses to a one-standard-deviation TFP shock (0.64 percent), identified recursively with TFP ordered first. The vector autoregression is estimated on quarterly US data over 1973Q1–2019Q4 and includes four lags and a constant. The variables are the log of utilization-adjusted TFP (Fernald, 2014); the proxy federal funds rate (Choi et al., 2022); the log PCE price index; log real GDP (Brave et al., 2019); the log real wage, measured as average hourly earnings of production and nonsupervisory employees deflated by the core PCE price index; the unemployment rate; and the excess bond premium (Gilchrist and Zakrajšek, 2012), whose response is not shown. Dark and light gray shaded areas indicate 95 percent and 99 percent bootstrap confidence bands, respectively.

C Aggregate evidence on the RBC mechanism

Figure A.1 reports the aggregate evidence summarized in the introduction, using quarterly US data from 1973Q1 to 2019Q4. We estimate a seven-variable structural VAR with four lags and a constant, using the data sources and variable definitions listed in the figure notes. Productivity shocks are identified recursively with utilization-adjusted TFP ordered first, allowing the other variables to respond contemporaneously to a TFP innovation. Panel A traces responses to a one-standard-deviation shock, which raises TFP by 0.64 percent on impact, while Panel B reports its contribution to forecast-error variances at horizons of zero, four, and eight quarters. At one- and two-year horizons, productivity shocks explain roughly one-tenth of the forecast-error variance of GDP and less than 5 percent of that of prices, wages, and unemployment.

D Data sources and data cleaning

This appendix describes the administrative sources and data cleaning procedures used to implement the empirical framework.

PRODCOM. We use PRODCOM to calculate quarterly changes in firm-level prices. PRODCOM — from the French PRODuction COMmunautaire — is a large-scale survey commissioned by Eurostat and administered in Belgium by the National Statistical Office. Its sampling strategy aims to cover at least 90% of domestic production value in each manufacturing industry (4-digit NACE codes). It surveys all firms operating in the country with either (i) at least 20 employees or (ii) total revenue above 4.5 million euros (European Commission 2014). Each month, firms must report product-specific physical quantities (e.g., volume, kg., m^2, etc.) and values in euros of production sold for all their manufacturing products. PRODCOM identifies products by an 8-digit PC code (e.g., 15.20.13.61 is "Men's sandals", 15.20.13.62 is "Women's sandals", and 15.20.13.63 is "Children's sandals"). The first four digits identify the industry (e.g., 15.20 is "Manufacture of footwear"). Sectors are broader groupings within the economy, defined by the first two digits of the product codes (e.g., CB is "Textile, apparel, leather and related products").

The raw data contain approximately 4,000 product headings across 13 manufacturing sectors. PC codes were revised several times between 1995 and 2023, including a major overhaul in 2008. We combine Eurostat conversion tables with firms' product portfolios to harmonize 8-digit product codes across consecutive months and 4-digit industry codes over time.[11] In most cases, the tables uniquely map 8-digit product codes across consecutive years. A limited number of mappings are many-to-one, one-to-many, or many-to-many. Many-to-one mappings are straightforward. We resolve most potentially ambiguous one-to-many and many-to-many cases using each firm's product basket.[12] We drop the limited number of observations for which the available information does not resolve the mapping.

We construct product-level prices (unit values) as domestic sales divided by domestic quantities sold. We first subtract export quantities and sales using Belgian Customs data for extra-EU trade ("Extrastat") and the Intrastat Inquiry for intra-EU trade. Eurostat's official conversion tables map the CN product codes in the trade data to PRODCOM product codes.

[11] The official conversion tables are available at the Eurostat webpage. We map NACE Rev.1 industry codes, used before 2008, to NACE Rev.2 codes, used from 2008.

[12] For example, consider a case where the official mapping indicates that product 11.11.11.11 in year t could map to either 22.22.22.21 or 22.22.22.22 in year $t+1$. Suppose two firms, f_1 and f_2, report in period t sales of product 11.11.11.11 in year t. If f_1 reports *only* sales of 22.22.22.21 and f_2 *only* reports sales of 22.22.22.22 in year $t+1$ we infer that we should map 11.11.11.11 to 22.22.22.21 for the former and 11.11.11.11 to 22.22.22.22 for the latter.

Most CN-to-PC mappings are one-to-one or many-to-one and are straightforward. We drop observations with one-to-many or many-to-many mappings.

Social security declarations and value added tax declarations. We use two administrative sources to measure intermediate-input outlays, investment flows, and labor expenses. First, quarterly Social Security declarations filed by each Belgian firm with the Department of Social Security in Belgium provide labor costs (the wage bill). We remove a fixed component of the wage bill that does not scale with hours worked. Second, periodic tax declarations report firm-level purchases of intermediate inputs and investment goods. We use the investment purchases to reconstruct the capital stock by the perpetual inventory method (PIM), initialized with the capital value reported in Annual Accounts submitted to the NBB.

E Tests of identification assumptions

This appendix details the evidence summarized in Section 5. Assumption 3 states that the endogenous vector $\boldsymbol{x}_{i,t} = \left[q_{i,t}, y_{i,t}, p_{i,t}, \ell_{i,t}, m_{i,t}, k_{i,t}, u_{i,t}\right]^\top$ contains exactly two independent unit roots. With $n = 7$ variables and $r = 2$ unit roots, the cointegrating rank is $R = n - r = 5$, and the five relations in Theorem 1 span the cointegrating space. Inventories and utilization supply the two stationarity conditions in Assumption 4. Appendix E.1 describes samples and tests; Appendices E.2 to E.3 examine the nine variables and competitors' price, the five relations, and the ranks of the expenditure and inputs systems. The panel unit-root null is that every firm's series has a unit root. Unless stated otherwise, rejection refers to the 5 percent level.

E.1 Samples, data treatments, and tests

Samples. Unit-root tests for the nine variables and competitors' price index use a balanced panel of 101 firms observed with all these series in all 116 quarters, from 1995:Q1 to 2023:Q4. Expenditure rank tests start from 102 firms observed in all 116 quarters with the seven expenditure-system series. Pedroni tests use 102 firms observed in all 116 quarters with the price and its six regressors. These panels differ from the estimation samples in Section 6, used only for the final diagnostics in Appendix E.3.

Data treatments. The variables' unit-root tests use raw log levels and log levels net of firm-by-quarter-of-year means, which also removes seasonality. Pricing cointegration and

expenditure rank tests use the same two treatments. Unit-root tests of the five long-run relations use raw components and components net of industry-by-quarter means.

Panel unit-root tests. Table A.1 reports three tests for the individual variables. The Im–Pesaran–Shin (IPS) test of Im et al. (2003) averages firm-level augmented Dickey–Fuller (ADF) statistics and reports the standardized W-t-bar statistic. The Fisher–ADF test combines the p-values of firm-level ADF tests (Dickey and Fuller, 1979; Maddala and Wu, 1999); we report the inverse-normal statistic of Choi (2001). The cross-sectionally augmented IPS (CIPS) test of Pesaran (2007) augments the firm-level regressions with cross-sectional averages to allow for a common factor and reports the t-bar statistic. In the levels regressions, all three tests include firm intercepts and four lagged differences of the tested series. The null is that every firm's series has a unit root; rejection does not establish stationarity for every firm.

E.2 Tests of stationarity and cointegration

Unit-root tests on the variables. Table A.1 reports the results in raw log levels (Panel A) and after removing firm-by-quarter-of-year means (Panel B). The rejection pattern is the same under both treatments. All three tests reject a unit root at the 1 percent level for production, intermediate expenditure, real intermediates, and utilization; none rejects for capital. For own and competitors' prices, IPS and Fisher–ADF do not reject, whereas CIPS rejects at the 1 percent level. For the wage bill, sales, and labor, IPS and Fisher–ADF reject at the 1 percent level, whereas CIPS does not reject at the tabulated 5 percent critical value.

In first differences, all three tests reject a unit root in all ten series under both treatments, using three lagged differences and firm intercepts. These results support excluding higher orders of integration. Separate within-firm AR(1) estimates also indicate substantial persistence: the coefficient is 0.997 for competitors' prices and 0.988 for capital under both treatments. These coefficients are distinct from the cross-firm medians of sample autocorrelations in Figure 2.

Unit-root tests on the wedges. Panels A and B of Table A.2 report IPS, Fisher–ADF, and CIPS tests with intercepts and four lagged differences, using relations constructed from raw and demeaned components. The null is a unit root, excluding a stationary long-run relation. All three tests reject for all five benchmark relations at the 1 percent level. The production-function Fisher–ADF statistic is −20.473. For the pricing wedge, it is −5.347. After industry-by-quarter demeaning, all three tests reject for all five benchmark relations at the 5 percent level. All these rejections are at 1 percent except CIPS for relative inputs, which rejects at 5 percent. All three

Table A.1: Panel unit-root tests on the nine log variables and competitors' price

	IPS, constant	Fisher–ADF	CIPS
Panel A: Raw log levels			
Production, $q_{i,t}$	-16.316^{***}	-17.176^{***}	-2.958^{***}
Sales, $y_{i,t}$	-5.025^{***}	-5.030^{***}	-1.927
Output price, $p_{i,t}$	4.897	3.509	-2.348^{***}
Wage bill, $w_{i,t} + \ell_{i,t}$	-3.901^{***}	-3.632^{***}	-1.826
Intermediate expenditure, $p^m_{i,t} + m_{i,t}$	-3.325^{***}	-3.202^{***}	-2.419^{***}
Labor, $\ell_{i,t}$	-3.720^{***}	-3.384^{***}	-1.937
Intermediates, $m_{i,t}$	-7.082^{***}	-7.231^{***}	-2.281^{***}
Capital, $k_{i,t}$	1.220	1.350	-1.269
Utilization, $u_{i,t}$	-16.680^{***}	-17.399^{***}	-2.833^{***}
Competitors' price, $p_{-i,t}$	13.848	13.317	-2.642^{***}
Panel B: Firm-by-quarter-of-year means removed			
Production, $q_{i,t}$	-15.772^{***}	-16.605^{***}	-2.891^{***}
Sales, $y_{i,t}$	-4.677^{***}	-4.659^{***}	-1.952
Output price, $p_{i,t}$	4.691	3.719	-2.366^{***}
Wage bill, $w_{i,t} + \ell_{i,t}$	-3.751^{***}	-3.466^{***}	-1.747
Intermediate expenditure, $p^m_{i,t} + m_{i,t}$	-3.067^{***}	-2.937^{***}	-2.463^{***}
Labor, $\ell_{i,t}$	-3.322^{***}	-2.960^{***}	-1.822
Intermediates, $m_{i,t}$	-6.708^{***}	-6.859^{***}	-2.310^{***}
Capital, $k_{i,t}$	0.993	1.285	-1.255
Utilization, $u_{i,t}$	-16.815^{***}	-17.538^{***}	-2.820^{***}
Competitors' price, $p_{-i,t}$	13.369	12.963	-2.595^{***}

Notes. The balanced panel contains 101 firms observed for 116 quarters with all nine variables and the competitors' price. All tests include firm intercepts and four lagged differences; CIPS also includes cross-sectional averages. IPS reports the W-t-bar statistic, Fisher–ADF the inverse-normal statistic, and CIPS the t-bar statistic. One, two, and three asterisks denote rejection of the unit-root null at the 10, 5, and 1 percent levels. For IPS and Fisher–ADF, the levels apply to the tests' p-values. For CIPS, they apply to the critical values of the truncated t-bar statistic that Pesaran (2007) tabulates.

tests also reject for the demeaned seven-variable pricing alternative.

Cointegration tests on the pricing equation. Panel C of Table A.2 reports the panel cointegration tests of Pedroni (1999, 2004), which does not impose the restrictions on the parameters. The null is no cointegration. All three statistics reject at the 1 percent level in both raw data and data net of firm-by-quarter-of-year means. Figure A.2 complements these tests with the distribution of the sample autocorrelations of the five relations, computed after

Table A.2: Tests on long-run relations and pricing cointegration tests

	IPS, constant	Fisher–ADF	CIPS
Panel A: Long-run relations			
Pricing wedge, $-\omega^{p}_{i,t}$	-5.300^{***}	-5.347^{***}	-2.216^{***}
Output wedge, $-a_{i,t}$	-19.557^{***}	-20.473^{***}	-3.459^{***}
Relative inputs, $m_{i,t} - \ell_{i,t}$	-3.850^{***}	-3.673^{***}	-2.552^{***}
Inventories, $q_{i,t} - y_{i,t}$	-21.063^{***}	-22.019^{***}	-3.541^{***}
Utilization, $u_{i,t}$	-14.595^{***}	-15.193^{***}	-2.828^{***}
Panel B: Long-run relations in deviation from industry-time FE			
Pricing relation	-4.939^{***}	-4.981^{***}	-2.285^{***}
Production function, $-a_{i,t}$	-19.609^{***}	-20.489^{***}	-3.365^{***}
Relative inputs, $m_{i,t} - \ell_{i,t}$	-4.798^{***}	-5.388^{***}	-2.113^{**}
Inventories, $q_{i,t} - y_{i,t}$	-20.402^{***}	-21.355^{***}	-3.434^{***}
Utilization, $u_{i,t}$	-22.181^{***}	-22.758^{***}	-3.622^{***}
Panel C: Cointegration tests (no parameter restrictions)			
	Raw statistic	Firm-seasonal statistic	
Pedroni modified PP	-45.048^{***}	-39.548^{***}	
Pedroni PP	-48.683^{***}	-45.438^{***}	
Pedroni ADF	-48.378^{***}	-46.188^{***}	

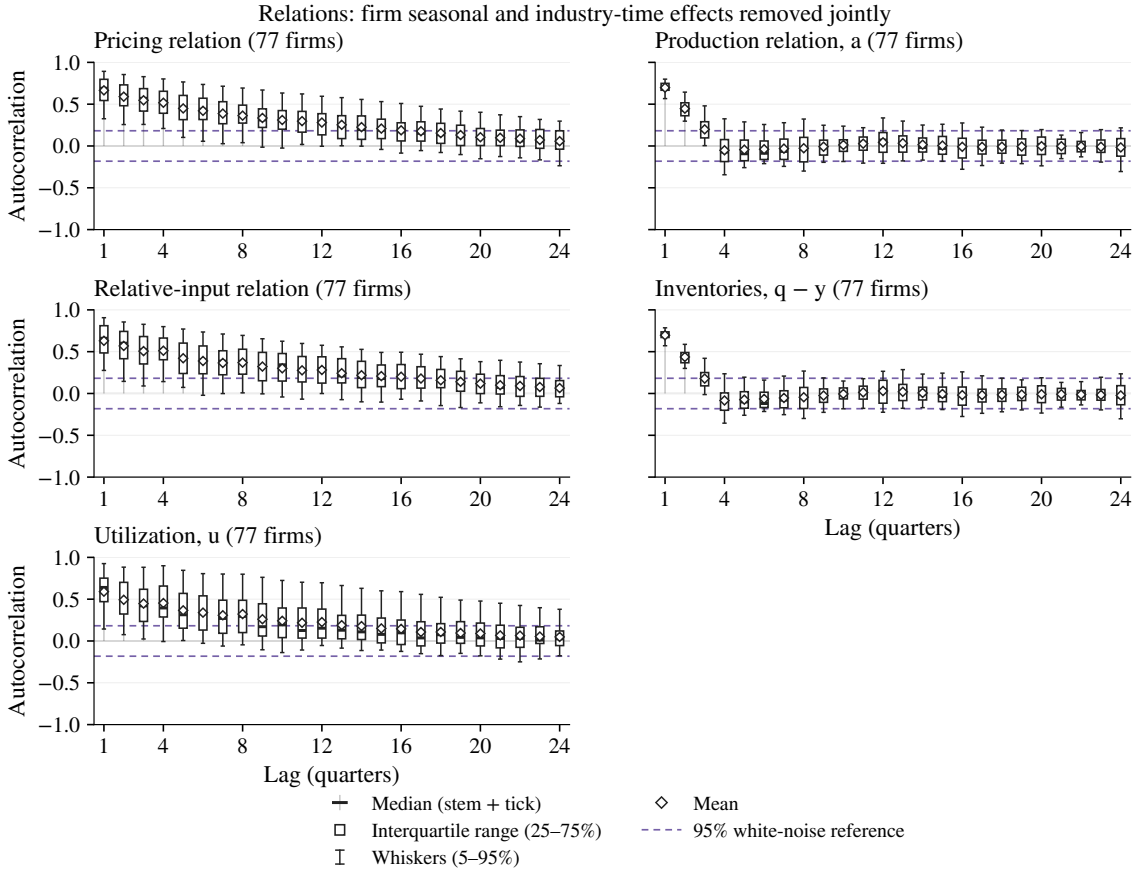
Notes. Panels A and B use the balanced panel of 77 firms whose log production–sales difference varies over time. Tests in Panels A and B include intercepts and four lagged differences; CIPS also includes cross-sectional averages. IPS reports the W-t-bar statistic, Fisher–ADF the inverse-normal statistic, and CIPS the t-bar statistic. One, two, and three asterisks denote rejection at the 10, 5, and 1 percent levels of the unit-root null in Panels A and B and of the no-cointegration null in Panel C. For IPS, Fisher–ADF, and the tests of Panel C, the levels apply to the p-values of the tests. For CIPS, they apply to the critical values of the truncated t-bar statistic that Pesaran (2007) tabulates. Panel C uses a balanced panel of 102 firms with the seven variables of the pricing equation and regresses prices on the wage bill, intermediate expenditure, sales, capital, utilization, and the competitors' price with free coefficients. These tests use their default lag and deterministic specifications. Their null is no cointegration. All panels span 116 quarters.

firm-by-quarter-of-year and industry-by-quarter fixed effects are jointly removed.

E.3 The rank of the system

Panel trace test. We assess rank directly with the panel-mean trace test of Larsson et al. (2001), applied separately to the system with expenditures ($w_t + \ell_{i,t}$ and $p^m_t + m_{i,t}$) and the system with real inputs ($\ell_{i,t}$ and $m_{i,t}$). The expenditure system replaces labor and intermediates in $\boldsymbol{x}_{i,t}$ with the wage bill and intermediate expenditure. Let $\boldsymbol{z}_{i,t}$ denote the seven-variable vector for each test.

Figure A.2: Autocorrelation of the long-run residuals



Notes. Sample autocorrelations at lags of one to 24 quarters, computed separately for 77 firms observed in all 116 quarters from 1995Q1 to 2023Q4, with a time-varying production–sales difference and all five long-run relations available: the pricing relation, the production-function residual, relative inputs, inventories, and utilization. The relations are evaluated at the benchmark estimated coefficients, with firm-by-quarter-of-year and industry-by-quarter fixed effects removed jointly. At each lag, the stem and horizontal tick mark the cross-firm median, the box spans the 25th–75th percentiles, the hollow diamond marks the mean, and the whiskers span the 5th–95th percentiles. Firms receive equal weight. The dashed lines are the approximate pointwise 95 percent white-noise reference bounds.

The statistic averages firms' trace statistics of Johansen (1991), each computed from the firm-level VECM

$$\Delta \boldsymbol{z}_{i,t} = \boldsymbol{c}_i + \Pi_i \boldsymbol{z}_{i,t-1} + \sum_{j=1}^{L} B_{i,j} \Delta \boldsymbol{z}_{i,t-j} + \boldsymbol{\epsilon}_{i,t}.$$

The firm's long-run matrix is Π_i, of dimension 7×7, and the $B_{i,j}$ are 7×7 short-run coefficient matrices. In the system with real inputs, Π_i is the firm-level counterpart of $\Lambda\delta^\top$ in Theorem 1.

The firm's innovation $\epsilon_{i,t}$ differs from the pooled innovation in Equation (9). Within each system, the rank R of Π_i is assumed common across firms. We test $R_0 = 0, 1, \ldots, 6$ sequentially under $H_0 : R \leq R_0$, selecting the first bound not rejected at 5 percent. We simulate the null distribution instead of using the asymptotic moments of Larsson et al. (2001).[13]

Table A.3 reports the last rejected and first retained rank bounds at each lag order. In the expenditure system, raw and firm-seasonal data both select ranks five, four, three, and zero at $L = 0$, 1, 3, and 7, respectively. Rank five is selected in two of the four specifications with at most one lagged difference; no expenditure specification selects a higher rank. In the inputs system, rank five is selected with zero lagged differences under both raw and defactored data, and with one lagged difference in defactored data: three of eight specifications. One lagged difference in raw data selects rank four; three lagged differences select rank three under both treatments, and seven select rank zero. With defactored data and zero lagged differences, the rank-five null is retained at a p-value of 0.070. Rank five is thus selected in three of the four specifications with at most one lagged difference, leaving 114 and 106 degrees of freedom per equation, and no inputs-system specification selects a higher rank.

Diagnostics of the estimated system. Finally, we compare the restricted system with an unrestricted VAR on the estimation sample in Section 6. Both use 80,921 firm-quarter observations, the same fixed effects, and 24 lagged differences. The restricted system has two imposed unit roots, no explosive roots, and a largest stable root of 0.939. The unrestricted VAR's spectral radius, the largest absolute companion root, is 0.969, with no root above 0.990. It is formally stationary but highly persistent. The restricted long-run matrix $\Pi \equiv \Lambda\delta^\top$ has two zero singular values; the unrestricted matrix's two smallest singular values are 0.065 and 0.025. With about 32 usable quarters per firm, within-firm estimation biases estimated persistence downward (Nickell, 1981). Applied to independent Gaussian random walks with the dimensions of the estimation sample, that is, 2,526 firms, 32 usable quarters, 25 lags in levels, and firm-by-quarter-of-year effects, it yields a largest root of about 0.94 and a long-run coefficient, the univariate analogue of a singular value of Π, of about -0.09, against true values of one and zero. The near-unit roots and the two

[13] The null distribution of the average trace statistic is simulated, not resampled from the data. Each of 400 simulated panels has the same number of firms and quarters of the data. In each simulated firm, $7 - R_0$ series are independent driftless Gaussian random walks and R_0 series are independent Gaussian white noise. The simulated data therefore have rank R_0, no short-run dynamics, and, before defactoring, no dependence across firms. The simulated panels are defactored whenever the data are, and each simulated firm's VECM has the same L and constant as in the data. The simulated panels are firm-seasonally demeaned whenever the data are. The p-value is the share of simulated averages at least as large as the observed one, so it is conditional on this simulated process. The column "Lagged differences" reports the number L of lags of $\Delta z_{i,t}$ in each firm's VECM, so that the VAR in levels has order $L + 1$.

Table A.3: Panel rank tests by system, lag order, and data treatment

	Lagged differences	Degrees of freedom	Firms	Last rejected R_0	First retained R_0	Selected rank
Panel A: Expenditure system						
Raw	0	114	78	4^{***}	5	5
Raw	1	106	78	3^{***}	4	4
Raw	3	90	78	2^{***}	3	3
Raw	7	58	76	–	0	0
Firm-seasonal	0	114	78	4^{***}	5	5
Firm-seasonal	1	106	78	3^{***}	4	4
Firm-seasonal	3	90	78	2^{***}	3	3
Firm-seasonal	7	58	77	–	0	0
Panel B: Inputs system						
Raw	0	114	77	4^{***}	5	5
Raw	1	106	77	3^{***}	4	4
Raw	3	90	77	2^{**}	3	3
Raw	7	58	76	–	0	0
Defactored	0	114	77	4^{***}	5^{*}	5
Defactored	1	106	77	4^{***}	5	5
Defactored	3	90	77	2^{***}	3	3
Defactored	7	58	77	–	0	0

Notes. Panel A reports the expenditure system $[q, y, p, w + \ell, p^m + m, k, u]$a; its underlying balanced panel contains 78 firms observed for 116 quarters whose log production–sales difference varies over time. Panel B reports the inputs system $[q, y, p, \ell, m, k, u]$. All specifications include an unrestricted constant. Firm-seasonal demeaning removes firm-by-quarter-of-year means. Defactoring removes the cross-sectional mean at each date. The column "Degrees of freedom" reports the degrees of freedom of each equation of the firm's VECM after the constant and the $7L$ lagged differences and before the seven lagged levels, $(116 - L - 1) - (7L + 1) = 114 - 8L$. One, two, and three asterisks denote rejection of $H_0 : R \leq R_0$ at the 10, 5, and 1 percent levels, that is, a simulated p-value of at most 0.10, 0.05, and 0.01. The selected rank applies the 5 percent level sequentially: it is the first bound that is not rejected at 5 percent. An asterisk on the first retained bound therefore indicates a rejection at 10 percent only. A dash indicates that the first bound, $R_0 = 0$, is retained, so that no bound is rejected in the sequential procedure.

small singular values of the unrestricted VAR are therefore what one would expect if the system has the two unit roots that the restriction imposes.

F Robustness for Section 6

F.1 Control-function estimates of the production function

Table A.4 reports the output elasticities estimated by Gagliardone et al. (2025) using the same data source and a control-function approach. Appendix B.3 of Gagliardone et al. (2025) describes the estimator, which builds on the two-stage routine of Gandhi et al. (2020); we refer to that

study for the identification argument. The first stage regresses the intermediate share in revenue on a polynomial in inputs and the output price, recovering the ex-post productivity shock and controlling for the markup. The second stage estimates output elasticities by generalized method of moments using the productivity law of motion, with labor and intermediates chosen flexibly and capital predetermined. A second control function in prices and market shares corrects for input expenditures deflated by industry price indices.

These estimates differ from ours in sample, data construction, and specification. Four differences concern the sample and data. First, the source panel in Gagliardone et al. (2025) covers 84 quarters, from 1999:Q1 to 2019:Q4, with 4,598 firms and 132,915 firm-quarter observations; the production function is estimated on yearly data. Our first stage uses 147,244 quarterly observations on 4,199 firms from a dataset spanning 1995:Q1 to 2023:Q4. Second, Gagliardone et al. (2025) retain each firm's longest continuous spell in each industry. We allow gaps in firms' histories, and each regression uses observations with the required lags. Third, their output is sales deflated by the firm's price index, corresponding to our $y_{i,t}$. Our output equation uses production, $q_{i,t}$, which adds the change in inventories to sales. Fourth, their capital elasticity applies to the stock of fixed tangible assets, $k_{i,t}$, whereas ours applies to capital services, $k_{i,t} + u_{i,t}$, incorporating utilization. Both papers use the same labor and intermediate-input sources: Social Security wage bills and VAT purchases of materials and services, each deflated by a price index.

Two further differences concern specification. First, their production function is Cobb–Douglas with sector-specific elasticities; ours is CES with one elasticity vector for all firms. Second, their first stage uses the intermediates first-order condition, which links the output elasticity of intermediates to their revenue share adjusted for the markup. Our estimator does not require this condition, as discussed in Section 6.1.

F.2 Dynamics of the estimated system

The VECM in Panel B of Table 4 has 1,176 coefficients on lagged differences. Table A.5 summarizes their implied dynamics without listing the coefficients.

Panel A reports the persistence profiles of the five long-run relations, following Pesaran and Shin (1996). Let Ψ_h denote the response of $\boldsymbol{x}_{i,t+h}$ to the reduced-form innovation $\boldsymbol{\epsilon}_{i,t}$ of Equation (9), and let Σ denote the covariance matrix of the innovations. The persistence profile of the j-th relation, $\delta_j^\top \boldsymbol{x}_{i,t}$, is

$$\mathrm{PP}_j(h) = \left(\frac{\delta_j^\top \Psi_h \, \Sigma \, \Psi_h^\top \delta_j}{\delta_j^\top \Sigma \, \delta_j} \right)^{1/2}.$$

The profile measures the standard deviation of the relation's response at horizon h to a shock

Table A.4: Control-function estimates of the output elasticities

	Output elasticities			*Returns to scale*	
Sector	Labor θ^L	Intermediates θ^M	Capital θ^K	Short-run $\theta^L + \theta^M$	Long-run $\theta^L + \theta^M + \theta^K$
CA	.214	.752	.026	.966	.992
CB	.200	.764	.028	.964	.992
CC	.230	.731	.033	.960	.994
CE	.217	.720	.039	.937	.976
CG	.219	.768	.008	.987	.996
CH	.188	.776	.028	.963	.991
CJ	.177	.782	.036	.960	.996
CK	.247	.684	.054	.932	.986
CL	.243	.725	.036	.968	1.004
Aggregate	.212	.753	.026	.965	.992

Notes. The table reports the Cobb–Douglas output elasticities estimated by Gagliardone et al. (2025), their Table A.2, for each manufacturing sector of the A*38 classification: CA food products, beverages, and tobacco; CB textiles, apparel, and leather; CC wood and paper products, and printing; CE chemicals; CG rubber and plastics, and other non-metallic mineral products; CH basic metals and fabricated metal products; CJ electrical equipment; CK machinery and equipment n.e.c.; CL transport equipment. The short-run returns to scale are the sum of the elasticities of the variable inputs, labor and intermediates. The long-run returns to scale add the elasticity of capital. The sector CI, computer, electronic, and optical products, has too few observations for the estimation, and Gagliardone et al. (2025) set its elasticities to the average of those of CH and CJ. The last row reports their aggregate estimates.

with the innovations' covariance, relative to its impact value. It is independent of structural-shock identification. It equals one on impact and converges to zero because the relation is stationary. Its half-life is the first horizon after which it remains at or below one half.

Panel B reports the companion-matrix roots of the levels representation with the largest moduli. A root λ has a half-life of $\ln(0.5)/\ln|\lambda|$ quarters; a complex pair with argument ϕ has a period of $2\pi/|\phi|$ quarters.

Three features stand out. First, the pricing wedge, relative inputs, and utilization fall below half their impact deviation within one quarter. The pricing and relative-input profiles then decay slowly: they are .18 and .18 after eight quarters, and .05 and .04 after twenty-four.

Second, the production-function and inventory profiles initially decay more slowly. They exceed .74 after two quarters and fall to about .16 by the fourth.

Third, the largest stable root is .939, with a half-life of 11.0 quarters; the second largest is

Table A.5: Dynamics of the estimated system

	Panel A: persistence profiles of the long-run relations						
	$h = 1$	$h = 2$	$h = 4$	$h = 8$	$h = 12$	$h = 24$	Half-life
Pricing relation, $\omega^{p}_{i,t}$	.384	.345	.242	.176	.125	.046	1
Production function, $-a_{i,t}$	.811	.747	.163	.087	.052	.020	4
Relative inputs, $m_{i,t} - \ell_{i,t}$	.441	.384	.266	.185	.126	.045	1
Inventories, $q_{i,t} - y_{i,t}$	.829	.761	.159	.075	.011	.017	4
Utilization, $u_{i,t}$	.346	.231	.090	.014	.020	.037	1

	Panel B: leading roots of the companion matrix				
	Modulus	Real	Imaginary	Half-life	Period
Roots 1–2	1.000	1.000	.000	∞	–
Root 3	.939	.939	.000	11.0	–
Root 4	.921	.921	.000	8.5	–
Roots 5–6	.914	.908	±.103	7.7	55.5
Roots 7–8	.910	-.158	±.896	7.4	3.6
Roots 9–10	.909	-.895	±.158	7.3	2.1
Roots 11–12	.908	-.894	±.159	7.2	2.1
Stable roots above .90	26				

Notes. Panel A reports the persistence profile $\mathrm{PP}_j(h)$ of each long-run relation at a horizon of h quarters, and its half-life in quarters. Panel B reports the roots of the companion matrix of the VECM in levels, with 25 lags, ordered by modulus. A complex pair occupies one row, with the imaginary part reported as ±. The two unit roots are imposed by the rank restriction. The half-life of a root is $\ln(0.5)/\ln|\lambda|$, and the period of a complex pair is $2\pi/|\phi|$, where ϕ is its argument. The covariance matrix Σ uses the 80,636 observations at which all seven residuals are available. The entries are point estimates; no standard errors are computed.

.921, with a half-life of 8.5 quarters. The next roots form complex pairs with moduli between .908 and .914. In total, 26 of the 173 stable roots have moduli above .90. Many roots have similar moduli, as is typical of a long lag polynomial, so we do not interpret individual pairs' periods as distinct cycles.

F.3 The estimated law of motion of TFPQ

The VECM in Panel B of Table 4 implies a 25-lag vector autoregression in levels. Expressed in the Cholesky-ordering coordinates, its TFPQ equation is

$$a_{i,t} = \sum_{j=1}^{25} \left(c_j^k k_{i,t-j} + c_j^a a_{i,t-j} + c_j^r r_{i,t-j} + c_j^\ell \ell_{i,t-j} + c_j^m m_{i,t-j} + c_j^u u_{i,t-j} + c_j^p p_{i,t-j} \right) + \varepsilon_{i,t}^a, \quad \text{(A.31)}$$

where $r_{i,t} = a_{i,t} + p_{i,t} + y_{i,t} - q_{i,t}$ is TFPR. All variables are deviations from firm-by-quarter-of-year and industry-by-quarter fixed effects. The quantitative model in Section 8 takes Equation (A.31) as the true TFPQ law of motion. Table A.6 reports its 175 coefficients.

Table A.6: Coefficients of the estimated law of motion of TFPQ

Lag j	Capital $k_{i,t-j}$	TFPQ $a_{i,t-j}$	TFPR $r_{i,t-j}$	Labor $\ell_{i,t-j}$	Materials $m_{i,t-j}$	Utilization $u_{i,t-j}$	Price $p_{i,t-j}$
1	−.053	.832	−.419	.016	.163	.024	.210
2	−.015	.070	.040	.023	−.024	.058	−.088
3	.051	−.048	.105	−.018	−.020	−.015	−.087
4	−.022	−.707	.501	−.101	−.096	−.005	−.593
5	.058	.579	−.373	.075	.077	−.034	.459
6	−.033	.087	−.012	−.015	.012	−.023	.051
7	.025	−.053	.075	.005	−.033	.088	−.063
8	−.018	−.489	.353	−.028	−.018	−.063	−.406
9	.036	.395	−.261	−.032	.022	.012	.283
10	.006	.070	−.014	.082	.001	−.043	.063
11	−.005	−.042	.050	−.036	.009	.043	−.089
12	−.053	−.302	.192	−.009	−.048	.050	−.175
13	.067	.241	−.158	−.011	.025	−.006	.185
14	.042	.049	−.012	.005	.014	.000	.048
15	−.104	−.025	.004	.027	−.024	.023	−.027
16	−.014	−.207	.145	.020	−.009	−.029	−.140
17	.054	.161	−.095	−.034	.013	.077	.109
18	−.032	.029	−.010	−.046	.015	−.084	.029
19	.020	−.012	−.020	.045	−.035	.008	−.033
20	.015	−.144	.075	.025	−.020	.033	−.095
21	−.011	.111	−.056	−.049	.026	−.009	.119
22	.034	.014	.031	.023	.011	.048	−.049
23	−.077	−.005	−.011	−.007	−.043	−.019	.021
24	.081	−.081	.045	.016	.007	.001	−.035
25	−.060	.059	−.053	−.009	.023	−.064	.013
Sum	−.008	.582	.120	−.032	.048	.070	−.292

Notes. Coefficients c_j^x of Equation (A.31). The entry in row j and column x multiplies $x_{i,t-j}$. The coefficients are the second row of TA_jT^{-1}, where A_j is the matrix of coefficients on lag j in the levels representation of the VECM (9), and T maps $\boldsymbol{x}_{i,t}$ into $(k_{i,t}, a_{i,t}, r_{i,t}, \ell_{i,t}, m_{i,t}, u_{i,t}, p_{i,t})$. They are linear combinations of the estimated coefficients of the VECM. The last row sums each column over the 25 lags. The entries are point estimates; no standard errors are computed.

G Bayesian estimation and historical decomposition

We estimate the shock processes on the firm panel, holding technology and adjustment-cost parameters at the values used to match the impulse responses. Dynare supplies the model solution, prior densities, posterior sampler, and smoother. The panel dimension enters through sample construction and the likelihood. We pool the likelihood across firms with common parameters but separate histories of states and shocks, then compute the historical decomposition separately for each history.

G.1 Panel and observables

Sample. We start from the 2,526 firms in the second-stage VECM estimation and use their available histories, including quarters excluded from the VECM regression because lagged regressors are missing. We retain one unit per VAT group: the unit with the most observations or, in a tie, the lowest identifier. This leaves 2,313 units before growth rates are constructed and fixed effects removed.[14]

The observables are quarterly log differences of production, sales, the price, labor, materials, and capital, measured in log points. Utilization is endogenous in the model but unobserved in the likelihood. We compute differences only between consecutive quarters of the same unit and jointly remove firm-by-quarter-of-year and industry-by-quarter fixed effects from each growth rate. For labor, the firm-by-quarter-of-year effects are estimated separately before and after the measurement change in 2003. We exclude the growth rate in 2003Q1 from all six series before removing fixed effects. If any of the six regressions drops an observation as a singleton, we remove it from all six and repeat the regressions on the common sample until it stops changing.

Spells. A spell is a maximal sequence of consecutive quarters within a firm with all six observables available. A missing observation in any series ends the spell; the next complete sequence starts a new one, even for the same firm. The final sample contains 141,164 firm-quarter observations from 2,311 firms in 4,909 spells. Lengths range from one to 83 quarters, with a median of 27. We retain every spell, including those lasting one quarter.

Each spell has its own initial state. The filter restarts at its beginning, carrying no information across gaps or the 2003 break. The likelihood therefore treats spells of the same firm as separate histories. Firm and calendar-time identifiers determine fixed effects and spell boundaries; after this transformation, the filtering recursion is indexed by quarters since the spell began.

G.2 Model and measurement

Common parameters. The baseline estimates 15 parameters: three persistences for the labor, materials, and investment wedges; the standard deviations of six structural innovations; five measurement-error standard deviations; and the shared price-sales measurement-error loading. The price wedge is i.i.d. Perceived TFPQ persistence and demand persistence remain fixed at their calibrated values, while the TFPQ and demand innovation standard deviations

[14] VAT groups are recovered by linking units with identical wage and materials-price observations in at least one quarter and taking the connected components of these links. The representative is selected among units in the second-stage sample.

are re-estimated jointly with those of the wedges. These parameters, like technology and adjustment-cost parameters, are common across firms. The baseline structural innovations are mutually independent.

At each candidate parameter vector, we recompute firms' decision rules under the perceived AR(1) processes and combine them with the estimated 25-lag TFPQ law. This law determines realized productivity, including its response to past endogenous variables; firms' forecasts remain those implied by the perceived processes. The estimated productivity-law coefficients remain fixed throughout estimation.

Measurement errors. Let $d_{i,t}^{x}$ denote the observed log growth rate of variable x after removing the fixed effects. The observation equations are

$$d_{i,t}^{x} = \Delta\widehat{x}_{i,t} + v_{i,t}^{x} - v_{i,t-1}^{x}, \qquad x \in \{q, \ell, m, k\}, \tag{A.32}$$

$$d_{i,t}^{p} = \Delta\widehat{p}_{i,t} + v_{i,t}^{py} - v_{i,t-1}^{py}, \tag{A.33}$$

$$d_{i,t}^{y} = \Delta\widehat{y}_{i,t} - \kappa_{py}\left(v_{i,t}^{py} - v_{i,t-1}^{py}\right). \tag{A.34}$$

The five errors $v_{i,t}^{q}$, $v_{i,t}^{py}$, $v_{i,t}^{\ell}$, $v_{i,t}^{m}$, and $v_{i,t}^{k}$ are i.i.d. over time, mutually independent, and independent of structural innovations. They enter log levels, so their contribution to growth rates is MA(1). Price and sales share one error, with loadings one and $-\kappa_{py}$. For $\kappa_{py} > 0$, their measurement errors are perfectly negatively correlated. The estimated κ_{py} determines their relative magnitude, not their correlation. All measurement errors enter the model as explicit states and innovations.

G.3 Panel likelihood and posterior

Initialization and pooling. Index spells by s and quarters within each spell by $t = 1, \ldots, T_s$. Stack the six observables in $\boldsymbol{d}_{s,t}$, and let $\boldsymbol{z}_{s,t}$ collect the model states, including those for productivity dynamics, measurement errors, and observed growth rates. For a common parameter vector ϑ, the system is

$$\boldsymbol{z}_{s,t} = A(\vartheta)\boldsymbol{z}_{s,t-1} + B(\vartheta)\boldsymbol{\varepsilon}_{s,t}, \qquad \boldsymbol{d}_{s,t} = Z\boldsymbol{z}_{s,t}, \qquad \boldsymbol{\varepsilon}_{s,t} \sim \mathcal{N}\big(0, \Sigma(\vartheta)\big). \tag{A.35}$$

The matrix Z selects the six observables. The vector $\boldsymbol{\varepsilon}_{s,t}$ contains the six structural innovations and the five measurement errors. We initialize every spell at a zero predicted mean and the same stationary prediction covariance P_1, which solves

$$P_1 = AP_1A^{\top} + B\Sigma B^{\top}. \tag{A.36}$$

We recompute the covariance at each candidate parameter vector. Every spell's first observation enters the likelihood; no initial quarters are discarded.

Let $\boldsymbol{v}_{s,t}$ denote the one-step prediction error and F_t its covariance. Conditional on the common parameters, we treat spells as independent. The log likelihood is

$$\log L(\vartheta) = -\frac{1}{2}\sum_{s=1}^{N}\sum_{t=1}^{T_s}\left[6\log(2\pi) + \log|F_t| + \boldsymbol{v}_{s,t}^{\top}F_t^{-1}\boldsymbol{v}_{s,t}\right]. \tag{A.37}$$

The likelihood sums individual firm-quarter contributions without dividing by spell length. Firms with longer observed histories therefore contribute more observations.

Priors and estimation. The posterior combines the pooled likelihood with one prior density for the common parameters:

$$p(\vartheta \mid \{\boldsymbol{d}_{s,t}\}) \propto p(\vartheta)L(\vartheta). \tag{A.38}$$

The prior enters once, rather than once per firm or spell. Wedge-process priors are reported in the main text. The six structural standard deviations have inverse gamma priors with two degrees of freedom. For TFPQ and demand, the prior medians equal the calibrated innovation standard deviations. For each wedge, the median is the innovation standard deviation that would generate the observed variance of the corresponding growth rate if that wedge acted alone, evaluated at the prior medians of the persistences.

Posterior estimates. Table A.7 reports priors and posterior estimates for all fifteen parameters. The wedge-process rows repeat Table 7 and are discussed in Section 8.2. Here we discuss the remaining estimates and the precision of all fifteen.

Posterior means are close to posterior modes except in the materials and labor blocks. For nine of the fifteen parameters, the difference is less than .06 posterior standard deviations. The six exceptions are the three parameters in each block.

The materials block comprises wedge persistence ρ_m, innovation standard deviation σ_m, and measurement-error standard deviation σ_{ν^m}. In posterior standard deviations, the mean exceeds the mode by .47 for ρ_m and .30 for σ_{ν^m}, and falls below it by .45 for σ_m. These parameters are strongly correlated in the posterior: $-.96$ between ρ_m and σ_m, .62 between ρ_m and σ_{ν^m}, and $-.72$ between σ_m and σ_{ν^m}. The posterior of ρ_m is right-skewed. Its mode, .014, lies 1.5 posterior standard deviations above zero, the lower bound of its support.

The labor block comprises wedge persistence ρ_ℓ, innovation standard deviation σ_ℓ, and measurement-error standard deviation σ_{ν^ℓ}. In posterior standard deviations, the mean exceeds the mode by .16 for ρ_ℓ and .13 for σ_{ν^ℓ}, and falls below it by .15 for σ_ℓ. These parameters are also strongly correlated in the posterior: $-.99$ between ρ_ℓ and σ_ℓ, .95 between ρ_ℓ and σ_{ν^ℓ}, and $-.97$ between σ_ℓ and σ_{ν^ℓ}. The posterior of ρ_ℓ is right-skewed. Its mode, .024, lies 2.1 posterior standard

Table A.7: Priors and posterior estimates of all estimated parameters

	Prior				Posterior			
Parameter	Distribution	Support	Mean	SD	Mode	Mean	5%	95%
			Wedge processes					
ρ_ℓ	Beta	$(0, 1)$	.500	.200	.024	.026	.009	.047
σ_ℓ	Inverse gamma	$(0, \infty)$	.066	-	.0822	.0819	.0786	.0848
ρ_m	Beta	$(0, 1)$	.500	.200	.0136	.0178	.0050	.0345
σ_m	Inverse gamma	$(0, \infty)$	.148	-	.1046	.1037	.1002	.1065
ρ_ι	Beta	$(-1, 1)$	.500	.300	$-.211$	$-.211$	$-.222$	$-.200$
σ_ι	Inverse gamma	$(0, \infty)$	.418	-	1.830	1.830	1.796	1.864
σ_p	Inverse gamma	$(0, \infty)$	.159	-	.0420	.0420	.0417	.0423
			TFPQ and demand					
σ_a	Inverse gamma	$(0, \infty)$	.713	-	.5175	.5176	.5156	.5195
σ_φ	Inverse gamma	$(0, \infty)$	.094	-	.0616	.0616	.0612	.0621
			Measurement errors					
σ_{ν^q}	Beta	$(.001, .671)$	.224	.158	.0605	.0605	.0564	.0645
σ_{ν^ℓ}	Beta	$(.001, .097)$	.033	.023	.0348	.0349	.0331	.0369
σ_{ν^m}	Beta	$(.001, .218)$	.073	.051	.1371	.1372	.1363	.1382
σ_{ν^k}	Beta	$(.001, .055)$	.019	.013	.0251	.0251	.0249	.0253
$\sigma_{\nu^{py}}$	Beta	$(.001, .233)$	.078	.055	.1235	.1235	.1230	.1241
κ_{py}	Uniform	$(0, 2)$	1.000	.577	.0853	.0854	.0733	.0973

Notes. The fifteen parameters are estimated jointly on the firm panel, as described in Appendix G.3. The rows for the wedge processes repeat Table 7. The Beta priors are rescaled to the reported support, and those of the five measurement-error standard deviations are Beta$(1, 2)$. The inverse gamma priors have two degrees of freedom, so their standard deviations are not finite. The mode is the posterior mode. The mean and the 5th and 95th percentiles are computed from the 100,000 draws retained from four Metropolis–Hastings chains.

deviations above zero. Mean-mode differences are larger in the materials block.

TFPQ and demand innovation standard deviations are close to their main-text calibrations. The TFPQ innovation standard deviation, σ_a, is calibrated at .483 to match TFPQ's impact response to its own shock. The demand innovation standard deviation, σ_φ, is calibrated at .064 from the impact response of the demand shifter implied by empirical sales and price responses to the demand shock (Section 8.1 and Table 6). The posterior mode of σ_a is .518, 7.1 percent above its calibrated value; that of σ_φ is .062, about 4 percent below. The 90 percent posterior intervals, $[.516, .519]$ and $[.0612, .0621]$, exclude the calibrated values. One interpretation is that the VECM impact responses and the panel likelihood of the six growth rates imply similar innovation scales. The model impulse responses in Section 8.1 use the calibrated values; the variance decomposition in Appendix G.4 uses the posterior modes.

Measurement error is small in production and sales but large in the price and materials. At the posterior mode, it accounts for 1.4 percent of the model's variance of observed production growth and .30 percent of that of observed sales growth. Its shares are 93.6 percent for observed price growth and 77.6 percent for observed materials growth. For labor and capital, they are 26.0 and 23.1 percent. Price and sales share one measurement error, whose standard deviation, $\sigma_{\nu^{py}}$, has a posterior mode of .123. The posterior mode of its sales loading, κ_{py}, is .0853. By Equations (A.33) and (A.34), the error moves measured sales by $-\kappa_{py}$ times its effect on the measured price. The standard deviation of the sales component, $\kappa_{py}\sigma_{\nu^{py}}$, is therefore .0105.

G.4 Forecast-error variance decomposition

We compute the forecast-error variance decomposition (FEVD) from the estimated model. The panel enters through estimation of the common parameters. Conditional on these parameters, every firm has the same FEVD, regardless of its observed history, spell length, or smoothed shocks. The denominator is the variable's model-implied forecast-error variance.

Variables and horizons. We decompose the error-free log levels of production, sales, the price, labor, materials, capital, utilization, and measured marginal cost. Marginal cost is constructed from model variables as in Equation (16). The likelihood uses growth rates, but the FEVD uses levels to match the VECM objects. Measurement errors affect parameter estimation but do not enter these levels, so their variance shares are zero.

For a parameter vector ϑ, let $\Psi_j^x(r;\vartheta)$ denote the response of variable x after r quarters to a unit innovation j. These responses follow the realized law of motion, including estimated productivity dynamics in Equation (A.31). Conditional on the model state at $t-1$, the forecast error for $x_{s,t+h}$ is

$$x_{s,t+h} - \mathbb{E}_{t-1}x_{s,t+h} = \sum_{r=0}^{h}\sum_{j=1}^{6}\Psi_j^x(r;\vartheta)\varepsilon_{s,t+h-r}^j. \tag{A.39}$$

The expectation uses the realized law of motion. Horizon h therefore denotes a forecast $h+1$ quarters ahead: $h=0$ is the impact, or one-quarter-ahead, forecast error. We compute shares through $h=24$ and report $h=0,4,8,24$.

Variance shares. The six structural innovations are mutually independent in the baseline and independent over time. Squaring Equation (A.39) and taking expectations therefore gives the share

of innovation j:

$$\mathrm{FEVD}_j^x(h;\vartheta) = 100\frac{\Sigma_{jj}(\vartheta)\sum_{r=0}^{h}\left[\Psi_j^x(r;\vartheta)\right]^2}{\sum_{k=1}^{6}\Sigma_{kk}(\vartheta)\sum_{r=0}^{h}\left[\Psi_k^x(r;\vartheta)\right]^2}. \tag{A.40}$$

The shares sum to 100 when forecast-error variance is positive. Because capital is predetermined, its impact forecast-error variance is zero and its shares at $h = 0$ are undefined. The figure therefore leaves this bar empty.

Figure 10 in Section 8.2 evaluates Equation (A.40) at the posterior mode. The bars distinguish TFPQ, demand, the labor and materials wedges reported together, the price wedge, and the investment wedge. In the baseline, the labor-materials segment sums two independent contributions. If their innovations are allowed to be correlated, the implementation retains their covariance and reports their joint contribution as one segment.

References

Scott A. Brave, R. Andrew Butters, and David Kelley. A new "big data" index of U.S. economic activity. *Federal Reserve Bank of Chicago Economic Perspectives*, 43(1):1–30, 2019. doi: 10.21033/ep-2019-1.

In Choi. Unit root tests for panel data. *Journal of International Money and Finance*, 20(2):249–272, 2001.

Jason Choi, Taeyoung Doh, Andrew Foerster, and Zinnia Martinez. Monetary policy stance is tighter than federal funds rate. Economic Letter 2022-30, Federal Reserve Bank of San Francisco, 2022.

David A. Dickey and Wayne A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431, 1979.

European Commission. Quality report on PRODCOM 2014 annual data. PRODCOM Working Group, 2014.

John G. Fernald. A quarterly, utilization-adjusted series on total factor productivity. Working Paper 2012-19, Federal Reserve Bank of San Francisco, 2014.

Luca Gagliardone, Mark Gertler, Simone Lenzu, and Joris Tielens. Anatomy of the phillips curve: micro evidence and macro implications. *American Economic Review*, 115(11):3941–3974, 2025.

Amit Gandhi, Salvador Navarro, and David A Rivers. On the identification of gross output production functions. *Journal of Political Economy*, 128(8):2973–3016, 2020.

Simon Gilchrist and Egon Zakrajšek. Credit spreads and business cycle fluctuations. *American Economic Review*, 102(4):1692–1720, 2012. doi: 10.1257/aer.102.4.1692.

Kyung So Im, M. Hashem Pesaran, and Yongcheol Shin. Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115(1):53–74, 2003.

Søren Johansen. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica: journal of the Econometric Society*, pages 1551–1580, 1991.

Rolf Larsson, Johan Lyhagen, and Mickael Löthgren. Likelihood-based cointegration tests in heterogeneous panels. *The Econometrics Journal*, 4(1):109–142, 2001.

G. S. Maddala and Shaowen Wu. A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics*, 61(S1):631–652, 1999.

Stephen Nickell. Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426, 1981. doi: 10.2307/1911408.

Peter Pedroni. Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, 61(S1):653–670, 1999.

Peter Pedroni. Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econometric theory*, 20(3):597–625, 2004.

M. Hashem Pesaran. A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22(2):265–312, 2007.

M. Hashem Pesaran and Yongcheol Shin. Cointegration and speed of convergence to equilibrium. *Journal of Econometrics*, 71(1-2):117–143, 1996.