Institute for International Economic Studies
Monograph Series No. 139

Stockholm
University

Frictions and Flexibility in Production

Essays in Energy Economics

Thore Petersen

Doctoral Thesis in Economics at Stockholm University, Sweden 2026

Frictions and Flexibility in Production
Essays in Energy Economics

Thore Petersen

Academic dissertation for the Degree of Doctor of Philosophy in Economics at Stockholm University to be publicly defended on Tuesday 9 June 2026 at 13.00 in Hörsal 3, Frescati Södra Huset B, Universitetsvägen 10B.

Abstract

Can the plants turn green?

What is the potential for substitution from fossil fuels to electricity? I answer this question with microdata from the German manufacturing sector, where fossil fuels account for 70% of primary energy consumption. I document large heterogeneity in the shares of fossil fuels and electricity across plants even within narrow categories of plants. This variation is difficult to explain with observable plant characteristics including location, industry, or products produced, which suggests plants have flexibility in the mix of energy sources. Fossil fuels and electricity respond differently to transitory plant-level demand shocks: For a given change in output, the response of electricity is three times larger than that of fossil fuels. To reproduce this finding, I develop a dynamic model of production with an adjustment cost for fossil fuels. In such a model it is not optimal for plants to fully adjust to transitory shocks, leading to a downward bias in estimates of the elasticity of substitution with canonical methods. I estimate the model using the simulated method of moments, and find an elasticity of substitution of 5, substantially higher than the literature. This implies that a tax on fossil fuels is more effective: A given reduction in fossil fuel use can be achieved at half the cost in foregone output compared to a model with an elasticity from the literature. German plants can, thus, turn green.

Long-run elasticities from short-run variation

Elasticity estimates are central to applied economics. A robust empirical finding is that long-run estimates exceed short-run ones. Under adjustment costs and stationary prices, agents optimally respond only partially to price shocks, attenuating short-run elasticity estimates below the structural long-run parameter. I derive a closed-form correction factor linking the two. It depends on three parameters: a calibrated discount factor, the persistence of the relative price, and the persistence of the input mix, the latter two estimable without additional data requirements. The persistence of the choice variable is a sufficient statistic for the adjustment friction, obviating structural estimation. I apply the method to intermediate input substitution in Indian manufacturing, and estimate a correction factor of 2.

Financial frictions and aggregate risk exposure

I study the welfare effect of preemptive industrial policy against aggregate supply chain risk when entrepreneurs choose technologies subject to a collateral constraint and the government lacks commitment. Laissez-faire is constrained efficient: without intervention, technology choices are socially optimal given the collateral constraint. When the government cannot commit to a redistribution rule before agents choose, agents anticipate the planner's response and distort their technology choices, generating welfare losses that offset the direct benefit of redistribution. A welfare decomposition confirms that this distortion cost dominates the redistribution gain in a majority of empirically relevant parameterizations, so no-commitment intervention reduces welfare relative to laissez-faire. This reverses only for the case of an almost complete supply cutoff.

Keywords: *Energy Economics, Energy Transition, Substitution, Measurement.*

Stockholm 2026
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-254277

ISBN 978-91-8107-632-5
ISBN 978-91-8107-633-2
ISSN 0346-6892

Stockholm University

Department of Economics

Stockholm University, 106 91 Stockholm

FRICTIONS AND FLEXIBILITY IN PRODUCTION

Thore Petersen

Stockholm University

Frictions and Flexibility in Production

Essays in Energy Economics

Thore Petersen

ISBN print 978-91-8107-632-5
ISBN PDF 978-91-8107-633-2
ISSN 0346-6892

Cover image:
The Fighting Temeraire (1839), J. M. W. Turner, National Gallery, London.

Back image:
Paolo Capelli.

Printed in Sweden by Universitetsservice US-AB, Stockholm 2026

FRICTIONS AND FLEXIBILITY IN PRODUCTION: ESSAYS IN ENERGY ECONOMICS

Thore Petersen

Doctoral Dissertation
Institute for International Economic Studies
Stockholm University

Abstracts

Can the plants turn green?

What is the potential for substitution from fossil fuels to electricity? I answer this question with microdata from the German manufacturing sector, where fossil fuels account for 70% of primary energy consumption. I document large heterogeneity in the shares of fossil fuels and electricity across plants even within narrow categories of plants. This variation is difficult to explain with observable plant characteristics including location, industry, or products produced, which suggests plants have flexibility in the mix of energy sources. Fossil fuels and electricity respond differently to transitory plant-level demand shocks: For a given change in output, the response of electricity is three times larger than that of fossil fuels. To reproduce this finding, I develop a dynamic model of production with an adjustment cost for fossil fuels. In such a model it is not optimal for plants to fully adjust to transitory shocks, leading to a downward bias in estimates of the elasticity of substitution with canonical methods. I estimate the model using the simulated method of moments, and find an elasticity of substitution of 5, substantially higher than the literature. This implies that a tax on fossil fuels is more effective: A given reduction in fossil fuel use can be achieved at half the cost in foregone output compared to a model with an elasticity from the literature. German plants can, thus, turn green.

Long-run elasticities from short-run variation

Elasticity estimates are central to applied economics. A robust empirical finding is that long-run estimates exceed short-run ones. Under adjustment costs and stationary prices, agents optimally respond only partially to price shocks, attenuating short-run elasticity estimates below the structural long-run parameter. I derive a closed-form correction factor linking the two. It depends on three parameters: a calibrated discount factor, the persistence of the relative price, and the persistence of the input mix, the latter two estimable without additional data requirements. The persistence of the choice variable is a sufficient statistic for the adjustment friction, obviating structural estimation. I apply the method

to intermediate input substitution in Indian manufacturing, and estimate a correction factor of 2.

Financial frictions and aggregate risk exposure

I study the welfare effect of preemptive industrial policy against aggregate supply chain risk when entrepreneurs choose technologies subject to a collateral constraint and the government lacks commitment. Laissez-faire is constrained efficient: without intervention, technology choices are socially optimal given the collateral constraint. When the government cannot commit to a redistribution rule before agents choose, agents anticipate the planner's response and distort their technology choices, generating welfare losses that offset the direct benefit of redistribution. A welfare decomposition confirms that this distortion cost dominates the redistribution gain in a majority of empirically relevant parameterizations, so no-commitment intervention reduces welfare relative to laissez-faire. This reverses only for the case of an almost complete supply cutoff.

Acknowledgments

I remember the day I arrived to Stockholm in August 2020. That sunny afternoon I ran along Brunsviken, and felt content and excited to be able to call this place home for the next few years. Little did I know what I would be in for, but I enjoyed many more sunsets over Brunsviken.

My cohort and I were immediately immersed in the Ph.D. program. The offices were mostly empty due to the Covid-19 pandemic, so we made the seventh floor our own. Quickly (and not entirely voluntarily), I became one of the poster children of the Stockholm Economics Ph.D. program through Ann-Sofie and Per's social media campaigns.[1]

A major step was joining the IIES in my third year. I should have known that it was not a coincidence that SJ introduced the EuroNight 345/346 connection between Stockholm and Hamburg right as I joined. That train would (less or more reliably) shuttle me between Stockholm and the Research Data Center in Hamburg many, many times, for me to figure out how German manufacturing plants use energy.

Thankfully, it is not the night train that defines the Ph.D. experience, but the people along the way.[2]

I want to thank my supervisor Per Krusell for his delicious fresh mackerel with butter and dill, and his lived aversion to pompousness (best reflected in his footwear). Intellectually, his comments were sharp, deep, and sometimes frustratingly prescient. I have spent hours trying

[1] If you are interested in joining an economics Ph.D. program, watch youtube.com/watch?v=doyScIk6KBE to learn what Stockholm has to offer!

[2] Although the patience learned on the platform proved useful for the job market.

to disprove Per's intuition, only to realize that he was right all along (I hear this is a rite of passage for his students though).

I am grateful for Joshua Weiss's open door and generosity with his time. Discussions with him were always insightful: he was probing any argument or model from angles I didn't even know existed. He was typically two steps ahead, despite my several hours of head start with the problem at hand. If only he were to accept that coffee must be hot.

John Hassler's expertise on everything climate and energy economics helped sharpen my research questions and approaches.

The IIES Macro Group additionally consisted of Timo Boppart, Brian Higgins, the late Paul Klein, Kieran Larkin, Kurt Mitman, and Yimei Zou, who all patiently listened to my presentations throughout the years and provided helpful comments.

Mitch Downey listened, commented, and improved my papers and presentations. He has the rare ability to present serious criticism in a friendly and constructive way that leaves the recipient motivated to improve, rather than demoralized. His diverse and creative modes of commuting reflect his inspiring attitude towards life in general.

Ann-Sofie Kolm has always been present, even before I started the program, always caring and supportive.

I want to thank Christina Lönnblad for her donation of Enhörnis, my office's stuffed spirit animal, and of course for her efficient administration of the IIES. Thank you also to Tobias Kjellberg and Alexander Johansson.

I remember arriving in Stockholm and getting to know my cohort. Among this smart and only slightly nerdy group of people, I immediately felt at home. I want to thank Huen Tat Au-Leung, Chek Yin Choi, Ida Haavi, Jinci Liu, Alexandre Mendonça, Alex Näsström, and Johan Orrenius for being great company. Also Merve Demirel for enforcing her "one-economics-joke-per-day" rule, to prevent us drifting too deeply into our bubble; Jonatan Riberth for being a (notoriously difficult to acquire) Swedish friend; Tiago Bernardino for his unwavering smile and academic pragmatism; Juan Llavador Peralt for short walks, deep insights, and sometimes shared frustration; and Jens Oehlen for long walks, surprising insights, and the gossip of course.

Over the years, I have shared my office, and thus discussions ranging from the mundane to the profound (typically the former), with Martina Dosser (great hints on coloring Easter eggs), Agneta Berge, Fredrik Paues, Jiaqi Luo, and Carolina Lindholm. All of you seemed to prefer to work from home,[3] but I am thankful for the time we did spend together in the office.

Beyond my cohort and office, I want to thank Jakob Beuschlein for looking at my graphs and deriving solutions to my econometrics questions on my whiteboard; as well as Anton Arbman Hansing, Alireza Azampour, Monir Bounadi, Chloé Castagnet, Simon Handreke, Stefan Hinkelmann, Philipp Hochmuth, Maija Kaartinen, Isabella Maaßen, Filip Mellgren, Thomas Mikaelsen, Artschil Okropiridse, Jonas Overhage, Stephan Schneider, and Leiv Storesletten for comments, conversations, discussions, and company.

I want to thank Maximilian Konradt for being one of the reasons I started the Ph.D. in the first place, and for pointing me towards the German data that became the foundation of Chapter 1.

Finally, I want to thank my family Elke, Wilhelm, and Ann-Kathrin Petersen for their unconditional support, and visits throughout. But most of all, my partner-in-Ph.D. and life, Carolin Seiferth. You knew what to say when I was struggling, and were amazing company when I was not. You were my reminder that there is more.

Thank you.

Thore Petersen
Stockholm, Sweden
April 2026

[3] Wait a minute. . .

Contents

Introduction

Substitution between inputs is core to production theory. John Hicks (in *The Theory of Wages*, 1932) and Joan Robinson (in *The Economics of Imperfect Competition*, 1933) introduced the *elasticity of substitution* to formally describe the degree to which one input can take the place of another in the production of a given output. The concept is a property of technology: it tells us how an economy or firm will respond to changes in the relative cost of its inputs. Nearly a century later, the same parameter takes center stage in one of the major challenges of our time: the transition to a low-carbon economy. The success of a transition away from fossil fuels depends on whether electricity from renewable sources can substitute for them in production.

Substitution is not instantaneous. Equipment is long-lived, supply chains are established, technology choices are sunk. There is a gap between the response observable in the short run and the response available in the long run, and its size matters. How we measure the elasticity depends on the time horizon of the data at hand. And how a transition unfolds—its cost and speed—depends not only on the long-run value but on how quickly the economy can move toward it. When short-run substitution is genuinely difficult, the question shifts from how much the economy can reorganize to what policy can do while it cannot.

This thesis comprises three self-contained chapters, each taking up one aspect of this set of questions. Chapter 1 asks how large the substitution potential is between fossil fuels and electricity in German manufacturing. Chapter 2 asks whether long-run elasticities can be measured using short-run variation when adjustment is costly. Chapter 3 asks whether, when substitution in response to a shock is infeasible, preemptive policy can compensate for it.

In the first chapter, *Can the plants turn green?*, I study the electrification potential of the German manufacturing sector. The central finding is a within-plant elasticity of substitution between fossil fuels and electricity of 5, substantially larger than existing micro-level estimates of around 1.5.

Using plant-level microdata from the German manufacturing census, I show that fossil fuels and electricity respond differently to changes in production output. Instrumenting changes in output with plant-level demand shocks, electricity use adjusts almost three times as strongly as fossil fuel use, and fossil fuel use exhibits higher persistence over time. This asymmetry is inconsistent with the standard assumption of a static energy input choice, and is consistent with the presence of adjustment frictions specific to fossil fuel use.

I develop a structural production function model with adjustment costs for fossil fuel use to study the implications of this finding. In such a model, canonical estimators of the elasticity of substitution are downward biased: for local shocks and short time horizons, a low elasticity and an adjustment cost are observationally similar. The structural model imposes additional moment restrictions on the time-series behavior of inputs, which allows separate identification of the elasticity and the friction. Structural estimation yields the within-plant elasticity of 5.

I illustrate the implications for climate policy using Germany's sectoral emission targets. A 40% reduction in fossil fuel use can be achieved with an 11% tax on fossil fuel prices, causing a 7% contraction in output. Under the lower elasticity from existing estimates, the same reduction requires a 23% tax and causes a 14% output contraction.

Finally, I document large cross-sectional heterogeneity in the energy mix across plants, even within narrowly defined products. This heterogeneity implies an additional extensive margin of adjustment through reallocation across plants, augmenting the within-plant intensive margin.

The within-plant elasticity of 5 is thus a lower bound on the aggregate potential for substitution.

The finding that adjustment frictions in fossil fuel use lead to underestimation of the elasticity by canonical methods raises a general methodological question. Adjustment frictions are not unique to energy: the empirical literature documents elasticity estimates that increase with the time horizon of identification in many settings, consistent with the presence of adjustment frictions. Short-run estimates obtained with canonical methods may therefore systematically understate the corresponding long-run elasticities across many applications.

In the second chapter, *Long-run elasticities from short-run variation*, I build on that point and develop a linear estimation framework to infer the long-run elasticity of substitution between inputs from short-run variation in input use, when the input choice is subject to adjustment costs.

A common finding in the empirical literature is that elasticity estimates increase with the time horizon of the price change, there is a distinction between a short-run and a long-run elasticity. Both values are relevant for policy: the short-run elasticity governs the speed of the transition, while the long-run value describes its endpoint. The short-run elasticity is easier to estimate, since plausibly exogenous transitory price variation is more readily available than exogenous and persistent variation.

I show that in certain cases the short-run dynamics of input use are still informative about the long-run elasticity, and that the relationship between the two can be expressed in a closed-form correction factor. To derive that result, I model the choice of the ratio of inputs subject to a quadratic adjustment cost. Under a second-order approximation of a CES production function, the short-run elasticity is the product of an attenuation factor and the long-run elasticity. The key insight is that the persistence of the input ratio is a sufficient statistic for the adjustment friction, enabling separate identification of the elasticity and the friction. The attenuation factor has a closed-form expression in three parameters:

the persistence of the input ratio, the persistence of the relative price, and the discount factor. The first two can be estimated from the data; the third can be calibrated to standard values. In the presence of adjustment costs, the long-run elasticity is weakly larger than the short-run elasticity.

I show that the canonical within estimator in panels targets the short-run elasticity. The canonical long-run estimator, the between estimator, is generally inconsistent unless the level of prices is strictly exogenous across units. The correction approach is complementary: it targets the same long-run parameter under a different set of assumptions, relying on exogeneity in changes rather than levels of the price, but imposing more structure on agents' behavior.

I apply the method to intermediate input substitution in Indian manufacturing, using data from Peter and Ruane (2025). They estimate a short-run elasticity of around 0.5 and a long-run elasticity of around 2.5 over 7 years, using the Indian trade liberalization as a natural experiment. My correction factor implies a long-run elasticity 2 times the short-run estimate, smaller than their directly estimated ratio of around 5. Possible sources of the discrepancy include different identification assumptions, model structure, and the time horizon over which the long run is defined.

The first two chapters study how adjustment frictions attenuate the economy's short-run response to shocks, and what this implies for measurement and policy. They ask how large the response to a shock is, and find that it is larger than canonical estimates suggest. A complementary question arises when the response is severely limited: can policy intervention compensate? If production is hard to reallocate after a shock, a government may try to shape its composition before adverse events occur, rather than inducing reallocation after the fact. But such preemptive intervention introduces a new challenge: rational agents anticipate the government's response to their choices, potentially undoing the intended benefits.

The third chapter, *Financial frictions and aggregate risk exposure*, turns to a setting with a different kind of friction: a financial constraint. This constraint limits the reallocation of production across technologies after a shock, and the chapter asks whether preemptive policy intervention can improve welfare.

Entrepreneurs choose between a safe technology with stable productivity and a risky technology with higher expected productivity but exposure to adverse shocks. A collateral constraint limits firms' ability to borrow capital, capturing the limited scope for reallocation after a shock. A planner redistributes assets across technology types before production occurs, relaxing borrowing constraints for some firms at the cost of tightening them for others. Crucially, the planner lacks commitment: it chooses redistribution after observing the composition of technology choices, and agents anticipate this when making their decisions.

The laissez-faire allocation is constrained efficient. Comparing no-commitment intervention to laissez-faire, the welfare effect decomposes into two opposing forces: a redistribution gain from optimally reallocating assets holding technology composition fixed, and a distortion cost from the shift in technology composition induced by anticipated redistribution. The distortion arises because agents respond to anticipated redistribution through two channels: higher expected profits from relaxed constraints, and a direct wealth transfer. The planner internalizes only the first when choosing redistribution, creating a wedge that causes agents to overshoot toward the favored technology.

The distortion cost exceeds the redistribution gain for the large majority of parameterizations. Only when the risky technology produces almost nothing in the bad state does no-commitment intervention marginally improve welfare. This is an empirically extreme scenario corresponding to near-complete supply cutoff. In the empirically more relevant case of major price spikes but no complete cutoffs, the distortion cost dominates, and preemptive intervention reduces welfare, relative to a planner not intervening at all.

Chapters 1 and 2 show that canonical methods can miss the elasticity when frictions are present, and that the difference runs in a specific direction: short-run variation understates long-run flexibility. Correcting for this changes both the number and the policy conclusion that follows from it. Chapter 3 takes the limit case—a friction so severe that post-shock reallocation is effectively impossible—and asks whether pre-shock policy can compensate. The answer depends on commitment: without it, anticipation of the policy distorts the very choices the policy was meant to shape.

Frictions are not nuisances to be averaged out. Their structure—adjustment costs, persistence, financial constraints—determines what we can measure and what a government can credibly do. Taking the friction seriously, rather than absorbing it into an aggregate response, is what changes the policy answer in each chapter.

Chapter 1

Can the plants turn green?

I thank my advisors Per Krusell and Joshua Weiss for a steady stream of critical questions and input. I am also grateful for comments and feedback from John Hassler, Timo Boppart, Torsten Persson, Kieran Larkin, Brian Higgins, Yimei Zou, Kurt Mitman, Mitch Downey, Amalia Repele, Veronica Salazar Restrepo, and other participants of the IIES Macro Group, and the IIES Brown Bag. During my work with the German manufacturing census, Steffi Dierks, Denise Henker, Thies-Hinnerk Krasmann, Frederik Liedtke, Vanessa Marquard, Michael Rößner, Janne Timmermann, Diane Zabel, and Benedikt Zapf of the Research Data Centres were very helpful and forthcoming. I am grateful for financial support from *Jan Wallanders och Tom Hedelius stiftelse samt Tore Browaldhs stiftelse*, *Stiftelsen Elisabeth och Herman Rhodins minne*, and *Stiftelsen Carl Mannerfelts fond*.

1.1 Introduction

Fossil fuels are both pervasive and problematic. They fuel the economy, but releasing their energy necessarily also releases carbon dioxide (CO_2), the main driver of climate change. Electricity from renewables is an alternative source of energy with no immediate CO_2 emissions. Can the economy meet its energy demands with such non-emitting sources?

Fuel requirements are typically embodied, meaning a particular piece of equipment requires a particular fuel. Gas turbines, electric motors, and petrol engines have their names for a reason. But the same process can often be performed by different equipment requiring another fuel: All these three pieces of equipment generate mechanical energy, or movement, but from different fuels. The question about the transition away from fossil fuels is thus a question about the development and adoption of technologies.

I study the use of fossil fuels and electricity in the census of German manufacturing plants. The manufacturing sector accounts for almost 30% of primary energy consumption in Germany, and fossil fuels comprise more than 70% of its consumption, see figures 1.1 and 1.2. I show that there is substantial heterogeneity in this energy mix across plants: Among the subset of plants that produce a single product (6-digit resolution), there is four times more variation across plants than within plant over time. In the full sample, among plants in the same 4-digit industry, there is almost five times more variation in the cross-section than within plant over time. This variation can not be explained by observable characteristics like location, year of entry, or the scope of production (proxied as the share of intermediate inputs in value-added). Plant size can explain some of the variation, larger plants use relatively less fossil fuels. A plant at the 80th percentile of the distribution of revenue has a 5.5 percentage points (p.p.) lower fossil fuel share in the energy mix compared to plant at the 20th percentile on average.

Finally, I show that the factor demands for fossil fuels and electricity respond differently to demand shocks. I instrument changes in physical

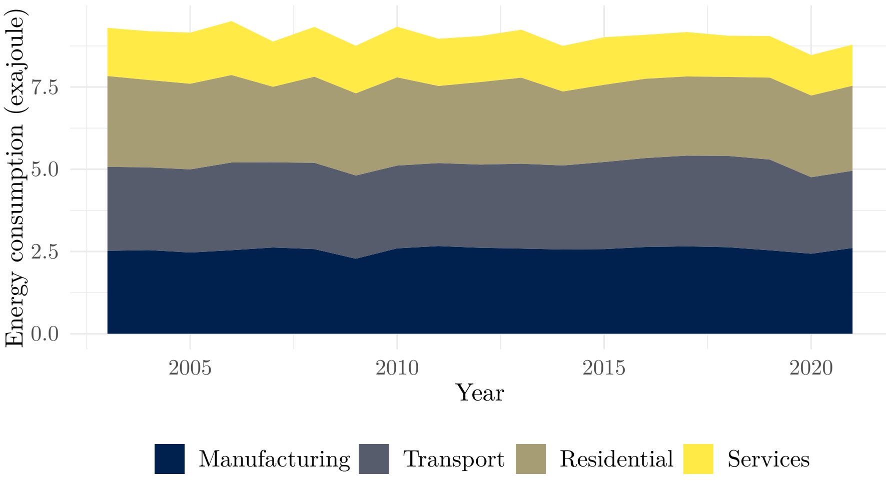


Figure 1.1: Primary energy consumption over time by sector in Germany. Own calculations based on Arbeitsgemeinschaft Energiebilanzen e.V. (2025).

output with plant-level demand shocks, and estimate the elasticity of energy use. Fossil fuel use is inelastic. For a 1% change in output, fossil fuel use changes by only 0.16% in the first year, and 0.24% in the second. Electricity use is more elastic. For a 1% change in output, electricity use changes by 0.62% and 0.64% in the first and second year respectively. Electricity use thus responds almost three times as strongly as fossil fuel use. A homothetic production function, as commonly assumed in the literature, would imply equal elasticities, which I can reject.

These results are novel and inconsistent with the common assumption of the choice of energy inputs as static and without dynamic considerations. To study the implications of this finding, I develop a dynamic model of heterogeneous plants that can reproduce these empirical findings. Plants differ in their productivity and a clean energy share parameter, both of which are permanent types. They produce a homogeneous output good by combining clean and dirty energy in a constant elasticity of substitution production function with decreasing returns to scale. I add one critical ingredient: a dynamic adjustment cost for dirty energy use.

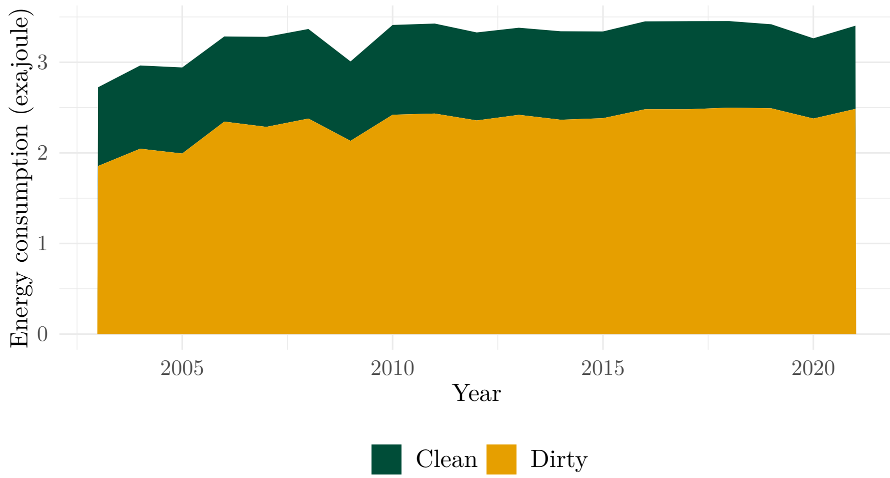


Figure 1.2: Primary energy consumption over time by clean and dirty sources within the manufacturing sector in Germany. Clean energy sources include electricity and renewables, dirty energy are fossil fuels. Own calculations based on microdata provided by Research Data Centre of the Statistical Offices of the Federal States (2023).

This is a reduced-form representation of the fact that fossil fuel equipment is subject to technical constraints, like so-called ramp times, that make adjustments costly. I assume that clean energy use can be adjusted freely. There is one source of exogenous variation in the model: an idiosyncratic stochastic demand process. Plants choose their inputs to maximize the net present value of profits. The distribution of plants over types is determined by a selection mechanism at entry. A potential entrant draws a productivity and clean share parameter from independent distributions, and must pay a fixed cost to enter. The fixed cost increases with the clean share parameter, to match the empirical correlation. There are two margins of substitution between clean and dirty energy in the model: A within-plant intensive margin, where a given plant substitutes between clean and dirty energy, and a between-plant extensive margin, where plants with different clean shares are selected at entry.

I estimate the model and can quantitatively reproduce the empirical findings. The elasticity of substitution between clean and dirty energy at the micro-level is estimated as 5.1, substantially larger than other micro-level estimates in the literature.[1] I show that in the presence of the adjustment costs, it is optimal for plants not to react fully to changes the relative price of energy or demand shocks, which introduces a downward biased estimate of the "deep" elasticity of substitution with the canonical approach. Due to the lack of data on firm- or plant-level prices of energy in my setting, I can not directly compare the estimates. The adjustment cost payments are small, representing less than 0.5% of a plant's total costs on average. Yet, they are sufficient to generate the observed difference in elasticity estimates.

I show the relevance of the findings for climate policy by conducting two policy experiments in the estimated model: an entry subsidy for clean plants, and a tax on dirty energy. The entry subsidy lowers the net fixed cost of entry for plants with a higher clean share parameter draw. It is not effective at increasing the aggregate clean share of energy use. The policy acts only at the entry margin by construction, but aggregate energy consumption is primarily driven by very productive firms, far above the entry cutoff. Spending one entire period's output on the entry subsidy increases the aggregate clean energy share by less than 0.1%. The policy is expansionary, meaning that it even increases total use of both clean and dirty energy.

A tax on dirty energy is effective at reducing dirty energy use. A 40% reduction in aggregate dirty energy use (corresponding to Germany's sectoral target for 2030, relative to 2018) can be achieved with a permanent 11% tax on dirty energy prices. The policy strictly increases marginal costs, and thus leads to a contraction in aggregate output of 7%, while generating revenue of around 1.5% of output. I show that the effectiveness of the tax critically depends on the value of the elasticity of substitution between clean and dirty energy. In a counterfactual calibration in line

[1]The average micro-level elasticity over industries in Jo (2024) is around 1.5. The aggregate elasticity in Papageorgiou et al. (2017) is around 3. Stern (2012) estimates a value around 1.

with existing micro-level estimates at a value of 1.5, the tax must be 23% to achieve the same 40% reduction in dirty energy use, leading to a contraction in output of 14%. The results suggest that the plants can turn green to a larger extent than previous estimates suggest, lowering the cost of climate change mitigation.

To map from the empirical analysis to the model, I classify energy as "clean" and "dirty". I consider all electricity consumed by a plant as clean, irrespective of how it was generated. While most electricity was generated from fossil sources during my sample, electrification is a necessary (but not sufficient) condition for a production process to run without emissions. If a plant electrifies, it can in principle be powered by renewables. I classify all fossil fuels as dirty.

Related Literature This paper speaks to the literature on climate macroeconomics, in particular the literature on substitution between "clean" and "dirty" inputs. Seminal papers in this literature (Acemoglu et al., 2012, 2016; Golosov et al., 2014) both stress the importance of the substitutability between clean and dirty inputs, and lament the lack of empirical estimates. The value of the elasticity of substitution determines the path and level of optimal policy, and the transition path of the economy. The transition to a clean economy is faster and cheaper, the higher the elasticity of substitution. Casey (2024) shows that the distinction between a short- and long-run elasticity is relevant for total emissions along the transition path. I contribute to the literature by providing estimates of the elasticity of substitution between clean and dirty inputs in a case for which this distinction is relatively clear: energy. By documenting the large heterogeneity in the energy mix across plants even within product, I show that there is substantial latent potential for aggregate substitution, even through mere technology adoption, rather than innovation. The within-plant elasticity can be interpreted as the short-run elasticity, while the long-run elasticity is augmented both by a demand reallocation channel between plants (Oberfield & Raval, 2021), and the selection at entry.

The paper is closely related to the literature on input substitution. A timeless focus is on the elasticity of substitution between capital and labor (Antràs, 2004; Chirinko, 2008). Diamond et al. (1978) provide an important non-identification result: in the presence of factor-augmenting technical change, the elasticity of substitution is not identified from timeseries data. A large literature devises identification strategies or structural models to estimate the substitutability between intermediate inputs (Barrot & Sauvagnat, 2016; Boehm et al., 2023; Peter & Ruane, 2025). The findings are typically a low short-run elasticity of substitution, that increases with the time horizon. A much smaller literature employs similar methods to estimate the elasticity of substitution between types of energy: Jo (2024) derives an estimation equation from a CES specification, and instruments for changes in the relative price of clean and dirty energy. Papageorgiou et al. (2017) non-linearly estimate a CES production function on sector-level data. Stern (2012) conducts a meta-analysis of estimates at different levels of aggregation. These papers report estimates of the elasticity of substitution of around 3 from cross-sectional variation, and around 1 from time-series variation. The different values are interpreted as a short- and long-run elasticity. Kaartinen and Prane (2024) and Leclair (2025) develop structural models and calibrate them to microdata, to study the substitution between a large set of fuels in production. Another approach is the estimation of translog cost functions (Arnberg & Bjørner, 2007; Bousquet & Ladoux, 2006; Hyland & Haller, 2018). The elasticity of substitution is implied by the estimated cross-price elasticities, but it is not typically calculated. I contribute to the literature by providing a structural estimate, and documenting possibly a large potential for substitution, given the large heterogeneity.

Lastly, there is a growing literature analyzing energy use in manufacturing microdata. Barrows and Ollivier (2018) show that across-firm composition effects drive energy intensity in Indian manufacturing. Rottner and von Graevenitz (2024) show that aggregate emission intensity in German manufacturing decreased due to compositional effects, while the intensity increased at the product level. Marin and Vona (2021) find negative employment effects in response to energy price increases in the French manufacturing sector. Hawkins-Pierot and Wagner (2025) and Linn (2008) document technology lock-in effects for energy intensity. A recent strand of the literature estimates causal effects of environmental policy. A focus is on the response of CO_2 emissions to carbon taxation, or emission trading systems (Andersson, 2019; Colmer et al., 2024; Dechezleprêtre et al., 2023; Gerster & Lamp, 2024; Martin et al., 2014; Martinsson et al., 2024; Shapiro & Walker, 2018). A robust finding is a reduction of the emissions of treated firms or plants, without adverse effects on economic activity. I contribute to the literature by studying energy use at the micro level, and add by considering the heterogeneity between plants.

How is energy used in manufacturing? Figure 1.3 breaks down energy consumption in German manufacturing by source and application. In 2023, 98.5% of fossil fuel use was for thermal applications, predominantly process heat. 67% of electricity use is for mechanical applications, mostly machine drives (for example pumps, compressors, drills, conveyors, or fans). 24% of electricity use is for thermal applications. The remaining 9% include lighting and IT systems. 76% of primary energy consumption is for thermal applications, and 21% for mechanical applications. The remainder includes lighting and IT.

Energy Primer Energy comes in different forms: thermal, mechanical, and electrical, among others. A common aggregate measure is primary energy consumption. It is defined as the heating potential or energy content of the fuels consumed, for example in units of joule (J) or watt-hours (Wh). When converting between different forms of energy there are losses. The useful energy content of a fuel is almost always lower than its heating potential. For example, a typical gas turbine that generates mechanical power from natural gas has an efficiency of around 40%: only 40% of the heating potential of the gas are converted

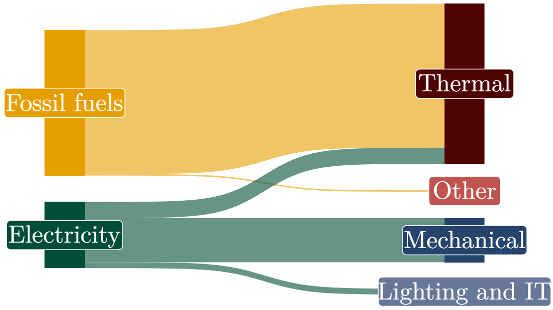


Figure 1.3: Primary energy consumption by source and application in German manufacturing in 2023. The width of a bar represents its share of total primary energy consumption. Fossil fuels are almost exclusively used in thermal applications, for process heat. Electricity is mostly used in mechanical applications, but also for heating and cooling. Own calculations for the year 2023 based on Fraunhofer ISI (2025).

into useful mechanical energy, the remainder is lost as waste heat. In an increasing number of applications that waste heat is recovered to increase overall efficiency to up to 80–90%. Converting electrical energy into mechanical energy is more efficient, with typical efficiencies for electric motors above 90%. Producing thermal energy from electricity using heat pumps can be done with implicit efficiencies even above 100%, since they concentrate and move heat instead of generating it. To substitute fossil fuels with electricity, only the useful energy content must be replaced, not the entire primary energy consumption. In this paper, I study primary energy consumption, which is reported in the data. With different processes, there is not a single conversion factor from primary to final energy consumption.

1.2 Data

1.2.1 Source and coverage

I use confidential microdata from the census of German manufacturing plants. The data is provided by the Research Data Center of the German statistical agency Destatis, and made available under the name AFiD (*Amtliche Firmendaten für Deutschland*, official firm data for Germany). The unit of observation is a plant (*Betrieb* in German), defined as a geographically bounded unit of production that belongs to a firm. Some variables are available only at the firm (*Unternehmen*) level, where a firm is the smallest independent legal entity required to keep accounts.

The data covers the universe of plants belonging to a firm with 20 or more employees in the manufacturing sector in Germany. I combine the modules on production, energy use, and employment, revenue, and investment (Research Data Centre of the Statistical Offices of the Federal States, 2023).[2] The data covers the years 1995 to 2021, except for the energy module, which starts in 2003.

Sample restrictions I restrict the sample to the manufacturing sector, corresponding to NACE Rev. 2 section C. I remove observations for which the following variables are within the top and bottom 2.5% of observations with strictly positive values by 2-digit industry: (i) dirty energy over clean energy, (ii) clean energy per worker, (iii) dirty energy per worker, and (iv) output per worker. I retain observations for which dirty energy use is zero, as some plants use only clean energy.

1.2.2 Variable construction

Energy use The energy module records the heating potential in kilowatt-hours (kWh) of fuel consumed at the plant for 10 fuel categories. The largest fuel categories by consumption are natural gas, electricity, and coal products, accounting for 30%, 25%, and 16% of total energy use on average over time, respectively.[3] I aggregate the energy use into clean and dirty: I define clean energy as the sum of electricity and renewables, and dirty energy as the sum of fossil fuels, as well as district heat. Some plants generate electricity from fossil fuels on-site, and the data does not

[2]All results in this paper are own calculations based on the AFiD data.

[3]The remaining fuel categories and shares are: other oil products (9%), other gas products (6%), district heat (5%), renewables (4%), waste (2%) and heating oil (2%).

distinguish between the use of fossil fuel in production and for electricity generation. For those plants, I estimate the kWh of dirty energy used in production by subtracting the kWh of electricity generated on-site from the total kWh of dirty energy consumed.[4]

Output The production module records both price and quantity for each product a plant produces in a given year. This level of detail allows me to construct a plant-level price index, to generate an accurate measure of real output. I construct a Törnqvist price index at the plant level (described in Eslava et al., 2004).[5] Due to the arbitrary basis, the level of Törnqvist-deflated output is not comparable across plants. The index does provide a more accurate measure of the changes in real output

[4]The relation between generated electricity, $E^{c\ \mathrm{gen}}$, and dirty energy use for generation, $E^{d\ \mathrm{gen}}$, is $E^{c\ \mathrm{gen}} = \eta^{\mathrm{elec}} E^{d\ \mathrm{gen}}$, where $\eta^{\mathrm{elec}} \in (0,1)$ is the electrical efficiency of generation. The remaining heating potential is converted into heat in the process, $E^{d\ \mathrm{heat}} = \eta^{\mathrm{heat}} E^{d\ \mathrm{gen}}$, with $\eta^{\mathrm{elec}} + \eta^{\mathrm{heat}} = 1$ by the first law of thermodynamics. Almost all plants with on-site generators employ combined heat and power generation (CHP), so use that heat in production. The adjusted dirty energy heating potential available for production is then

$$\begin{aligned} E^{d\ \mathrm{prod}} &= E^d - E^{d\ \mathrm{gen}} + E^{d\ \mathrm{heat}} = E^d - E^{d\ \mathrm{gen}} + \eta^{\mathrm{heat}} E^{d\ \mathrm{gen}} \\ &= E^d - \left(1 - \eta^{\mathrm{heat}}\right) E^{d\ \mathrm{gen}} = E^d - \left(1 - \eta^{\mathrm{heat}}\right) \frac{E^{c\ \mathrm{gen}}}{\eta^{\mathrm{elec}}} \\ &= E^d - E^{c\ \mathrm{gen}}. \end{aligned}$$

This assumes that the recovery rate of useful energy from $E^{d\ \mathrm{heat}}$ is equal to the conversion efficiency, which is the case on average. Engineering estimates for both are around 0.7–0.9, depending on exact application.

[5]The Törnqvist index is a chained price index. The change in the index P_t for a basket of goods G from period t to $t+1$ is

$$\begin{aligned} \Delta \log P_{t+1} &= \sum_{g \in G} \bar{s}_{gt+1} \Delta \log P_{gt+1}, \\ \bar{s}_{gt+1} &= \frac{s_{gt+1} + s_{gt}}{2}, \end{aligned}$$

where P_{gt} is good g's price, and s_{gt} is its quantity share of the basket in period t. The index price in period t is then

$$P_t = \exp\left(\sum_{\tau=1}^{t} \Delta \log P_\tau\right).$$

It is implicitly normalized to the first year in which the plant is active.

though, compared to industry-level deflators. For a measure of output that is comparable across plants, I deflate the nominal value of production and revenue, see below.

Deflators I deflate revenue and the nominal value of production to 2015 Euro using producer price indices (PPIs) at the 2-digit industry level from the Federal Statistical Office of Germany (Destatis). I deflate expenditure on intermediate inputs to 2015 Euro using the corresponding price series at the 2-digit industry level from the EU KLEMS database (Bontadini et al., 2023).[6]

Industry and product classification The industry and product classifications change several times during the sample period. I harmonize all classifications to the versions used in the latest years in the sample. For industry, the mapping between old and new schemes is not unique for most industries. I map all 4-digit industry identifiers to WZ2008 (equivalent to NACE Rev. 2) according to the following rule. Among plants that are active before and after the change in classification, I copy the new classification to the previous years. For plants that are active only under the old classification, I assign the most common transition from the plants that are active in both classifications. For products, I map the 9-digit product identifiers to the GP2019 classification scheme (corresponding to PRODCOM up to 8 digits, with Germany-specific details at the 9th digit). At this resolution, there is a unique mapping between the old and new classifications for most products. For ambiguous cases, I choose the first product listed in the official correspondence table.

Intermediate inputs The expenditure on intermediate inputs or materials is recorded only at the firm level, and not in all years for all firms. The variable is part of the cost structure survey (*Kostenstrukturerhebung*). This survey is conducted every year for firms with 500 or more employees.

[6]PPI: Destatis table number 61241-0003; intermediate input prices: EU KLEMS national accounts series `II_PI`.

Among smaller firms, a sample stratified by 4-digit industry and number of employees is drawn, such that a total of around 16,000 firms are included. Smaller firms are included every four years on average. Among multi-plant firms, I assign intermediate inputs expenditure according to the share of total nominal production value.[7]

Entry The data does not directly record the year of entry of a plant or firm. For plants that enter after the first year of the sample, 1995, I define their year of entry as the year in which they first appear. Given the inclusion criteria, this is the actual year of entry for plants that are part of a firm with 20 or more employees. For plants belonging to smaller firms, it measures the year in which the firm grows to at least 20 employees. There is a level difference between these sets of plants: plants of existing firms have on average fewer employees at entry. Both follow very similar trajectories over time.

Summary statistics Table 1.1 presents summary statistics of the sample for the main variables used in the analysis.

Variable	N	Mean	SD	Median	10th Pct.	90th Pct.
E^c	248590	4305614	20398274	717022	107978	8279847
E^d	248590	7737987	45403063	547884	101782	10426251
$p_y Y$	248590	30212108	88538350	9613197	2455404	66202875
$\log E^d/E^c$	248590	-0.09	1.20	0.06	-1.71	1.30
$E^d/(E^c+E^d)$	248590	0.49	0.23	0.52	0.15	0.79
E. exp./VA	186771	0.06	0.06	0.04	0.01	0.13

Table 1.1: Summary statistics for clean and dirty energy use (E^c, E^d in kWh), revenue $p_y Y$ (in 2015 Euro), the log of the ratio of dirty to clean energy, $\log E^d/E^c$, the dirty energy share, $E^d/(E^c+E^d)$, and total energy expenditure over value added (E. exp./VA, available only at the firm-level).

Energy prices The data does not record plant- or firm-level prices of energy. Prices for energy are determined on national markets in Germany.

[7] 94% of firms have one plant. 82% of plants belong to a single-plant firm. Both shares are stable over time.

There exists a single wholesale spot market for electricity, with some regional variation in price surcharges. Natural gas is not traded centrally, but prices are similar across the country (Bundesnetzagentur, 2021). There are level differences in the prices of electricity and natural gas by consumption band: larger consumers pay lower prices per kWh. The relative price is constant over the consumption bands though (own calculations based on Destatis, 2023). Figure 1.4 shows price indices for electricity, and the major fossil fuels used in manufacturing: natural gas, coal, and fuel oil.[8] All prices are increasing over time. The prices of fossil fuels are correlated, and are more volatile than the price of electricity, but no divergence in the trends is apparent. This is consistent with the stability of the aggregate energy mix.

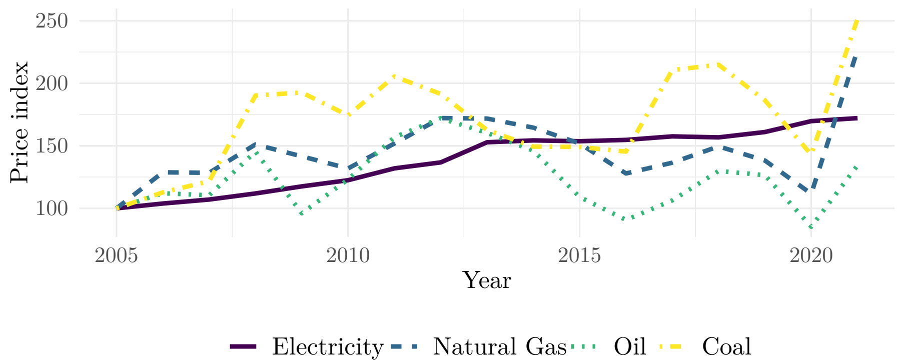


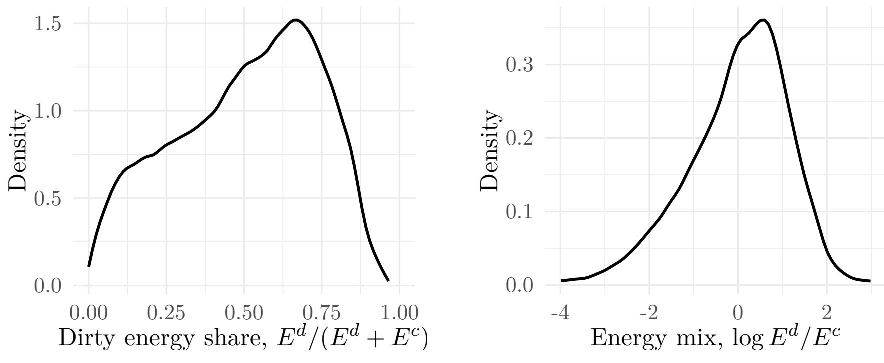
Figure 1.4: National price indices of the four main energy sources in German manufacturing. All indices are trending upwards over time. The prices of fossil fuels are correlated, and more volatile than the price of electricity, but no difference in trend is apparent. Data from Destatis (2023).

[8]The price indices are from Destatis (2023). The series are "Elektrischer Strom bei Abgabe an gewerbliche Anlagen" for electricity, "Erdgas, bei Abgabe an die Industrie" for natural gas, "Heizöl leicht" for fuel oil, and "Steinkohle" for coal.

1.3 Empirical Results

1.3.1 Variation in the energy mix

There is substantial variation in the energy mix. I calculate the share of dirty energy in total energy use, $E^d/(E^c + E^d)$ as a measure of the energy mix. This share has a mean of 0.49 and a standard deviation of 0.23. Another measure is the log ratio of dirty to clean energy use, $\log(E^d/E^c)$, which has a mean of -0.01 and a standard deviation of 1.2. I will use this log-ratio as the primary measure for its convenient numerical properties in most subsequent analysis.



(a) Distribution of dirty energy share. **(b)** Distribution of energy mix.

Figure 1.5: Kernel density estimates of (a) dirty energy share and (b) energy mix across manufacturing plants over the whole sample. 8% of plants use clean energy only, which are excluded from these density estimators.

What drives this variation? Observable characteristics of the plants can explain only a small fraction of it. I estimate the regression

$$\log \frac{E_{it}^d}{E_{it}^c} = \text{fixed effects} + \eta_{it}, \tag{1.1}$$

for several sets of fixed effects: year, industry at the 2- and 4-digit level, product at the 6-digit level, district (*Landkreis*)[9], and plant. The

[9] There are 401 *Landkreise* in Germany. The average number of plants per district is 95, the median 68.

results are reported in table 1.2. Table 1.B.1 additionally reports sample size and Bayesian Information Criterion (BIC) to demonstrate that the explanatory power of plant fixed effects is not due to overfitting.

Fixed Effects	adj. R^2
Industry (2-digit)	0.14
Industry (4-digit)	0.27
Industry (4-digit) by Year	0.28
Product (6-digit)	0.35
District	0.03
District + Industry (4-digit)	0.28
Plant	0.87

Table 1.2: Adjusted R^2 for regressions of energy mix measure $\log E^d_{it}/E^c_{it}$ on different sets of fixed effects. For the models with district effects and product effects, I restrict the sample to districts with at least 5 plants, and products produced by at least 5 plants.

In the full sample, the industry a plant belongs to explains a quarter of the variation in the energy mix. Among plants that produce only a single product, the product they produce explains a little over one third of the variation in the energy mix. The district a plant is located in explains very little by itself, and adds only marginally to the explanatory power of industry. The year also explains little variation. Plant fixed effects stand out: they explain 87% of the variation in the energy mix.

Under the assumption of an additive hierarchical model ($\log E^d_{it}/E^c_{it} = \text{fe}_i + \text{fe}_{\text{group}(i)} + \varepsilon_{it}$), I can calculate the ratio of the variances from the R^2 values. The ratio of the between-plant to within-plant variance can then be calculated as

$$\frac{R^2_{\text{plant}} - R^2_{\text{group}}}{1 - R^2_{\text{plant}}},$$

where R^2_{plant} is the R^2 from plant fixed effects, and R^2_{group} is the R^2 from a grouping level (e.g., industry). Using this formula, I find that the variation in the energy mix between plants relative to within plant over time is about 4 times larger among producers of the same product, and

Dependent Variables:	$\log E_{it}^d/E_{it}^c$	$\log E_{it}^d/E_{it}^c$
Model:	(1)	(2)
$\log Y_{it}$	-0.1603	-0.1019
	(0.0078)	(0.0063)
Fixed-effects		
Plant	Yes	No
Year	Yes	No
4-digit Industry by Year	No	Yes
Fit statistics		
Observations	248,590	248,590
R^2	0.92	0.36
within-R^2	0.013	0.014

Table 1.3: Conditional correlation between energy mix and plant size. Standard errors are clustered by plant and year.

4.6 times larger among plants in the same 4-digit industry.

1.3.2 Energy mix-size correlation

One observable characteristic that can explain variation in the energy mix is the size of a plant. Larger plants have a higher share of clean energy in their energy mix. This holds within plant over time, and in cross-section controlling for different sets of fixed effects.

I estimate the conditional correlation with the regression equation

$$\log \frac{E_{it}^d}{E_{it}^c} = \beta \log Y_{it} + \delta_{\text{fe}(i,t)} + \nu_{it}, \tag{1.2}$$

where $\delta_{\text{fe}(i,t)}$ are different sets of fixed effects, and Y_{it} is PPI-deflated revenue. The results are reported in table 1.3.

1.3.3 Differential dynamics of clean and dirty energy use

Estimation approach

To understand how clean and dirty energy enter production, I estimate their factor demand elasticities: how much does the use of each input change in response to changes in output? To estimate the factor demand elasticity, I regress the change in factor use on the change in output,

$$\Delta_k \log X_{it} = \beta_k \Delta_k \log Y_{it} + \gamma_{\iota(i,t)} + \xi_{r(i,t)} + \epsilon_{it}, \tag{1.3}$$

where $\Delta_k z_{it} = z_{it} - z_{it-k}$, and X_{it} are production inputs. I include fixed effects $\gamma_{\iota(i,t)}$ for 2-digit industry-by-year, and $\xi_{r(i,t)}$ for district-by-year. This equation suffers from several sources of bias: (i) simultaneity bias (Marschak & Andrews, 1944), as inputs and outputs are jointly chosen; (ii) attenuation bias, as productivity shocks introduce variation in Y_{it} without corresponding changes in X_{it}; (iii) omitted variable bias, as idiosyncratic input price shocks may change both factor demand and the optimal scale of production.

To circumvent these problems, I construct a demand shock instrument for the change in output. For each plant i and year t, I calculate the leave-one-out change in real output within the plant's 4-digit industry $\mathcal{I}_{it}$:

$$\text{shock}_{it}^{k} = \log \sum_{j \in \mathcal{I}_{it} \setminus \{i\}} Y_{jt} - \log \sum_{j \in \mathcal{I}_{it} \setminus \{i\}} Y_{jt-k}. \tag{1.4}$$

Exclusion restriction This instrument directly addresses the simultaneity and attenuation biases, as it is uncorrelated with plant-level shocks by construction. It addresses the omitted variable bias for the same reason: conditional on industry-by-year and state-by-year fixed effects, the instrument is uncorrelated with idiosyncratic input price shocks. The exclusion restriction could be violated for plants that have sufficiently high shares in either the input or output markets, such that their behavior influences prices. Concentration is low in the data in general, and I further address the issue by restricting the analysis to 4-digit industries with at least 50 plants each year.

Identifying variation The factor demand elasticity is then identified by variation at the 4-digit industry-by-year level, controlling for 2-digit industry-by-year and state-by-year fixed effects. The 2-digit industry-by-year effects control for aggregate shocks and structural trends, while the state-by-year effects absorb regional shocks.

Results

I estimate equation (1.3) using 2SLS for clean and dirty energy E^c, E^d, respectively, up to 2-year differences. The first stage results are strong, with a coefficient estimate of 0.22 in the first year, and 0.20 in the second, and Wald statistics of 146 and 145, respectively. Appendix table 1.B.2 reports the first stage results.

The factor demand elasticity estimates are reported in Table 1.4. Dirty energy is not very responsive to demand shocks in general. This contrasts with the response of clean energy, which is much more elastic.[10]

Heterogeneity and robustness The results for both the first and second stage are broadly symmetric for positive and negative demand shocks (appendix table 1.B.4). I estimate the regression separately by 2-digit industry, and find very similar results in all the industries that have sufficiently many plants for the first stage to be strong. The effects appear to be linear: when including a quadratic term in both stages, only the linear terms are statistically significant, and are similar in magnitude to the linear specification (appendix table 1.B.5). I repeat the exercise on a subsample of plants that produce only a single product at the 6-digit level. I construct the demand shock instrument analogously at

[10]For illustration, I also estimate the factor demand elasticities for labor and materials inputs. The results are presented in appendix table 1.B.3. The response of labor is between those of clean and dirty energy. The elasticity of materials is 1, they respond one-for-one to output changes.

Dependent Variables:	$\Delta_1 \log E^d$ (1)	$\Delta_1 \log E^c$ (2)	$\Delta_2 \log E^d$ (3)	$\Delta_2 \log E^c$ (4)
$\Delta_1 \log Y$	0.1641	0.6198		
	(0.0592)	(0.0321)		
$\Delta_2 \log Y$			0.2362	0.6411
			(0.0586)	(0.0357)
Fixed-effects				
Year by 2-digit Industry	Yes	Yes	Yes	Yes
Year by District	Yes	Yes	Yes	Yes
Fit statistics				
Observations	275,455	304,305	249,927	276,385
Wald (1st stage)	146.15	153.27	144.65	158.02

Table 1.4: 2SLS regression results for equation (1.3), factor demand elasticity for dirty and clean energy at 1-year and 2-year differences. Standard errors are clustered at the plant and 4-digit industry by year level. The dependent variable is the change in factor use, $\Delta_k \log X_{it}$, where k is 1 or 2 years. The key independent variable is the change in output, $\Delta_k \log Y_{it}$. This change in output is instrumented with the change in the leave-one-out output within a plant's 4-digit industry. Fixed effects for year by 2-digit industry and year by district are included.

the product-by-year level. The point estimates are very similar to those in the full sample, although less precise due to the smaller sample size.

1.4 Model

1.4.1 Environment and technology

I develop a dynamic partial equilibrium model of infinitely-lived plants. Plants are heterogeneous in their productivity and a clean energy share parameter, both of which are permanent types. They produce a homogeneous output good by combining clean and dirty energy. Dirty energy is subject to an adjustment cost: changing the level of dirty energy use between periods is costly. Clean energy is chosen freely. The prices of clean and dirty energy are exogenous and constant. The only source of variation to a plant are idiosyncratic demand shocks, modelled as a stochastic process for a plant's output price.

The distribution of plants is determined through an entry margin. Potential entrants draw productivity and clean share types from independent distributions. To enter, they must pay an entry cost that scales with their clean share type. A plant does enter if the expected present discounted value of profits exceeds the entry cost, given its draw of productivity and clean share type.

Plants combine clean and dirty energy, E^c and E^d, in a constant elasticity of substitution (CES) production function to produce energy services E,

$$E(E^c, E^d; b) = \left[b(E^c)^{\frac{\sigma-1}{\sigma}} + (1-b)(E^d)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{1.5}$$

where $b \in (0,1)$ is the plant's clean energy share parameter, and $\sigma > 0$ is the elasticity of substitution between clean and dirty energy.

The final good Y is then produced with a decreasing returns to scale (DRS) production function from energy services:

$$Y(E^c, E^d; A, b) = A^{1-\alpha} E(E^c, E^d; b)^{\alpha}, \tag{1.6}$$

with productivity $A > 0$ and returns to scale parameter $\alpha \in (0, 1)$.

Dirty energy is subject to a quadratic adjustment cost:

$$\phi(E^d, E^d_{-1}) = \frac{\phi_1}{2} \frac{1}{E^d_{-1}} \left(E^d - E^d_{-1}\right)^2, \tag{1.7}$$

where E^d_{-1} is the previous period's dirty energy use, and $\phi_1 > 0$ governs the scale of the adjustment cost.

The idiosyncratic price for the final good, p_y, follows an AR(1) process in logs,

$$\log p'_y = \rho_y \log p_y + \sigma_y \epsilon', \quad \epsilon' \sim \mathcal{N}(0, 1), \tag{1.8}$$

with persistence $\rho_y \in (-1, 1)$ and shock standard deviation $\sigma_y > 0$. The shocks ϵ' are i.i.d. over time and across plants, and drawn from a standard normal distribution.

I consider a partial equilibrium model for the following reasons. Germany imports the vast majority of its fossil fuels (except for coal), and electricity is traded on a large European market. Plants are price-takers, and I am abstracting from equilibrium effects on energy prices.

1.4.2 Equilibrium

I consider a steady state partial equilibrium, with one-time entry. The equilibrium is given by a distribution of plants over types (b, A), such that all entering plants maximize their expected present discounted value of profits given the exogenous prices of clean and dirty energy, and the entry condition is satisfied.

1.4.3 Plant problem

Plants maximize the expected present discounted value of profits, by choosing clean and dirty energy inputs. I am solving for their value function and policy function for dirty energy use, E^d.

Within period profits are

$$\pi(E^c, E^d, E^d_{-1}, p_y; b, A) = p_y Y(E^c, E^d; A, b) - p_c E^c - p_d E^d - \phi(E^d_{-1}, E^d), \tag{1.9}$$

where p_c and p_d are the prices of clean and dirty energy, respectively. The Bellman equation is

$$V(E^d_{-1}, p_y; b, A) = \max_{E^c, E^d} \pi(E^c, E^d, E^d_{-1}, p_y; b, A) + \beta \mathbb{E}_{p'_y|p_y} V(E^d, p'_y; b, A), \tag{1.10}$$

where $\beta \in (0, 1)$ is the discount factor.

1.4.4 Entry

There is a fixed mass of potential entrants normalized to 1. Each draws a clean share type $b \in (0, 1)$, and a productivity type $A > 0$ from the independent distributions G_b and G_A.

After entering, a plant can freely choose its initial dirty energy input E^d, such that its value function after entry is

$$V_{\text{entry}}(b, A) = \max_{E^d} V(E^d, \bar{p}_y; b, A), \tag{1.11}$$

where $\bar{p}_y$ is the unconditional mean of the idiosyncratic price process.

The entry cost is

$$f^e(b) = f^e_0 \exp(f^e_1 b), \tag{1.12}$$

where $f^e_0 > 0$ governs the scale, and f^e_1 represents the semi-elasticity of the entry cost with respect to the clean share: a 1 percentage point increase in b increases the entry cost by approximately f^e_1 percent.

A plant enters, if the value of entry exceeds the entry cost,

$$V_{\text{entry}}(b, A) \geq f^e(b). \tag{1.13}$$

1.4.5 Solution

The Bellman equation has the following scaling property in A (see appendix section 1.A.1):

$$V\left(\mu E^d_{-1}, p_y; b, \mu A\right) = \mu V(E^d_{-1}, p_y; b, A), \quad \forall \mu > 0. \tag{1.14}$$

This implies $V(E^d_{-1}, p_y; b, A) = AV(A^{-1}E^d_{-1}, p_y; b, 1) = AV(\tilde{E}^d_{-1}, p_y; b, 1)$: the value function is linear in A when appropriately scaling E^d_{-1}. Thus, I need to solve it only for $A = 1$, and can then rescale the value and policy functions by simply multiplying by some A'.

This carries through to the entry value function, which is then $V_{\text{entry}}(b, A) = AV_{\text{entry}}(b, 1)$. With this, I define a cutoff for entry in productivity for each clean share type b, $\bar{A}(b)$:

$$\bar{A}(b) = \frac{f^e(b)}{V_{\text{entry}}(b, 1)}. \tag{1.15}$$

A plant with a draw (b, A) enters iff $A \geq \bar{A}(b)$.

I specify G_A as a Pareto distribution with scale parameter $A_{\min}$ and shape parameter γ. Only the relative scale between $A_{\min}$ and the scale of the entry cost f^e_0 matters, so without loss of generality I normalize $A_{\min} = 1$. For now suppose that f^e_0 is such that $\bar{A}(b) \geq A_{\min} = 1$ for all b. This ensures that the selection mechanism is active for all levels of b. Then, the conditional distribution of $A|b$ among entering plants is a left-truncated version of G_A. In the case of a Pareto, the truncated distribution is also a Pareto with the same shape parameter γ and scale parameter equal to the point of truncation, $\bar{A}(b)$.

The density of the conditional distribution of entering plants' clean share types, $g_b(b|A \geq \bar{A}(b))$, is given by Bayes' rule (with a slight abuse of notation):

$$g_b(b|A \geq \bar{A}(b)) = \Pr(b|A \geq \bar{A}(b)) = \frac{\Pr(A \geq \bar{A}(b)|b)\Pr(b)}{\Pr(A \geq \bar{A}(b))}. \tag{1.16}$$

Since the distribution of A is Pareto with unit scale, $\Pr(A \geq \bar{A}(b)|b) = \left(\bar{A}(b)\right)^{-\gamma}$. Note that the denominator is the integral of the numerator over the support of b. It equals the share of potential entrants that do enter. The distribution of b among entering plants is then proportional to the ex-ante distribution, down-scaled by the entry cutoff at each b raised to the power of the Pareto shape parameter.

1.4.6 Illustration

To illustrate the interaction between the elasticity of substitution and adjustment costs, I simulate impulse response functions of energy use to a positive demand shock for different values of σ and ϕ_1. Figure 1.6 plots the results. The shock is a one standard deviation increase in the idiosyncratic price p_y, with persistence of 0.2.

In the case of no adjustment costs, $\phi_1 = 0$, both energy inputs respond equally to the shock, regardless of the elasticity of substitution. This is because the model effectively reduces to a sequence of static problems, and the magnitude of the response is determined by the output response to the demand shock only. The optimal ratio between both inputs is determined only by their relative prices, which remains constant. With adjustment costs, the responses differ: In general, dirty energy responds less than clean energy. For a given level of adjustment costs, a higher elasticity of substitution leads to a larger difference in responses. Vice versa, for a given elasticity of substitution, higher adjustment costs lead to a larger difference in responses. With an increasing adjustment cost parameter, the response of dirty energy becomes more muted, but also more persistent. It is this behavior that I exploit to separately identify the elasticity of substitution and adjustment costs in the estimation.

1.4.7 Adjustment costs in the estimation of the elasticity of substitution

Adjustment costs introduce a dynamic consideration into the plant's choice of energy inputs. Let $b = 1/2$ for simplicity, and consider the

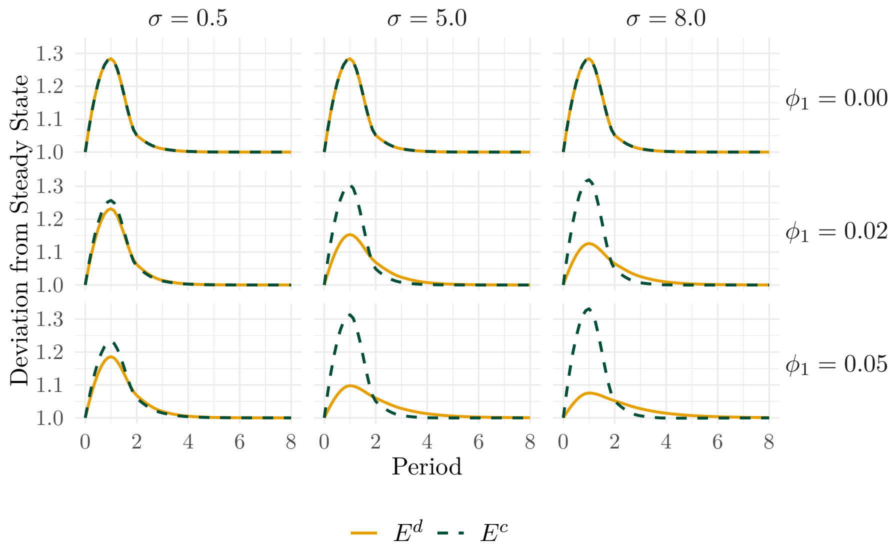


Figure 1.6: Smoothed impulse response functions of clean and dirty energy inputs E^c and E^d in response to a 1 standard deviation demand shock with a persistence of 0.2 in period 1. Without adjustment costs, $\phi_1 = 0$, the response of both inputs is the same: The model is effectively a sequence of static problems, and the magnitude of the response is determined by the output response to the demand shock only. For a positive adjustment cost of dirty energy, $\phi_1 > 0$, the responses of clean and dirty energy diverge: Clean energy always responds more than dirty. This difference is increasing in the elasticity of substitution between the two inputs, σ, and the adjustment cost parameter, ϕ_1. The adjustment cost parameter further determines the persistence of the response of dirty energy: it is increasing in ϕ_1.

production function in equation (1.5)

$$E(E^c, E^d; b = 1/2) = \left[(E^c)^{\frac{\sigma-1}{\sigma}} + (E^d)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}.$$

With input prices p_c and p_d, the first order conditions, and the optimal ratio of energy inputs in logs is

$$0 = \frac{\partial E(E^c, E^d)}{\partial E^c} - p_c,$$
$$0 = \frac{\partial E(E^c, E^d)}{\partial E^d} - p_d,$$

$$\implies \log \frac{E^d}{E^c} = \sigma \log \frac{p_c}{p_d}.$$

This last equation is commonly used in the literature to estimate the elasticity of substitution between two inputs.

Consider now a simplified version of the dynamic model with adjustment costs,

$$\begin{aligned} V(E^d_{-1}, p_y) = \max_{E^c, E^d} & \; p_y Y(E^c, E^d) - p_c E^c - p_d E^d - \phi(E^d_{-1}, E^d) \\ & + \beta \mathbb{E}_{p'_y | p_y} V(E^d, p'_y). \end{aligned}$$

The first-order conditions for E^c and E^d are

$$\begin{aligned} 0 &= p_y \frac{\partial Y(E^c, E^d)}{\partial E^c} - p_c, \\ 0 &= p_y \frac{\partial Y(E^c, E^d)}{\partial E^d} - p_d - \frac{\partial \phi(E^d_{-1}, E^d)}{\partial E^d} + \beta \frac{\partial \mathbb{E} V(E^d, p'_y)}{\partial E^d}. \end{aligned}$$

Compared to the static case, there are two additional terms in the first-order condition for E^d.

Suppose for simplicity that p_y follows iid shocks, such that $\mathbb{E}V(E^d, p'_y) = \mathbb{E}V(E^d)$. Further suppose that the plant has had a series of identical shocks, such that it is at steady state with respect to its dirty energy use, $E^d_{-2} = E^d_{-1}$. If the plant is at steady state, then

$$\left. \frac{\partial \mathbb{E} V(E^d)}{\partial E^d} \right|_{E^d = E^d_{-1}} = 0.$$

Given concavity of the value function, any deviation $E^d \neq E^d_{-1}$ reduces the expected value. Additionally, the marginal adjustment cost is always positive for $E^d \neq E^d_{-1}$. Thus, at steady state, the plant chooses a value of E^d closer to E^d_{-1} than in the absence of adjustment costs. A plant reacts less to changing prices when there are adjustment costs.

In simulations, this seems to extend to the dynamic case with persistent shocks as well. There are no trends in the model, so on average plants are close to their steady state.

This has an important implication for the estimation of the elasticity of substitution. If there are adjustment costs and the estimation approach based on the static first-order conditions is used, the estimate will be biased downwards relative to the true elasticity of substitution. That estimate then represents a reduced-form parameter, which combines the true elasticity of substitution, the adjustment cost parameter, and the plant's expectations about the price trajectory. It is not necessarily informative about the true elasticity, which governs the potential for substitution to permanent price changes.

1.5 Identification and Estimation

I estimate the parameters of the model to replicate empirical results from section 1.3. The scaling property of the value function allows me to separate the estimation of the dynamic within-plant part of the model, and the cross-sectional entry part.

1.5.1 Within-plant dynamics

Table 1.5 lists the parameters that govern the within-plant dynamics.

Parameter	Description	Identified by
σ	Elast. of subst. between E^d, E^c	Ratio of factor demand elasticities
ϕ_1	Adj. cost parameter	Autocorrelation in $\Delta \log E^d$, $\Delta \log E^c$
ρ_y	Demand shock persistence	Autocorrelation in revenue
σ_{p_y}	Demand shock volatility	Within-firm variance in revenue
b	Energy share parameter	Distribution of energy mix
α	Returns to scale parameter	External
β	Discount factor	External
p_c, p_d	Energy prices	External

Table 1.5: Within-plant model parameters and identification.

A crucial assumption is that b is fixed over time for each plant. This is necessary to identify the elasticity of substitution (Diamond et al., 1978).

Elasticity of substitution and adjustment cost The difficulty is in disentangling the adjustment cost parameter and the elasticity of substitution. For transitory shocks, they have similar effects on the dynamics of dirty energy use: It could be that dirty energy is adjusted much less than clean energy either because there is a slight adjustment cost and they are highly substitutable, or the adjustment cost is very high, such that even though they are not very substitutable, it is not adjusted much.

One solution is to consider permanent shocks and long differences. Over an increasingly long horizon the adjustment cost becomes less relevant. This approach is infeasible in my case, since there is no persistent change in the relative price of clean to dirty energy over the sample period.

Instead, I separate the two by considering the dynamics of the changes of dirty energy use compared to clean energy use and output. Suppose there is some persistence in the demand shocks, $\rho_y \in (0, 1)$, and that there is some adjustment cost, $\phi_1 > 0$. Then, compare the two limiting cases of σ in the CES aggregator of clean and dirty energy: In the Leontief case ($\sigma \to 0$), dirty energy use must adjust the same as clean energy use in relative terms, the plant must bear the adjustment cost if it wants to increase output. In contrast, in the perfect substitutes case ($\sigma \to \infty$), dirty energy use will not adjust at all, while clean energy use will adjust the same as output. The relationship is monotonic for intermediate values of σ; the relative response of dirty energy use to that of clean energy use and output varies between these two extremes.

Now consider the case of some $\sigma > 0$, and let the adjustment cost ϕ_1 vary. In the case of no adjustment cost, the plant will maintain the optimal energy mix, and adjust clean and dirty energy by the same relative amount. The persistence in both energy inputs is then the same as that of output. In the case of infinite adjustment cost, dirty energy use will not change at all. Dirty energy use will be perfectly persistent, while clean energy use will adjust in line with output. This relationship between the relative response of dirty energy use to that of clean energy use and output, and the level of ϕ_1 is also monotonic.

Combining these two arguments, for a given relative response of energy use to an output shock, there is a locus of combinations of σ and ϕ_1 that are consistent with it. The parameters are then separately identified from the persistence of the *changes* of the use of either energy: Conditional on the persistence of the demand shock, dirty energy use is more persistent; the plant will smooth the adjustment. In contrast, clean energy use exhibits lower persistence in its changes: it will be adjusted optimally, given the current level of dirty energy. Figure 1.7 illustrates this identification argument.

Demand process The demand shock parameters ρ_y and σ_y are identified from the auto-correlation and variance in plant-level revenue, conditional on the production function parameters.

Distribution of clean share parameters Assuming the plant is in a steady state, the optimal energy mix satisfies

$$\log \frac{E^d}{E^c} = \sigma \left[\log \frac{p_c}{p_d} + \log \frac{1-b}{b} \right]. \tag{1.17}$$

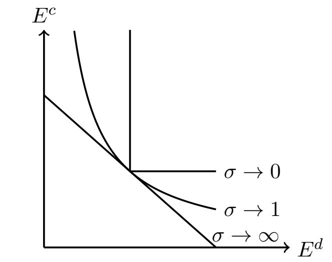
Conditional on the relative price and σ, the distribution of the clean share parameter b is identified from the distribution of the log of the energy mix in the data.

1.5.2 Cross-section

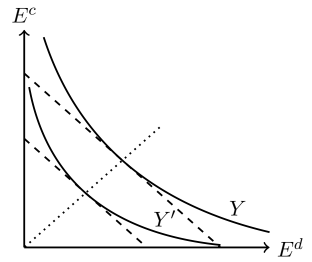
Table 1.6 lists the parameters that govern the entry and cross-sectional part of the model.

Parameter	Description	Identified by
G_A	Distr. of productivity	Observed marginal distr. of revenue
G_b	Distr. of energy share param.	Observed marginal distr. of energy mix
f_0^e	Entry cost scale	Normalized
f_1^e	Entry cost semi-elasticity w.r.t. b	Cross-sect. corr. between size and mix

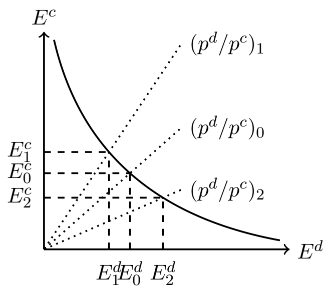
Table 1.6: Across-plant model parameters and identification.



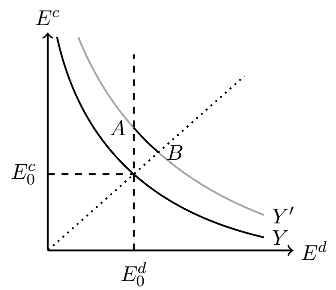
(a) CES production isoquants at different levels of the elasticity of substitution σ.



(b) Optimal input mix at point of tangency between isoquant (solid) and isocost lines (dashed).



(c) σ is identified by exogenous variation in the relative input prices p^d/p^c and the resulting changes in the optimal input mix.



(d) With adjustment cost in E^d, the plant leaves the optimal input ray in response to a demand shock to a point on the segment AB.

Figure 1.7: Illustration of the identification of the elasticity of substitution σ in the case with and without adjustment costs. Panel (a) shows CES production isoquants at different levels of σ. The curvature of the isoquants identifies σ. Panel (b) shows how the optimal input mix is determined by the tangency between isoquants and isocost lines. The set of those points lies on a ray from the origin, the slope of which depends on the relative input prices and σ. In the absence of adjustment costs, σ is identified from exogenous variation in relative input prices and the resulting changes in the optimal input mix, as shown in panel (c). Panel (d) illustrates the case with adjustment cost in E^d: In the case of infinite adjustment costs, the plant moves to point A on the new isoquant. In the case of zero adjustment costs, the plant moves to point B. For intermediate adjustment costs, the plant moves to a point between A and B. This allows tracing out the curvature of the isoquant, and thereby identifying σ, from changes in output.

Productivity distribution I specify G_A as Pareto with shape parameter γ and scale parameter $A_{\text{min}} = 1$. From section 1.4.5 we have that a plant enters if its productivity exceeds the entry threshold $\bar{A}(b)$, given its draw for b. The distribution of entering plants' productivity is then a truncated Pareto, with shape parameter γ and scale parameter $\bar{A}(b)$.

The average revenue of a plant depends on its productivity A, as well as its clean share b, through the marginal cost. A result from the theory of regular variation is that since the between-plant differences in marginal cost from b are bounded, the tail index of the revenue distribution is the same as that of the productivity distribution. Therefore, I can estimate γ directly from the revenue distribution, without the need to control for b.

I estimate the shape parameter γ to match the top k sales shares of plants in the data: I residualize revenue by industry and year fixed effects, then, for $k \in \{3, 4, 5, 10, 15, 20, 50\}$ I calculate the share of the residualized revenue accounted for by the top k plants in year. The shares are stable over time, and I take the average over years as the target moments.

Clean share distribution I specify the distribution G_b as Beta with shape parameters α_b, β_b. Conditional on $\bar{A}(b)$ and γ, the ex-ante distribution G_b is identified from the ex-post distribution of clean shares, as in equation (1.16). I fit a Beta distribution to the observed ex-post density of the distribution of clean shares in the data, $g_b(b|A \geq \bar{A}(b))$. The ex-ante distribution is then given by rearranging equation (1.16):

$$g_b(b) \propto g_b(b|A \geq \bar{A}(b))\bar{A}(b)^{\gamma}. \tag{1.18}$$

If $G_b(b|A \geq \bar{A}(b))$ is Beta, and given the scaling term $\bar{A}(b)^{-\gamma}$, the density g_b is not exactly Beta in general. The Beta distribution is a good approximation of the observed ex-post distribution, and assuming a Beta for the ex-ante distribution leads to a simulated sample that can be well-approximated by a Beta as well.

Entry cost The size of the pool of potential entrants is not identified, and neither is the scale of the entry cost f^e_0. I normalize the size of the pool of potential entrants, and make the assumption that the scale of the entry cost is such that $\bar{A}(b) \geq 1 = A_{\min}$ for all $b \in (0, 1)$. This ensures that the selection mechanism is active for all clean share types.

The semi-elasticity of the entry cost with respect to the clean share, f^e_1, is identified from the relationship between plants' clean share and their size in the data. The target moment is the regression coefficient in equation (1.2), the cross-sectional energy mix-size correlation.

1.5.3 Target moments

Table 1.7 lists the empirical target moments, their standard errors, and the corresponding simulated moments. The loss function is the weighted sum of squared deviations between simulated and empirical moments, with weights given by the inverse of the squared standard errors.

Description	Informs	Data (95% CI)	Model
Ratio of fact. dem. elast. E^d/E^c, $\Delta t = 1$	σ, ϕ_1	0.26 (0.08, 0.44)	0.42
Ratio of fact. dem. elast. E^d/E^c, $\Delta t = 2$	σ, ϕ_1	0.37 (0.19, 0.55)	0.48
Autoregressive coef. $\Delta \log E^d$	σ, ϕ_1	0.18 (0.14, 0.22)	0.28
Autoregressive coef. $\Delta \log E^c$	σ, ϕ_1	0.11 (0.05, 0.17)	0.00
Autoregressive coef. $\Delta \log p_y Y$	ρ_y	0.16 (0.11, 0.21)	0.04
Std. dev. revenue	σ_y	0.08	0.10
Cross-sect. corr. energy mix and size	f^e_1	-0.10 (-0.11, -0.09)	-0.10
Rev. share of top $\{3, 4, 5, 10, 15, 20\}$ plants	G_A	{0.028, 0.033, 0.038, 0.057, 0.070, 0.080}	{0.030, 0.036, 0.040, 0.057, 0.069, 0.078}
Distribution of $\log E^d/E^c$	G_b	$\mathcal{N}(0.06, 1.15^2)$	$\mathcal{N}(0.06, 1.15^2)$

Table 1.7: Estimation targets with confidence intervals, and simulated moments.

Moment calculation

Ratio of factor demand elasticities The regression equations correspond to their empirical counterpart, equation (1.3), without the fixed effects:

$$\Delta_k \log E^d_{it} = \beta^d_k \Delta_k \log Y_{it} + \epsilon^d_{it},$$

$$\Delta_k \log E^c_{it} = \beta^c_k \Delta_k \log Y_{it} + \epsilon^c_{it}.$$

I estimate these with 2SLS like the empirical counterpart, using the true simulated demand shock as instruments for the change in output. I then take the ratio β^d_k/β^c_k as the target moment, for $k \in \{1, 2\}$.

Persistence in changes I estimate the persistence in the changes of energy use and output from the following autoregressive equations:

$$\begin{aligned}
\Delta \log E^d_{it} &= \rho_{E^d} \Delta \log E^d_{i,t-1} + \epsilon^d_{it}, \\
\Delta \log E^c_{it} &= \rho_{E^c} \Delta \log E^c_{i,t-1} + \epsilon^c_{it}, \\
\Delta \log[p_{y,it} Y_{it}] &= \rho_Y \Delta \log[p_{y,it} Y_{i,t-1}] + \epsilon^Y_{it}.
\end{aligned}$$

These estimation equations suffer from Nickell bias, so I instrument for the lagged difference using the level of the second lag (Anderson-Hsiao IV):

$$\Delta \log E^d_{i,t-1} = \tilde{\beta}_d \log E^d_{i,t-2} + \tilde{\epsilon}^d_{it},$$

and similar for clean energy and output.

Standard deviation of revenue I calculate the standard deviation of within-plant revenue as the mean squared residual of the regression of the persistence in revenue, ϵ^Y_{it}.

Cross-sectional energy mix-size correlation I estimate the cross-sectional energy mix-size correlation from the following regression equation:

$$\log \frac{E^d_{it}}{E^c_{it}} = \beta_{\text{size}} \log\left[p_{y,it} Y_{it}\right] + \epsilon_{it}.$$

1.5.4 Implementation

Within-plant model

First, I solve the plant's dynamic programming problem by value function iteration over a three-dimensional grid of E^d_{-1}, p_y, and b.[11] The results are the value function, policy functions for E^d, and optimal choices for E^c, given the policy for E^d. I simulate a panel of plants, and calculate the relevant moments from the simulated data, using the following algorithm:

1. Set the number of firms $N = 30,000$ and time periods $T = 120$ (the first 100 are burn-in), to match the empirical panel dimensions.

2. Draw N clean share types $\{b_i\}$ from the grid, such that the distribution matches the observed empirical distribution.[12]

3. Generate N discretized demand shock processes $\{p_{y,it}\}$ of length T each.

4. Initialize $\{E^d_{it=0}\}$ as the steady state level given b_i and $p_{y,it=0} = 1$ (The steady state is the fixed point in the policy function for E^d).

5. Solve the model forward for each plant i and time $t = 1, \ldots, T$:

   (a) Given $E^d_{i,t-1}$, $p_{y,it}$, and b_i, interpolate the policy functions to get optimal E^d_{it} and E^c_{it}.

   (b) Calculate output Y_{it}.

   (c) Go to the next period.

6. Discard the first 100 periods as burn-in.

[11] I discretize the process for p_y using the Rouwenhorst method.

[12] I match the empirical distribution of $\log E^d/E^c$ as follows. Suppose there is no adjustment cost, $\phi_1 = 0$. Then, from the optimal input mix equation, $\log E^d/E^c = \sigma[\log p_c/p_d + \log(1-b)/b]$. In the data, $\log E^d_{it}/E^c_{it} \sim \mathcal{N}(0, 1.1^2)$. Assuming the relative price is constant, this implies that b follows a logit-normal distribution, with variance depending on the elasticity of substitution σ. I draw sampling weights for each grid value of b, such that the weighted distribution of b in the sample matches the implied logit-normal distribution.

7. Calculate target moments from the simulated data, as described in section 1.5.3.

To estimate the parameters, I first run a coarse grid search over the parameter space. Then, with the best two results from the grid search as initial values, I run a simulated annealing algorithm, subject to parameter bounds. The algorithm converges to a point well within the parameter bounds.

Cross-sectional model

For the cross-section, I need a sample of types (b, A) from the conditional distribution of entering plants. Given the sample, I then simulate the dynamic decisions of the plant, as in the within-plant estimation.

First, I generate a sample of the conditional distribution $(b, A)|A > \bar{A}(b)$ using a two-stage accept-reject algorithm. Note that $\bar{A}(b)$ depends on the entry cost parameters. I normalize the scale f_0^e such that $\inf_b \bar{A}(b) = 1.1 > A_{\min} = 1$, to ensure that there is selection at all values of b. Then, I estimate f_1^e to match the energy mix-size correlation in the data. Conditional on a candidate value for f_1^e, and the estimation of the within-plant parameters, I generate a sample as follows:

1. Approximate the observed conditional distribution of b as a logit-normal distribution, as in the within-plant estimation, with mean $\log p^d/p^c$ and standard deviation $1.1/\sigma$, where σ is the estimated elasticity of substitution. This is the proposal distribution for b.

2. Draw a value from the proposal distribution for b. Accept the draw with a probability proportional to $\bar{A}(b)^\gamma$. This gives a sample from the unconditional distribution of b.

3. Pair the value of b with a draw for A from the unconditional Pareto distribution. Accept the pair if $A \geq \bar{A}(b)$. The result is a sample from the conditional distribution $(b, A)|A \geq \bar{A}(b)$.

4. Repeat until the sample size is $N = 30,000$.

Then, given the sample of types, I simulate the dynamic decisions of each plant as in the within-plant estimation in section 1.5.4. To incorporate the productivity draws, I multiply the value and policy functions by A_i for each plant i: I adjust the state variable $E^d_{-1,it} = A_i \tilde{E}^d_{-1,it}$, the policy function $E^d_{it} = A_i \tilde{E}^d_{it}$, the optimal choice of $E^c_{it} = A_i \tilde{E}^c_{it}$, and the value function $V_{it} = A_i \tilde{V}_{it}$, where the tilde variables are from the solution for $A = 1$. Output is then $Y_{it} = A_i^{1-\alpha} E(E^c_{it}, E^d_{it}; b_i)^\alpha$.

I calculate the target moment, the energy mix-size correlation, from the simulated data as in section 1.5.3. Conditional on the estimation of the within-plant parameters, and on the directly estimated Pareto shape parameter, I estimate f^e_1 using a univariate bounded minimization algorithm, to minimize the distance between the simulated and empirical moments.

Finally, to estimate the unconditional distribution of b, I parameterize a Beta distribution as follows. For each draw of b in the simulated sample, I calculate the weight $w(b) = \bar{A}(b)^\gamma$. Then, I calculate the weighted mean and variance of b in the sample, and solve for the Beta shape parameters α_b, β_b using the method of moments formulas.

1.5.5 Results

Table 1.8 presents the estimated parameter values. The externally estimated parameters are $\alpha = 0.65$ (which would correspond to a demand elasticity of 2.9 in a monopolistic competition model with constant returns to scale production), $\beta = 0.96$ for annual data, and energy prices $p_c = 0.5$ and $p_d = 0.18$. The level of the prices is subject to normalization, and the relative price $p_c/p_d = 2.8$ corresponds to the average aggregate relative price over the sample period.

The estimated value for the adjustment cost parameter is such that in the simulated panel, the adjustment cost represent on average 0.1% of total costs. This small share is sufficient though to generate significant differences the response of dirty energy use, compared to clean in response to output shocks. The elasticity of substitution between clean and dirty energy is estimated as 5.1, which is on the higher end of estimates in the literature. I show in section 1.4.7 that the common estimation approach underestimates the deep elasticity of substitution when there are adjustment costs. This can explain my relatively high estimate.

Parameter	Description	Estimated value
σ	Elast. of subst. between E^d, E^c	5.10
ϕ_1	Adj. cost parameter	0.02
ρ_y	Demand shock persistence	0.01
σ_{p_y}	Demand shock volatility	0.13
G_A	Distr. of productivity	Pareto($\gamma = 1.38, A_{\min} = 1$)
G_b	Distr. of energy share param.	Beta($\alpha_b = 90, \beta_b = 32$)
f_0^e	Entry cost scale	17.55
f_1^e	Entry cost semi-elasticity w.r.t. b	0.38
α	Returns to scale parameter	0.65
β	Discount factor	0.96
p_c, p_d	Energy prices	0.5, 0.18

Table 1.8: Parameters and estimated values.

The estimated value for the Pareto tail parameter is in line with estimates from the literature (it implies somewhat thinner tails than Zipf's law, as estimated in Axtell, 2001; Gabaix, 2016). The distribution of clean share parameters has a mean of 0.74, and a standard deviation of 0.04. The entry cost semi elasticity parameter is such that a 1 p.p. increase in b leads to a 0.38% increase in the entry cost.

Untargeted moments I match an untargeted moment: the within-plant over-time correlation between energy mix and revenue. The empirical correlation is -0.16, the simulated correlation is -0.26. The remaining discrepancy might be related to attenuation bias in the empirical estimate due to measurement error in revenue. This moment is closely related to the elasticity of substitution and adjustment cost parameters, providing some additional validation for their estimation.

1.6 Policy Experiments

I conduct two policy experiments in the parameterized model. First, a subsidy for the entry cost of clean plants. Second, a tax on dirty energy purchases, similar to a carbon tax.

1.6.1 Simplified aggregation

I will conduct steady-state comparisons for the policy experiments. Thus, I will abstract from the dynamics of the model, in particular the demand shocks, and focus on the steady-state aggregates.

Let $E^{d\star}(b, A)$ be the steady-state value of dirty energy consumption for a plant with type (b, A), defined as the fixed point of the policy function at the unconditional mean of the demand shock, $p_y = 1$. Let $E^{c\star}(b, A)$ be the corresponding optimal steady-state clean energy consumption. The aggregates of clean and dirty energy consumption, and output are then given by

$$\begin{aligned}
E^c_{\text{agg}} &= \int_0^1 \int_{\bar{A}(b)}^{\infty} E^{c\star}(b, A) g_b(b) g_A(A) dA db, \\
E^d_{\text{agg}} &= \int_0^1 \int_{\bar{A}(b)}^{\infty} E^{d\star}(b, A) g_b(b) g_A(A) dA db, \\
Y_{\text{agg}} &= \int_0^1 \int_{\bar{A}(b)}^{\infty} A^{1-\alpha} E(E^{c\star}(b, A), E^{d\star}(b, A))^{\alpha} g_b(b) g_A(A) dA db.
\end{aligned}$$

Given the scaling property of the value and policy functions, and the Pareto distribution of A, these simplify to (see appendix section 1.A.2):

$$E^c_{\text{agg}} = \frac{\gamma}{\gamma - 1} \int_0^1 E^{c\star}(b, 1) \bar{A}(b)^{1-\gamma} dG_b(b) \tag{1.19}$$

$$E^d_{\text{agg}} = \frac{\gamma}{\gamma - 1} \int_0^1 E^{d\star}(b, 1) \bar{A}(b)^{1-\gamma} dG_b(b) \tag{1.20}$$

$$Y_{\text{agg}} = \frac{\gamma}{\gamma - 1} \int_0^1 E\left(E^{c\star}(b, 1), E^{d\star}(b, 1)\right)^{\alpha} \bar{A}(b)^{1-\gamma} dG_b(b). \tag{1.21}$$

Recall the definition of the cutoff productivity $\bar{A}(b)$ from equation (1.15):

$$\bar{A}(b) = \frac{f^e(b)}{V_{\text{entry}}(b,1)}.$$

Define the mass of entering plants as

$$m = \int_0^1 \int_{\bar{A}(b)}^{\infty} g_b(b) g_A(A) dA db = \int_0^1 \bar{A}(b)^{-\gamma} dG_b(b). \tag{1.22}$$

1.6.2 Entry subsidy for clean plants

Suppose a policymaker observes the (b, A) draws of the potential entrants. They offer a multiplicative subsidy $s(b) = \exp(-s_1 b)$ for the entry of clean plants. The entry costs are then

$$f^{e,\text{sub}}(b) = f^e(b)s(b) = f_0^e \exp[b(f_1^e - s_1)], \tag{1.23}$$

where s_1 represents the semi-elasticity of the subsidy with respect to the clean share. For values of s_1 up to f_1^e, the policymaker subsidizes part of the additional entry cost for a higher clean share. For values of s_1 larger than f_1^e, the entry cost to be paid by the plant decreases in the clean share.

Under the subsidy, the entry cutoff in productivity becomes

$$\bar{A}^{\text{sub}}(b) = \frac{f^{e,\text{sub}}(b)}{V_{\text{entry}}(b,1)} = \frac{f_0^e}{V_{\text{entry}}(b,1)} \exp[b(f_1^e - s_1)]. \tag{1.24}$$

The subsidy influences aggregates only through changes in the entry cutoff.

The total cost of the subsidy to the policymaker is

$$\begin{aligned} C^{\text{sub}} &= \int_0^1 \int_{\bar{A}^{\text{sub}}(b)}^{\infty} f^e(b)[1 - s(b)] g_A(A) g_b(b) dA db \\ &= \int_0^1 f^e(b)[1 - s(b)] \bar{A}^{\text{sub}}(b)^{-\gamma} dG_b(b). \end{aligned} \tag{1.25}$$

Figure 1.8 shows the results of the entry subsidy policy experiment.

I consider values $s_1 \in (0, 1.2 \times f_1^e)$: between no subsidy and a subsidy that fully reverses the clean share dependence of the entry cost, such that plants with a higher clean share have a lower entry cost.

The policy is expansionary: it increases the mass of entering plants, output, and both clean and dirty energy use. It increases the aggregate share of clean energy, but only minutely so. The policy acts only at the entry margin by construction. Given the Pareto distribution of productivity, aggregates are primarily driven by the most productive plants, which are not affected by the policy.

The policy is expensive: I plot the total cost of the subsidy as fraction of aggregate output.

1.6.3 Tax on dirty energy use

I consider a proportional tax τ_d on the price of dirty energy. The final price is then

$$p_d^{\text{tax}} = p_d(1 + \tau_d). \tag{1.26}$$

The tax is permanent and must be paid by all plants. I consider values $\tau_d \in (0, 1)$, so up to a doubling of the dirty energy price.

I solve for plants' problem under the new price including the tax, and then calculate aggregates for two cases: (i) holding the distribution of plants over (b, A) fixed at the baseline equilibrium, and (ii) allowing the distribution to adjust to the equilibrium under the new prices. Case (i) can be interpreted as the short-run effects of the tax, while case (ii) represents the long-run effects with compositional changes.

The results of the tax policy experiment allowing for compositional changes are shown in figure 1.9. As opposed to the entry subsidy, the tax on dirty energy use is contractionary, since it strictly increases marginal cost. The policy is very effective though at reducing the aggregate use of dirty energy.

The results for the case with the fixed distribution are shown in appendix section 1.C.1. They are similar quantitatively, the main difference being a smaller reduction in dirty energy use, and a smaller decrease in aggregate output, since the mass of firms remains constant.

Comparison to low-σ calibration

To demonstrate the impact of the value of the elasticity of substitution between clean and dirty energy, I repeat the policy experiment in a model with a different calibration. I decrease the value of the elasticity σ from 5.1 to 1.5 and the value of the adjustment cost parameter ϕ_1 to 0, keeping all other parameters fixed. The value 1.5 corresponds to the upper end of estimates of the micro-level elasticity in the literature (Jo, 2024). I solve the value and policy functions for the new parameterization, and compute the new steady-state aggregates under the tax on dirty energy use allowing for compositional changes as above. Figure 1.10 shows the results.

To illustrate, the German emission target for the manufacturing sector until 2030 corresponds to a reduction of dirty energy use by 40% relative to 2018 levels.[13] This same reduction can be implemented with a tax rate of 11% in the baseline calibration, and a tax rate of 23% in the low-elasticity calibration. The corresponding output reductions are 7% and 14%, respectively. This illustrates the importance of the elasticity of substitution between clean and dirty energy for the effectiveness and cost of policies aimed at reducing dirty energy use.

1.7 Conclusion

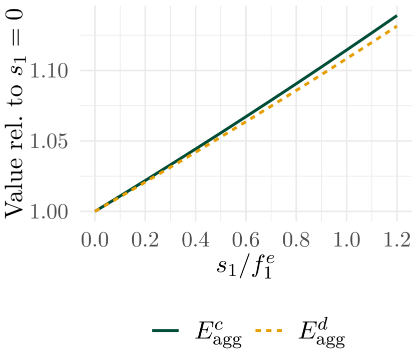
In this paper, I analyze the use of clean and dirty energy in production in the German manufacturing census. I document empirically, that (i) there is substantial heterogeneity in the energy mix across plants, (ii) this heterogeneity is difficult to explain with observables, (iii) some variation can be explained by size, larger plants use a higher share of clean energy,

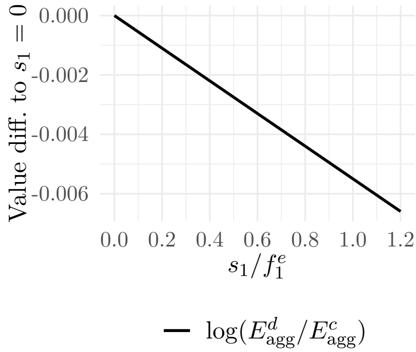
[13] See https://www.bmwi.de/Redaktion/DE/Artikel/Energie/klimaschutz-in-der-industrie.html (accessed 15 November 2025). I use 2018 as the last pre-COVID year for the comparison.

and (iv) the factor demands for clean and dirty energy respond differently to demand shocks, dirty energy use adjusts much less than clean to the same shock to output.

Based on these findings, I develop a dynamic model of heterogeneous plants with entry. Plants differ in productivity and in their technology to combine clean and dirty energy to produce output. A key ingredient of the model to match the data is an adjustment cost for dirty energy use. The model replicates both the within-plant dynamics, and the cross-sectional distribution of plants. I show that estimates of the elasticity of substitution between clean and dirty energy are downward biased if they ignore the presence of adjustment costs. My estimation of the model implies an elasticity of substitution of around 5, which is three times larger than estimates in the literature.

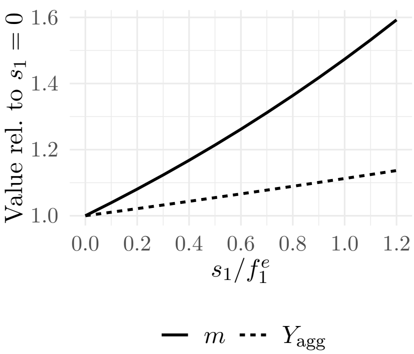
In the calibrated model, I study two policies: an entry subsidy for clean plants, and a tax on dirty energy use. The entry subsidy has only marginal effects: Aggregate use of clean and dirty energy is dominated by large plants, which are unaffected by the policy. A tax on dirty energy use is effective at reducing dirty energy use. I show that the effectiveness of the tax depends strongly on the elasticity of substitution between clean and dirty energy. Given the higher elasticity in my calibration, the tax has much lower contractionary effects on output for a given reduction in dirty energy use, compared to a model with an elasticity closer to estimates in the literature. My results suggest that the plants can turn green at a lower cost than previously thought.

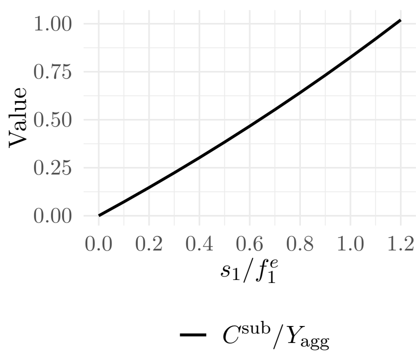




(a) Use of clean and dirty energy.

(b) Energy mix.

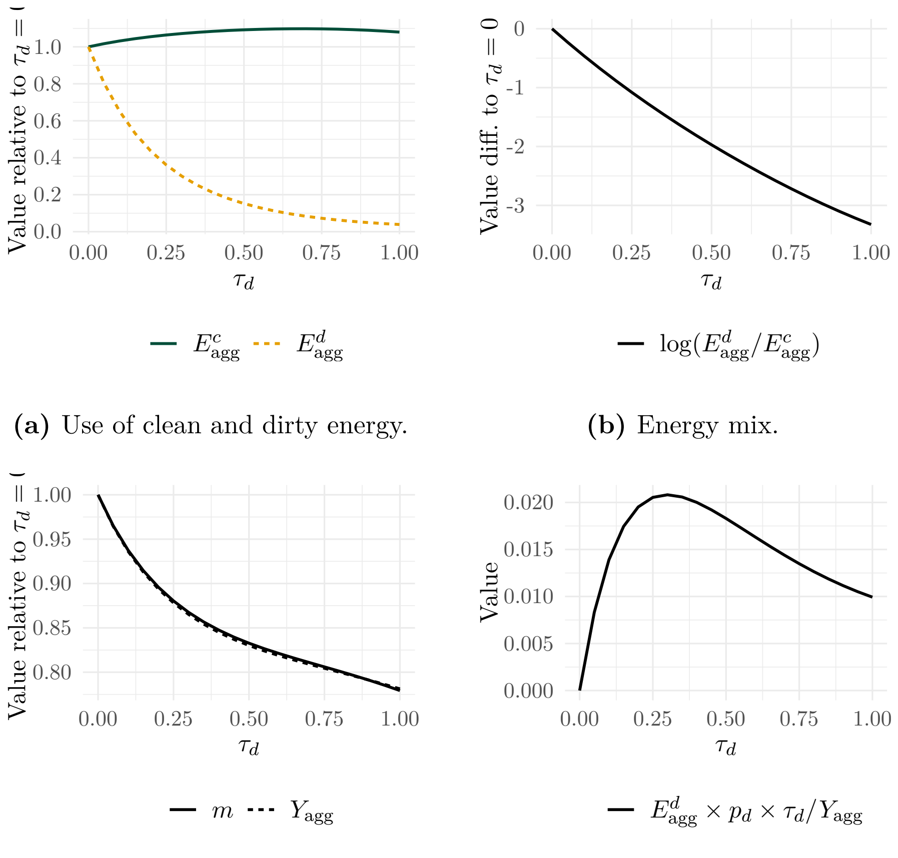




(c) Total output and mass of entering plants.

(d) Total subsidy spending as fraction of output.

Figure 1.8: Results of entry subsidy policy experiment. Panel (a) shows the response of clean and dirty energy use. The entry subsidy is expansionary, the use of both types of energy increases. There is a slight composition effect through the policy, dirty energy use increases relatively less. Panel (b) shows the negligible effect on the aggregate energy mix of the policy. Panel (c) shows total output and the mass of entering plants. The mass increases more than output, the subsidy allows plants with marginal productivity to enter. Panel (d) shows total subsidy spending as a fraction of output.



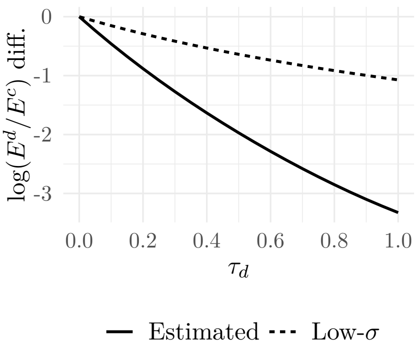
(a) Use of clean and dirty energy.

(b) Energy mix.

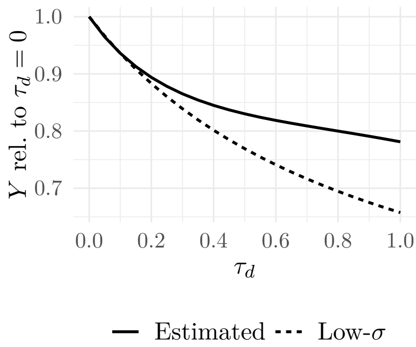
(c) Output and mass of entrants.

(d) Tax revenue as fraction of output.

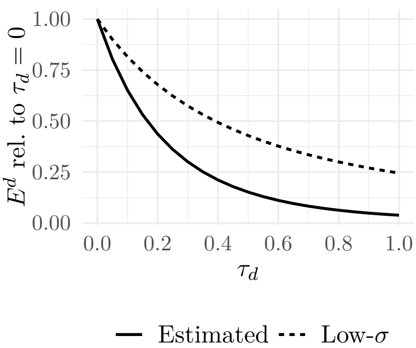
Figure 1.9: Results of tax on dirty energy use policy experiment, allowing for compositional changes. Panel (a) shows the response of clean and dirty energy use. Dirty energy use decreases strongly in the tax rate. Clean energy increases, partially offsetting the reduction in dirty energy use. Panel (b) shows the strong effect on the aggregate energy mix. Panel (c) shows total output and the mass of entering plants. Both decrease, the policy is contractionary. Panel (d) shows tax revenue as a fraction of output.



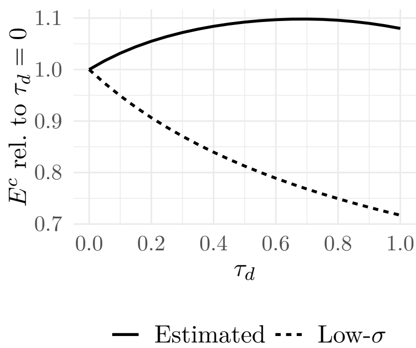
(a) Response of energy mix.



(b) Response of total output.



(c) Response of dirty energy use.



(d) Response of clean energy use.

Figure 1.10: Comparison of the effects of a dirty energy tax between the baseline calibration (solid lines) and the low-elasticity calibration (dashed lines). The baseline calibration uses the parameter values found in the structural estimation (section 1.5). The low-elasticity calibration sets the elasticity of substitution between clean and dirty energy to $\sigma = 1.5$ (Jo, 2024), and the adjustment cost parameter to $\phi_1 = 0$. This illustrates the impact of the value of the elasticity of substitution on policy effectiveness. In the calibrated model, the energy mix is much more responsive, since dirty energy use can easily be reduced, and substituted for with clean energy. The reflects in the smaller output losses associated with achieving the same reduction in dirty energy use.

References

Acemoglu, D., Aghion, P., Bursztyn, L., & Hémous, D. (2012). The environment and directed technical change. *American Economic Review*, *102*(1), 131–166.

Acemoglu, D., Akcigit, U., Hanley, D., & Kerr, W. (2016). Transition to clean technology. *Journal of Political Economy*, *124*(1), 52–104.

Andersson, J. J. (2019). Carbon taxes and CO2 emissions: Sweden as a case study. *American Economic Journal: Economic Policy*, *11*(4), 1–30.

Antràs, P. (2004). Is the U.S. aggregate production function Cobb-Douglas? New estimates of the elasticity of substitution. *Contributions in Macroeconomics*, *4*(1), 20121005.

Arbeitsgemeinschaft Energiebilanzen e.V. (2025). *Evaluation Tables on the Energy Balance for the Federal Republic of Germany 1990 to 2024* [https://ag-energiebilanzen.de/en/data-and-facts/evaluation-tables-on-the-energy-balance/, accessed 02.10.2025].

Arnberg, S., & Bjørner, T. B. (2007). Substitution between energy, capital and labour within industrial companies: A micro panel data analysis. *Resource and Energy Economics*, *29*(2), 122–136.

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science*, *293*(5536), 1818–1820.

Barrot, J.-N., & Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics*, *131*(3), 1543–1592.

Barrows, G., & Ollivier, H. (2018). Cleaner firms or cleaner products? how product mix shapes emission intensity from manufacturing. *Journal of Environmental Economics and Management*, *88*, 134–158.

Boehm, C. E., Levchenko, A. A., & Pandalai-Nayar, N. (2023). The long and short (run) of trade elasticities. *American Economic Review*, *113*(4), 861–905.

Bontadini, F., Corrado, C., Haskel, J., Iommi, M., & Jona-Lasinio, C. (2023, February). *EUKLEMS & INTANProd: Industry productivity accounts with intangibles - sources of growth and productivity trends: Methods and main measurement challenges* (Deliverable No. D2.3.1). Luiss Lab of European Economics.

Bousquet, A., & Ladoux, N. (2006). Flexible versus designated technologies and interfuel substitution. *Energy Economics*, *28*(4), 426–443.

Bundesnetzagentur. (2021). *Monitoring report 2020* (tech. rep.).

Casey, G. (2024). Energy efficiency and directed technical change: Implications for climate change mitigation. *The Review of Economic Studies*, *91*(1), 192–228.

Chirinko, R. S. (2008). σ: The long and short of it. *Journal of Macroeconomics*, *30*(2), 671–686.

Colmer, J., Martin, R., Muûls, M., & Wagner, U. J. (2024). Does pricing carbon mitigate climate change? Firm-level evidence from the European Union Emissions Trading System. *The Review of Economic Studies*, rdae055.

Dechezleprêtre, A., Nachtigall, D., & Venmans, F. (2023). The joint impact of the European Union emissions trading system on carbon emissions and economic performance. *Journal of Environmental Economics and Management*, *118*, 102758.

Destatis. (2023). *Daten zur Energiepreisentwicklung Dezember 2023* [https://www.destatis.de/DE/Themen/Wirtschaft/Preise/Publikationen/Energiepreise/statistischer-bericht-energiepreisentwicklung-5619001.html, accessed 14.03.2024].

Diamond, P., McFadden, D., & Rodriguez, M. (1978). Measurement of the elasticity of factor substitution and bias of technical change. In M. Fuss & D. McFadden (Eds.), *Applications of the theory of production* (pp. 125–147, Vol. 2). Elsevier.

Eslava, M., Haltiwanger, J., Kugler, A., & Kugler, M. (2004). The effects of structural reforms on productivity and profitability enhancing reallocation: Evidence from Colombia. *Journal of Development Economics*, *75*(2), 333–371.

Fraunhofer ISI. (2025). *Erstellung von Anwendungsbilanzen für die Jahre 2021 bis 2023 für die Sektoren Industrie und GHD* [https://ag-energiebilanzen.de/wp-content/uploads/2024/11/Anwendungsbilanz_Industrie_2023_final_20250324.pdf, accessed 02.10.2025].

Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives*, *30*(1), 185–206.

Gerster, A., & Lamp, S. (2024). Energy tax exemptions and industrial production. *The Economic Journal*.

Golosov, M., Hassler, J., Krusell, P., & Tsyvinski, A. (2014). Optimal taxes on fossil fuel in general equilibrium. *Econometrica*, *82*(1), 41–88.

Hawkins-Pierot, J. T., & Wagner, K. R. H. (2025, September). *Technology lock-in and costs of delayed climate policy* (Working Paper).

Hyland, M., & Haller, S. (2018). Firm-level estimates of fuel substitution: An application to carbon pricing. *The Energy Journal*, *39*(6), 71–98.

Jo, A. (2024). Substitution between clean and dirty energy with biased technical change. *International Economic Review*.

Kaartinen, M., & Prane, O. (2024, November). *Carbon pricing and fuel switching by firms: Theory and evidence* (Job Market Paper). Department of Economics, Stockholm University.

Leclair, E. M. (2025, January). *Balancing production and carbon emissions with fuel substitution* (Working Paper). University of Chicago.

Linn, J. (2008). Energy prices and the adoption of energy-saving technology. *The Economic Journal*, *118*(533), 1986–2012.

Marin, G., & Vona, F. (2021). The impact of energy prices on socioeconomic and environmental performance: Evidence from French manufacturing establishments, 1997–2015. *European Economic Review*, *135*, 103739.

Marschak, J., & Andrews, W. H. (1944). Random simultaneous equations and the theory of production. *Econometrica*, *12*(3/4), 143–205.

Martin, R., de Preux, L. B., & Wagner, U. J. (2014). The impact of a carbon tax on manufacturing: Evidence from microdata. *Journal of Public Economics*, *117*, 1–14.

Martinsson, G., Sajtos, L., Strömberg, P., & Thomann, C. (2024). The effect of carbon pricing on firm emissions: Evidence from the Swedish CO2 tax. *The Review of Financial Studies*, *37*(6), 1848–1886.

Oberfield, E., & Raval, D. (2021). Micro data and macro technology. *Econometrica*, *89*(2), 703–732.

Papageorgiou, C., Saam, M., & Schulte, P. (2017). Substitution between clean and dirty energy inputs: A macroeconomic perspective. *The Review of Economics and Statistics*, *99*(2), 281–290.

Peter, A., & Ruane, C. (2025, May). *The aggregate importance of intermediate input substitutability* (Working Paper No. 31233). NBER.

Research Data Centre of the Statistical Offices of the Federal States. (2023). *AFiD Official Firm Data for Germany* [Modules: *Produkte* (doi:10.21242/42131.2021.00.03.1.1.0), *Energieverbrauch* (doi:10.21242/43531.2021.00.03.1.1.0), *Industriebetriebe* (doi:10.21242/42111.2021.00.01.1.1.0), and *Industrieunternehmen* (doi:10.21242/42221.2021.00.01.1.1.0)].

Rottner, E., & von Graevenitz, K. (2024). What drives carbon emissions in German manufacturing: Scale, technique or composition? *Environmental and Resource Economics*.

Shapiro, J. S., & Walker, R. (2018). Why is pollution from US manufacturing declining? The roles of environmental regulation, productivity, and trade. *American Economic Review*, *108*(12), 3814–3854.

Stern, D. I. (2012). Interfuel substitution: A meta-analysis. *Journal of Economic Surveys*, *26*(2), 307–331.

Appendices

Appendix 1.A Derivations

1.A.1 Bellman equation scaling property

Static case Start with the static case, without adjustment cost and demand shocks. The problem is

$$\begin{aligned} V(A,b) &= \max_{E^c,E^d} p_y Y(E^c,E^d;A,b) - p_c E^c - p_d E^d, \\ Y(E^c,E^d;A,b) &= A^{1-\alpha} E(E^c,E^d;b)^{\alpha}, \\ E(E^c,E^d;b) &= \left[b(E^c)^{\frac{\sigma-1}{\sigma}} + (1-b)(E^d)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}. \end{aligned}$$

The claim in this case is

$$V(A,b) = AV(1,b).$$

Dual unit cost $p_E(b)$. Because E is homogeneous of degree 1 in (E^c,E^d), the minimum cost of producing $\bar{E}$ units is $p_E(b)\cdot\bar{E}$ where

$$p_E(b) = \min_{E^c,E^d\geq 0} p_c E^c + p_d E^d \quad \text{s.t.} \quad E(E^c,E^d;b) = 1.$$

The Lagrangian first-order conditions are, using $\partial E/\partial E^c = b(E/E^c)^{1/\sigma}$ and $\partial E/\partial E^d = (1-b)(E/E^d)^{1/\sigma}$,

$$\begin{aligned} p_c &= \mu b (E^c)^{-1/\sigma}, \\ p_d &= \mu(1-b)(E^d)^{-1/\sigma}, \end{aligned}$$

where the second equality uses $E=1$ on the constraint. Dividing eliminates μ and gives the optimal input ratio,

$$\frac{E^c}{E^d} = \left(\frac{b}{1-b} \cdot \frac{p_d}{p_c} \right)^{\sigma}.$$

Inverting each FOC gives conditional factor demands as functions of μ,

$$E^c = \left(\frac{\mu b}{p_c}\right)^{\sigma}, \qquad E^d = \left(\frac{\mu(1-b)}{p_d}\right)^{\sigma}.$$

Substituting into the constraint $E = 1$ and simplifying,

$$1 = b\left(\frac{\mu b}{p_c}\right)^{\sigma-1} + (1-b)\left(\frac{\mu(1-b)}{p_d}\right)^{\sigma-1} = \mu^{\sigma-1}\left[b^{\sigma}p_c^{1-\sigma} + (1-b)^{\sigma}p_d^{1-\sigma}\right],$$

so $\mu = [b^{\sigma}p_c^{1-\sigma} + (1-b)^{\sigma}p_d^{1-\sigma}]^{1/(1-\sigma)}$. Total expenditure at the optimum is $p_cE^c + p_dE^d = \mu^{\sigma}[b^{\sigma}p_c^{1-\sigma} + (1-b)^{\sigma}p_d^{1-\sigma}] = \mu$, giving

$$p_E(b) = \left[b^{\sigma}p_c^{1-\sigma} + (1-b)^{\sigma}p_d^{1-\sigma}\right]^{\frac{1}{1-\sigma}}.$$

Static profit and A-scaling. Optimizing over the scalar E using the dual cost,

$$\pi^{\star}(b, A) = \max_{E\geq 0} p_y A^{1-\alpha}E^{\alpha} - p_E(b)E.$$

The FOC $p_y\alpha A^{1-\alpha}E^{\alpha-1} = p_E(b)$ yields

$$E^{\star}(b, A) = A\left[\frac{p_y\alpha}{p_E(b)}\right]^{\frac{1}{1-\alpha}},$$

which scales linearly in A. Substituting back and collecting terms,

$$V(A, b) = \pi^{\star}(b, A) = A(1-\alpha)p_y^{\frac{1}{1-\alpha}}\left[\frac{\alpha}{p_E(b)}\right]^{\frac{\alpha}{1-\alpha}} = A \cdot V(1, b).$$

Output and both factor demands scale linearly in A, completing the static case.

Dynamic case The Bellman equation is

$$
\begin{aligned}
V(E^d_{-1}, p_y; b, A) = \max_{E^c, E^d} \; & p_y A^{1-\alpha} F(E^c, E^d; b) \\
& - p_c E^c - p_d E^d - \phi(E^d_{-1}, E^d) \\
& + \beta \mathbb{E}_{p'_y | p_y} V(E^d, p_y; b, A), \\
\phi(E^d_{-1}, E^d) = & \frac{\phi_1}{2E^d_{-1}} \left(E^d - E^d_{-1} \right)^2, \\
\log p'_y = & \rho \log p_y + \sigma_{p_y} \epsilon', \; \epsilon' \sim N(0,1).
\end{aligned}
$$

The claim is

$$
V(E^d_{-1}, p_y; b, A) = AV \left(A^{-1} E^d_{-1}, p_y; b, 1 \right).
$$

Consider first the felicity function:

$$
p_y A^{1-\alpha} F(E^c, E^d; b) - p_c E^c - p_d E^d - \frac{\phi_1}{2E^d_{-1}} \left(E^d - E^d_{-1} \right)^2 .
$$

If we can factor out A after adjusting the state variable E^d_{-1} by A^{-1}, we show that the felicity function scales linearly in A. So, what must μ be such that

$$
\begin{aligned}
& p_y A^{1-\alpha} F(E^c, E^d; b) - p_c E^c - p_d E^d - \frac{\phi_1}{2E^d_{-1}} \left(E^d - E^d_{-1} \right)^2 \\
& = A \left[p_y F(\mu E^c, \mu E^d; b) - p_c \mu E^c - p_d \mu E^d - \frac{\phi_1}{2E^d_{-1} A^{-1}} \left(\mu E^d - E^d_{-1} A^{-1} \right)^2 \right] ?
\end{aligned}
$$

The production function is homogeneous of degree α:

$$
F(\mu E^c, \mu E^d; b) = \mu^\alpha F(E^c, E^d; b).
$$

$\mu = A^{-1}$ satisfies the equation. This scales the within-period factor demands for E^c, and critically for E^d, which becomes the state variable in the next period. The scaling is consistent with the proposed adjustment of the state variable. This concludes the proof.

1.A.2 Aggregation

I present the aggregation for clean energy use. The steps are analogous for dirty energy use and output.

$$
\begin{aligned}
E^c_{\text{agg}} &= \int_0^1 \int_{\bar{A}(b)}^{\infty} E^{c\star}(b, A) g_b(b) g_A(A) dA db \\
&= \int_0^1 \int_{\bar{A}(b)}^{\infty} A E^{c\star}(b, 1) g_b(b) g_A(A) dA db \\
&= \int_0^1 E^{c\star}(b, 1) g_b(b) \left(\int_{\bar{A}(b)}^{\infty} A g_A(A) dA \right) db \\
\int_{\bar{A}(b)}^{\infty} A g_A(A) dA &= \frac{\gamma}{\gamma - 1} \bar{A}(b)^{1-\gamma} \\
\implies E^c_{\text{agg}} &= \frac{\gamma}{\gamma - 1} \int_0^1 E^{c\star}(b, 1) \bar{A}(b)^{1-\gamma} dG_b(b) \\
\implies E^c_{\text{agg}} &= \frac{\gamma}{\gamma - 1} \int_0^1 E^{c\star}(b, 1) \left(\frac{f^e(b) s(b)}{V_{\text{entry}}(b, 1)} \right)^{1-\gamma} dG_b(b)
\end{aligned}
$$

Appendix 1.B Additional Empirical Results

1.B.1 Variation

Sample	Fixed Effects	N	adj. R^2	BIC
Full	Industry (2)	600,417	0.14	1816950
Full	Industry (2) by Year	600,417	0.15	1818253
Full	Industry (4)	600,417	0.27	1721316
Full	Industry (4) by Year	600,417	0.28	1768868
Single Product	Product	281,380	0.35	810365
Single Product	Product by Year	281,380	0.37	909941
Full Geo	District (5)	600,404	0.03	1893801
Full Geo	District (5) + Industry (4)	600,404	0.28	1714386
Full Geo	District (5) by Year	600,404	0.04	1963530
Full	Plant	600,417	0.87	1537935

Appendix Table 1.B.1: Adjusted R^2, number of observations, and BIC for regressions of energy mix $\log E_{it}^d/E_{it}^c$ on different sets of fixed effects. For the models with district effects and product effects, I restrict the sample to districts with at least 5 plants, and products produced by at least 5 plants The BIC is only comparable within a given sample due to the different numbers of observations. It penalizes models with more parameters, i.e., more fixed effects. In the full sample, the plant fixed effect model has the lowest BIC, suggesting that its high explanatory power is not due to overfitting.

1.B.2 Factor demand elasticity

1.B.2.1 First stage

The first stage results for the baseline specification are shown in table 1.B.2. The coefficient estimates are strong, and statistically significant.

1.B.2.2 Labor and materials inputs

For illustration, I also estimate the factor demand elasticities for labor and materials inputs. Labor is defined as the number of employees at the plant. Materials are given by the deflated expenditure on intermediate inputs. The results are shown in table 1.B.3.

Dependent Variables:	$\Delta_1 \log Y$ (1)	$\Delta_2 \log Y$ (2)
shock[1]	0.2239	
	(0.0185)	
shock[2]		0.2026
		(0.0168)
Fixed-effects		
Year by 2-digit Industry	Yes	Yes
Year by District	Yes	Yes
Fit statistics		
Observations	275,455	249,927
Wald (1st stage)	146.15	144.65
R^2	0.10	0.11
within-R^2	0.0036	0.0036

Appendix Table 1.B.2: Regression results for the first stage of the 2SLS estimation of equation (1.3), estimated on the dirty energy (E^d) sample. Results are virtually identical for the clean energy (E^c) sample, except for the number of observations. Standard errors are clustered at the year and 4-digit industry level.

Dependent Variables:	$\Delta_1 \log L$ (1)	$\Delta_2 \log L$ (2)	$\Delta_1 \log M$ (3)	$\Delta_2 \log M$ (4)
$\Delta_1 \log Y$	0.3170 (0.0232)		1.065 (0.0986)	
$\Delta_2 \log Y$		0.4028 (0.0249)		1.031 (0.0987)
Fixed-effects				
Year by 2-digit Industry	Yes	Yes	Yes	Yes
Year by District	Yes	Yes	Yes	Yes
Fit statistics				
Observations	459,057	373,174	146,665	106,387
Wald (1st stage)	246.80	293.70	145.75	113.87

Appendix Table 1.B.3: 2SLS results for the factor demand elasticity of the number of employees, L, and deflated expenditure on intermediate inputs, M (regression equation (1.3)). Standard errors are clustered at the 4-digit industry level by year level. The responsiveness of the number of employees is between that of clean and dirty energy. Employees in Germany enjoy strong protections, so the firms may face adjustment costs in their labor input. Materials and intermediate inputs respond one-for-one with output.

Dependent Variables:	$\Delta_1 \log E^d$ (1)	$\Delta_1 \log E^c$ (2)	$\Delta_2 \log E^d$ (3)	$\Delta_2 \log E^c$ (4)
$\Delta_1 \log Y \times \mathbf{1}[\text{shock}^1 < 0]$	0.2694	0.6723		
	(0.0803)	(0.0380)		
$\Delta_1 \log Y \times \mathbf{1}[\text{shock}^1 > 0]$	0.0703	0.5732		
	(0.0813)	(0.0482)		
$\Delta_2 \log Y \times \mathbf{1}[\text{shock}^2 < 0]$			0.1889	0.5920
			(0.0759)	(0.0475)
$\Delta_2 \log Y \times \mathbf{1}[\text{shock}^2 > 0]$			0.2776	0.6816
			(0.0927)	(0.0522)
Fixed-effects				
Year by 2-digit Industry	Yes	Yes	Yes	Yes
Year by District	Yes	Yes	Yes	Yes
Fit statistics				
Observations	275,455	304,305	249,927	276,385
Wald (neg. shock)	31.52	34.22	33.26	33.84
Wald (pos. shock)	50.61	52.58	40.06	47.77

Appendix Table 1.B.4: 2SLS regression results for the factor demand elasticity, interacted with the sign of the shift-share shock. The output growth term is interacted with an indicator for whether the leave-one-out industry output shock is negative ($\text{shock}^k < 0$) or positive ($\text{shock}^k > 0$), where k denotes the horizon in years. Both interaction terms are instrumented with the corresponding interactions of the leave-one-out industry output. Standard errors are clustered at the plant and 4-digit industry by year level. Fixed effects for year by 2-digit industry and year by district are included.

1.B.2.3 Asymmetric response by shock sign

Table 1.B.4 shows the factor demand elasticities estimated separately for positive and negative output shocks, by interacting the instrument and the endogenous regressor with an indicator for the sign of the shock. The response of clean energy is generally symmetric. There is an asymmetry in the response of dirty energy at the 1-year horizon, which reverses at the 2-year horizon. The estimates are noisier than the main specification, but are broadly consistent.

Dependent Variables:	$\Delta_1 \log E^d$ (1)	$\Delta_1 \log E^c$ (2)	$\Delta_2 \log E^d$ (3)	$\Delta_2 \log E^c$ (4)
$\Delta_1 \log Y$	0.1788	0.6280		
	(0.0629)	(0.0318)		
$(\Delta_1 \log Y)^2$	-0.6914	-0.3596		
	(0.3566)	(0.2146)		
$\Delta_2 \log Y$			0.2391	0.6323
			(0.0580)	(0.0352)
$(\Delta_2 \log Y)^2$			-0.0547	0.1462
			(0.1790)	(0.1177)
Fixed-effects				
Year by 2-digit Industry	Yes	Yes	Yes	Yes
Year by District	Yes	Yes	Yes	Yes
Fit statistics				
Observations	275,455	304,305	249,927	276,385
Wald ($\Delta \log Y$)	74.71	78.08	75.10	82.34
Wald ($(\Delta \log Y)^2$)	26.60	25.31	46.28	48.14

Appendix Table 1.B.5: 2SLS regression results for the factor demand elasticity with a squared output growth term. The squared term tests whether the elasticity varies with the magnitude of the output shock. There are two endogenous regressors ($\Delta \log Y$ and $(\Delta \log Y)^2$), instrumented with the leave-one-out industry output shock and its square (shock^k and $(\text{shock}^k)^2$, where k denotes the horizon in years). Standard errors are clustered at the plant and 4-digit industry by year level. Fixed effects for year by 2-digit industry and year by district are included.

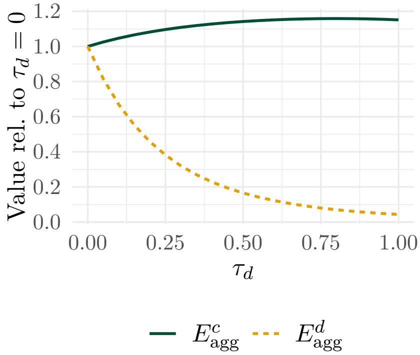
1.B.2.4 Nonlinearity in output growth

To test for non-linear effects, I estimate a regression with the square of the change in output, and the square of the shock as its instrument. Table 1.B.5 presents the results. The quadratic terms are not statistically significant, and the linear terms are in line with their baseline estimates.

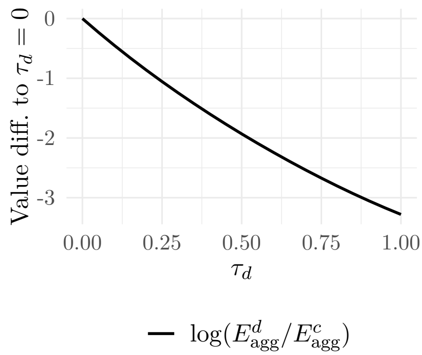
Appendix 1.C Additional Policy Experiments

1.C.1 Tax on dirty energy use

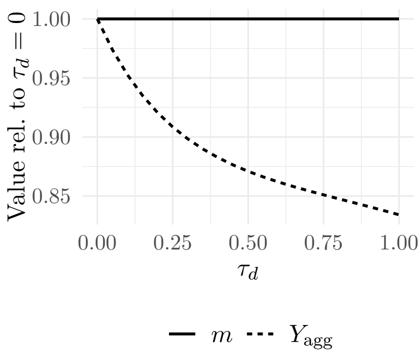
Figure 1.C.1 shows the results of the tax on dirty energy use policy experiment keeping the distribution fixed.



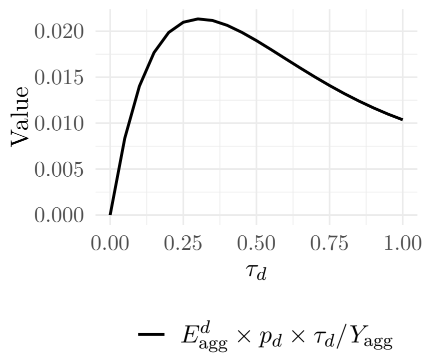
(a) Use of clean and dirty energy.



(b) Energy mix.



(c) Output and mass of entrants.



(d) Tax revenue as fraction of output.

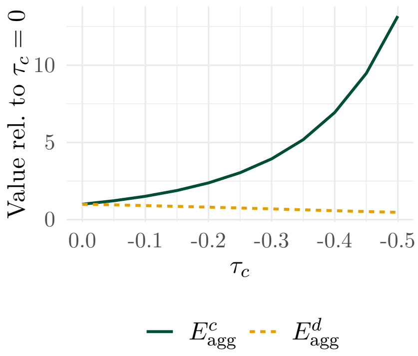
Appendix Figure 1.C.1: Results of tax on dirty energy use policy experiment, keeping the distribution fixed at baseline.

1.C.2 Subsidy on clean energy use

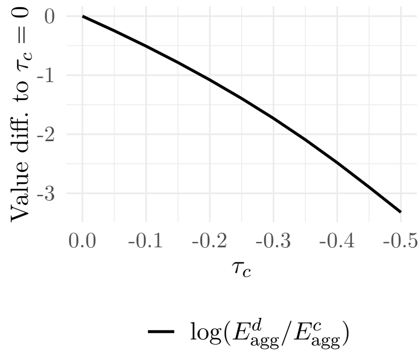
Consider a subsidy on clean energy use, such that the final price of clean energy is

$$p_c^{\text{sub}} = p_c(1 + \tau_c). \tag{1.C.1}$$

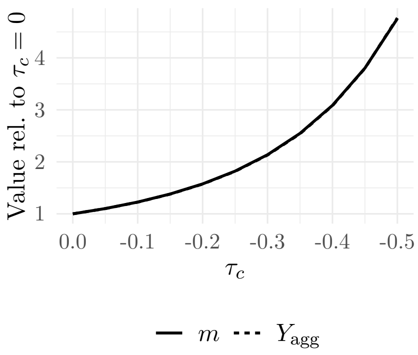
The subsidy is permanent and will be received by all plants. I consider values $\tau_c \in [-0.5, 0]$, up to a 50% subsidy on the price of clean energy. As in section 1.6.3, I solve the model under the new price and compute aggregates given the new equilibrium distribution of plants. Figure 1.C.2 shows the results of the subsidy on clean energy use policy experiment.



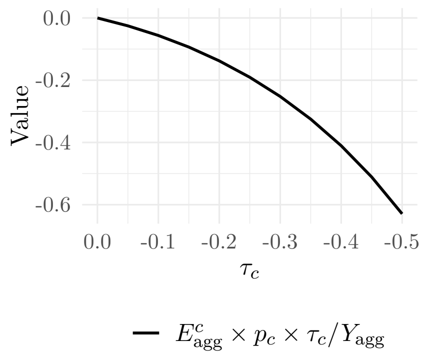
(a) Use of clean and dirty energy.



(b) Energy mix.



(c) Total output and mass of entering plants.



(d) Subsidy spending as fraction of output.

Appendix Figure 1.C.2: Results of subsidy on clean energy use policy experiment, allowing for compositional changes.

Chapter 2

Long-run elasticities from short-run variation

I want to thank my advisors Per Krusell and Joshua Weiss for guidance and feedback. Anton Arbman Hansing, and participants of the IIES Macro Group and Informal Macro Discussion buddies provided helpful comments.

2.1 Introduction

Elasticity estimates are central to applied and quantitative economics. They describe the response of an economy or agents to shocks, and are thus directly informative for counterfactual analysis and optimal policy. A common finding in the empirical literature is that for a persistent price change, an elasticity estimate increases with the time horizon before eventually levelling off. There is a distinction between a "short-" and a "long-run" elasticity, and the difference between them is typically attributed to adjustment frictions. Both time-horizons matter in welfare analysis: the short-run elasticity governs the speed of the transition, while the long-run value describes the target of the transition.

In general, the estimation of elasticities is difficult due to endogeneity problems; we can not simply regress prices on quantities. There are two main estimation approaches: Structural estimation with context-specific models, or reduced-form instrumental variable approaches. The latter generally relies on exogenous variation in the price. This implies that the short-run elasticity can be estimated in a larger number of settings: The difficulty of finding plausibly exogenous variation in prices increases with the time horizon and persistence of the shock. Fewer natural experiments with exogenous and persistent price changes exist, than there are relevant long-run elasticities.

In this paper, I propose an alternative estimation strategy using transitory time-series variation instead. I develop a dynamic model of the choice of the ratio of inputs in response to shocks to the relative price, subject to an adjustment cost. When specifying the model as a second-order approximation of a constant elasticity of substitution production function, and a quadratic adjustment cost, the short-run elasticity is the product of an attenuation factor due to the adjustment cost, and the long-run response. Crucially, this factor has a closed-form expression in three parameters: the persistence of the input ratio, the persistence of the relative price (both can be estimated from the data), and the discount factor (which can be calibrated to standard values).

There are no additional data requirements to implement this approach; it only requires running two regressions in addition to the short-run elasticity estimation, and choosing a value for the discount factor. Then the long-run elasticity is given as the short-run estimate divided by the attenuation factor.

I focus on a constant elasticity of substitution setting. For simplicity of exposition, I develop the model under the assumption of the relative price being the only stochastic variable. This assumption is most applicable in this choice of ratios, since we can abstract from productivity or demand shocks. In principle, when those shocks can be controlled for, the model generalizes to other isoelastic relationships (supply, demand, trade, ...). The generic use-case for the method is a panel of units in a stationary price regime. A policy counterfactual may ask how the economy would react to a permanent change in the level of the price, for example through a tax. The method then uses the response to transitory price shocks under the assumption of a specific adjustment structure to infer the elasticity, and thus how the economy would respond to permanent price changes in the long-run.

With generic adjustment frictions it is impossible to separate the elasticity from the adjustment friction: Is the short-run response small because the long-run elasticity is low, or because the adjustment friction is large? I show that under the approximated model in this paper, the persistence in the choice variable is a sufficient statistic for the frictions, which allows the separation.

The canonical approach to disentangle the elasticity and friction is to use cross-sectional variation: the between-estimator in panels. The implicit assumption is that all panel units face the same isoelastic relationship, and they are close to their "steady-state" values on average. The between-estimator then aims to identify that long-run relationship. In this paper, I show that for the between-estimator to be consistent, all variation in the price level between units must be exogenous. If variation in prices is endogenous (for example through market power), the between-estimator is inconsistent. The inconsistency can be severe, depending on the ratio of the exogenous to endogenous variation. For an IV strategy to alleviate the problem, we require exogenous cross-sectional variation in the instrument.

The correction approach developed in this paper is complementary to the between-estimator in panels. They rely on different sets of assumptions: The between estimator imposes less structure on the agent problem, but requires exogenous variation in the level of the price. The correction approach relies only on exogeneity in the changes of the price, but imposes more structure on the behavior of the agents.

I apply the estimation approach to the setting of Peter and Ruane (2025), who study intermediate input substitution in Indian manufacturing firms. Using the unexpected Indian trade liberalization in the 1990s as a natural experiment, the authors estimate the long-run elasticity over a 7-year horizon. They estimate the short-run elasticity with a complementary shift-share price IV. Using the same data, I estimate the relevant persistence parameters, and use the discount factor from their paper. The estimated correction factor implies that the long-run elasticity of substitution is 2 times larger than the short-run estimate.

Structure The paper proceeds as follows. Section 2.2 presents the model, and derives the correction factor. How to estimate and implement the correction is described in Section 2.3. In Section 2.4, I present the problem of the between-estimator under endogenous prices. I apply my correction method in the Setting of Peter and Ruane (2025) in Section 2.5. Section 2.6 presents limitations of the approach, before Section 2.7 concludes.

Literature Economists have long recognized and observed that individual agents and whole economies do not immediately respond to a persistent shock, but that there is an adjustment over time. Samuelson (1947) introduces, and Milgrom and Roberts (1996) generalize the so-called LeChatelier principle. The principle formalizes that the short-run response of a factor of production to a change in its price may be limited by binding constraints, such as another predetermined factor. For a persistent price change, these constraints tend to become slack over time due to adjustments, such that the long-run response is weakly larger than the short-run response.

The literature has since proposed different microfoundations for such "binding constraints": Atkeson and Kehoe (1999) propose a putty–clay model, in which the capital–energy ratio is fixed after the capital is installed. For the capital–energy ratio to change, the existing capital stock must be replaced over time. León-Ledesma and Satchi (2019) develop a model of factor-augmenting technical change, in which the production technology exhibits a certain elasticity of substitution between inputs in the short run, and a weakly larger elasticity for the long run, due to the possibility to develop factor-augmenting technology. Liu and Tsyvinski (2024) develop an input–output model with endogenous network formation. Agents incur a cost to adjust their network, which introduces a dynamically increasing response of the aggregate to a shock.

While the credible estimation of long-run elasticities is difficult, notable studies exist that exploit natural experiments to directly document the difference between the short- and long-run responses. Chirinko (2008) provides a review of estimates of the capital–labor elasticity of substitution, with short-run estimates around 0.25 and long-run estimates around 0.60. Boehm et al. (2023) show that Armington trade elasticities increase in magnitude from -0.76 after 1 year, to -2.1 after 10 years. Deryugina et al. (2020) estimate the elasticity of residential electricity demand as -0.09 after 6 months and -0.30 after 2 years. Peter and Ruane (2025) show that the elasticity of substitution between intermediate inputs increases from 0.52 at the 1-year horizon to 2.47 after 7 years.

The alternative to natural experiments are structural models. Chirinko and Mallick (2017) employ a low-pass filter to extract the long-run correlation between capital and labor. Apostolakis (1990) and Pindyck and Rotemberg (1983) study the own-price elasticity of energy in a model with adjustment costs to capital. Topel and Rosen (1988) estimate the housing supply elasticity in a model with forward-looking investors.

The interpretation of the between-estimator as the long-run response is well-established in the literature. Kuh (1959) is an early reference to explicitly state the connection. Baltagi and Griffin (1984) numerically compare different estimators, and show that the within-estimator corresponds to the short-run response, while the between-estimator corresponds to the long-run response under a distributed-lag data-generating process. Stern (2012) explicitly notes the larger estimates based cross-sectional variation in interfuel substitution elasticities, and reiterates the long-run interpretation. Pirotte (1999) formally shows that the static between-estimator converges in probability to the dynamic long-run coefficient.

The autoregressive distributed-lag (ARDL)/error-correction model (ECM) literature recovers the long-run multiplier from such dynamic regressions (Pesaran & Shin, 1999; Pesaran et al., 2001). The optimal policy derived in this paper takes the form of an ARDL(1,0) model, and can be written as an ECM. Under stationary prices, however, the ARDL long-run multiplier captures the behavioral steady-state gain rather than the structural elasticity: forward-looking agents who expect prices to mean-revert dampen their response, compressing the multiplier below the true elasticity. The correction developed in this paper provides the additional step of undoing this dampening, thereby recovering the structural elasticity from transitory price variation. Section 2.2.8 develops the connection.

There are several papers with models closely related to this paper's, but asking different questions. Sargent (1978) and Kennan (1979) both develop dynamic linear-quadratic models with quadratic adjustment costs and rational expectations. Both recover the structural elasticity via Euler-equation estimation. The correction formula provides a closed-form shortcut: the same identification without structural estimation.

A conceptually related approach is Chetty (2012). Under a model of fixed cost of adjustment, he derives bounds on the labor supply elasticity of individuals, based on the (lack of) response to a given tax change.

2.2 Model

This section develops a dynamic model of input choice under adjustment frictions. The leading example is a firm choosing its input mix in response to relative prices, where the elasticity of substitution governs the static response. Adjustment frictions cause a partial response, attenuating the estimated elasticity. The approximated model yields a linear policy rule, the coefficients of which depend on observable persistence parameters and a calibrated discount factor, but not on the structural cost parameters that generate the friction. These estimated persistence parameters can then be used to recover the long-run elasticity from the short-run estimate.

2.2.1 The Static Problem

A firm produces output $\bar{Y}$ using two inputs with a constant elasticity of substitution (CES) production function with constant returns to scale:

$$F(X_1, X_2) = \left(\tilde{\gamma}^{1/\sigma} X_1^{\frac{\sigma-1}{\sigma}} + (1-\tilde{\gamma})^{1/\sigma} X_2^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$$

The firm minimizes the cost of producing $\bar{Y}$:

$$\min_{X_1, X_2} \left\{ P_1 X_1 + P_2 X_2 \,\middle|\, F(X_1, X_2) = \bar{Y} \right\}$$

With constant returns to scale, the output constraint pins down both input levels from the input ratio. Define $x := \log(X_2/X_1)$ and $p := \log(P_1/P_2)$. Both X_1 and X_2 are uniquely determined by x and $\bar{Y}$, so the cost minimization reduces to choosing the scalar x. The first-order condition yields the optimal input ratio:

$$x^* = \sigma p + \gamma, \qquad \gamma := \log \frac{1-\tilde{\gamma}}{\tilde{\gamma}} \tag{2.1}$$

The elasticity of substitution σ governs the log-linear response of the input ratio to the relative price.

The CES cost function, expressed as a function of the scalar x, separates into a scale factor and a relative unit cost (appendix 2.A.1):

$$C(x; p, \bar{Y}) = P_2 \bar{Y} c(x; p) \tag{2.2}$$

where $c(x; p)$ is the relative unit cost function. Without loss of generality, let input 2 be the numéraire ($P_2 = 1$). Then c is the unit cost function, and $C = \bar{Y} c$.

2.2.2 Adjustment Frictions and the Dynamic Problem

Suppose the firm faces an adjustment cost for changing its input mix, $\Phi(x, x_{-1}; \bar{Y})$. Assume the function is increasing and convex in the difference between last period's and today's values, $x - x_{-1}$, and satisfies $\Phi(x, x) = 0$. It scales linearly with $\bar{Y}$, such that we can write $\Phi(x, x_{-1}; \bar{Y}) = \bar{Y} \phi(x, x_{-1})$.

The dynamic problem of the firm is then to minimize the expected discounted sum of production costs plus adjustment costs:

$$\min_{\{x_t\}} \mathbb{E} \sum_{t=0}^{\infty} \beta^t \left[c(x_t; p_t) + \phi(x_t, x_{t-1}) \right] \tag{2.3}$$

Since both $C = \bar{Y} c$ and $\Phi = \bar{Y} \phi$ carry the common factor $\bar{Y}$, it divides out of every term in the objective, reducing the problem to the unit-cost form shown. For each t, the minimized cost $c(x^*(p_t); p_t)$ depends only on the price realization and is not a choice variable. In the infinite-horizon problem, subtracting this term from each period's cost does not change the optimal policy: the firm's choice of $\{x_t\}$ is the same whether it minimizes total cost or total excess cost,

$$\min_{\{x_t\}} \mathbb{E} \sum_{t=0}^{\infty} \beta^t \left[c(x_t; p_t) - c(x^*(p_t); p_t) + \phi(x_t, x_{t-1}) \right],$$

where $x^*(p) = \sigma p + \gamma$. The problem is therefore equivalent to minimizing the excess cost of deviating from the static optimum at each t, plus the cost of adjustment. Assuming the log-relative price p follows a first-order Markov process, the problem has the following recursive Bellman equation:

$$V(x_{-1}; p) = \min_x \left\{ c(x; p) - c(x^*(p); p) + \phi(x, x_{-1}) + \beta \mathbb{E}[V(x; p') \mid p] \right\}. \tag{2.4}$$

2.2.3 The LQ Approximation

To derive the closed-form optimal policy, I transform the problem into a problem of the linear-quadratic (LQ) class of Bellman equations. For that, I need two approximations and one functional form assumption.

Second-order approximation. Taylor-expanding $c(x; p)$ around $x^*(p)$, the first-order term vanishes because $x^*(p)$ is defined as the minimizer of $c(\cdot; p)$,

$$c(x; p) \approx c(x^*(p); p) + \underbrace{c'(x^*(p); p)}_{=0}(x - x^*(p)) + \frac{1}{2} c''(x^*(p); p)(x - x^*(p))^2.$$

Subtracting the level term from both sides yields an approximation of the excess cost:

$$c(x; p) - c(x^*(p); p) \approx \frac{1}{2} c''(x^*(p); p)(x - x^*(p))^2.$$

The curvature $c''(x^*(p); p)$ is state-dependent, as it varies with p through the cost shares. For the CES case, $c''(x^*(p); p) = c^* s_1 s_2 / \sigma$, where c^* is the minimized unit cost and s_i are cost shares at the optimum (appendix 2.A.2). This precludes a closed-form solution, so I introduce the following additional approximation.

Constant-curvature approximation. Evaluate the curvature at the unconditional mean price:

$$\kappa := c''(x^*(\bar{p})), \qquad \bar{p} = \mu = \mathbb{E}[p_t]$$

This is an additional approximation beyond the second-order expansion. The gain is a linear policy rule and closed-form correction. The cost is approximation error when prices are far from $\bar{p}$. The approximation error is of the same order as the Taylor truncation error and does not degrade the overall approximation (appendix 2.A.6).

Quadratic adjustment cost. Specify the adjustment cost as:

$$\phi(x, x_{-1}) = \frac{\psi}{2}(x - x_{-1})^2$$

This is a functional form assumption. The parameter ψ measures the effective intensity of adjustment frictions. The quadratic specification is deliberately agnostic about the underlying mechanism: different micro frictions can generate the same observable adjustment speed, so the quadratic form is without loss for the correction.

AR(1) price process. Specify the price process as

$$p' = \rho p + (1 - \rho)\mu + \eta',$$

with $\eta' \sim N(0, \sigma_\eta^2)$. The log-relative price has persistence $\rho \in [0, 1]$, and unconditional mean μ.[1]

LQ Bellman equation. Combining these specifications, the Bellman equation becomes:

$$V(x_{-1}; p) = \min_x \left\{ \frac{\kappa}{2}(x - x^*(p))^2 + \frac{\psi}{2}(x - x_{-1})^2 + \beta \mathbb{E}[V(x; p') \mid p] \right\}, \tag{2.5}$$

$$x^*(p) = \sigma p + \gamma,$$

$$p' = \rho p + (1 - \rho)\mu + \eta', \; \eta' \sim N(0, \sigma_\eta^2).$$

[1]The method generalizes to stationary AR(k) processes, see appendix 2.A.8.

This is a linear-quadratic (LQ) tracking problem (Ljungqvist & Sargent, 2018, ch. 5): the firm tracks a moving target x^*, penalized for deviating from it (curvature κ) and for adjusting toward it (cost ψ).

2.2.4 Optimal Policy

Proposition 2.1 (Optimal Policy)**.** *The solution to the LQ tracking problem* (2.5) *is:*

$$x(x_{-1}; p) = \lambda x_{-1} + (1 - \lambda)\xi\sigma p + (1 - \lambda)\big[\gamma + \sigma\mu(1 - \xi)\big] \qquad (2.6)$$

where $\lambda \in (0, 1)$ *is the persistence parameter and* $\xi \in (0, 1)$ *is the forward-looking dampening factor:*

$$\xi = \frac{1 - \beta\lambda}{1 - \beta\lambda\rho} \qquad (2.7)$$

Proof. By undetermined coefficients on a quadratic value function; see appendix 2.A.3. □

The persistence parameter λ is the stable root of the Euler equation's characteristic polynomial. It has a closed-form expression:

$$\lambda = \frac{(\kappa + \psi + \beta\psi) - \sqrt{(\kappa + \psi + \beta\psi)^2 - 4\beta\psi^2}}{2\beta\psi} \qquad (2.8)$$

As $\psi \to 0$, $\lambda \to 0$ (no adjustment costs, immediate adjustment). As $\psi \to \infty$, $\lambda \to 1$ (prohibitive costs, no adjustment). Higher β decreases λ: more patient agents adjust faster, as they weight future deviations from the target more heavily.

The dampening factor ξ captures forward-looking behavior. It equals one when prices are permanent ($\rho = 1$), so the firm tracks the target fully. When prices are transitory ($\rho < 1$), the firm dampens its response: a price change that will revert is not worth fully tracking.

Corollary 2.1 (Sufficient Statistic)**.** *The structural cost ratio is deter-*

mined by the persistence parameter:

$$\frac{\kappa}{\psi} = \frac{(1-\lambda)(1-\beta\lambda)}{\lambda} \tag{2.9}$$

Consequently, the dampening factor ξ *depends only on* (λ, β, ρ).

The reduction of ξ from a function of $(\kappa, \psi, \lambda, \beta, \rho)$ to (λ, β, ρ) alone is a general property of LQ tracking problems. The stable root λ of the Euler equation creates a one-to-one mapping between λ and κ/ψ (given β), an identity that requires both the within-period first-order condition and the envelope condition (appendix 2.A.3). The implication is a sufficient-statistic interpretation: the observable adjustment speed λ encodes all information about the structural cost parameters needed for the correction. This property would break with (i) multiple adjustment margins or (ii) non-quadratic costs. Under AR(k) prices the sufficient-statistic result is preserved, but the correction formula generalizes: it involves the full companion matrix of the price process rather than the scalar ρ (see appendix 2.A.8).

In steady state at $p = \mu$, the optimal policy gives $\bar{x} = \sigma\mu + \gamma = x^*(\mu)$: the firm sits at its frictionless target when nothing is changing.

2.2.5 Approximation Order

Two alternative routes lead to the same LQ policy rule (2.6): a joint second-order approximation of the objective function around $(\bar{x}, \bar{p})$ (appendix 2.A.4) and a first-order linearization of the exact Euler equation around $(\bar{x}, \bar{p})$ (appendix 2.A.5). Both are perturbations around the same point on the optimum ridge $x^*(p)$ and produce identical policy rules (a general result in perturbation methods, Benigno & Woodford, 2012).

Using perturbation arguments, I show that the policy-rule approximation error is of order $O(\sigma_\eta^2)$ around $\sigma_\eta \to 0$ (appendix 2.A.6).

2.2.6 Short-Run and Long-Run Elasticities

Based on the optimal policy (2.6) there are two natural concepts for the elasticity of substitution at different time horizons.

Definition 2.1 (Short-Run Elasticity)**.** *The short-run elasticity is the impact response of the input ratio to the current price, holding the inherited state fixed:*

$$\sigma_{SR} := \left.\frac{\partial x}{\partial p}\right|_{x_{-1}} = (1-\lambda)\xi\sigma = \frac{(1-\lambda)(1-\beta\lambda)}{1-\beta\lambda\rho}\sigma$$

Two forces attenuate the impact response relative to σ. First, partial adjustment: the firm closes only a fraction $(1-\lambda)$ of the gap to the target each period. Second, forward-looking dampening: when prices are transitory ($\rho < 1$), the firm anticipates mean reversion and does not fully track the current target. Only when prices are permanent ($\rho \to 1$, so $\xi \to 1$) does the second force vanish.

Definition 2.2 (Long-Run Elasticity)**.** *The long-run elasticity is the steady-state response to a permanent shift in the price environment:*

$$\sigma_{LR} := \frac{d\bar{x}}{d\mu} = \sigma$$

In steady state the input ratio is constant, so adjustment costs are zero and the firm sits at its frictionless optimum. Since $\bar{x}(\mu) = \sigma\mu + \gamma$, the steady-state input ratio responds one-for-one with the frictionless target, no attenuation from either force, because neither operates when nothing is changing.

Geometric adjustment trajectory. The transition dynamics yield a testable prediction. If the firm knows a price change Δp is permanent, it adjusts by $(1-\lambda)\sigma\Delta p$ on impact, $\lambda(1-\lambda)\sigma\Delta p$ in the next period, and so on. The cumulative response is:

$$(1-\lambda)\sigma\Delta p\sum_{j=0}^{\infty}\lambda^j = \sigma\Delta p$$

Under the optimal policy, the impulse response profile to a permanent shock therefore follows a geometric decay with parameter λ. This is a testable prediction we could verify by estimating the time-profile of adjustments in an ideal empirical setting.

SR/LR ratio. The ratio of short-run to long-run elasticity summarizes the total attenuation:

$$\frac{\sigma_{SR}}{\sigma_{LR}} = \frac{(1-\lambda)(1-\beta\lambda)}{1-\beta\lambda\rho} \tag{2.10}$$

Table 2.1 reports limiting cases.

Table 2.1: Limiting cases of the SR/LR elasticity ratio

Limiting case	σ_{SR}/σ_{LR}	Interpretation
$\lambda \to 0$ (no frictions)	$\to 1$	Immediate full adjustment
$\lambda \to 1$ (prohibitive frictions)	$\to 0$	No contemporaneous response
$\rho \to 1$ (permanent prices)	$\to (1-\lambda)$	Only partial adjustment attenuates
$\rho \to 0$ (i.i.d. prices)	$\to (1-\lambda)(1-\beta\lambda)$	Both forces attenuate

2.2.7 Attenuation and Correction

The analysis uses time-series (within) variation: the relevant estimating equation regresses the input ratio on the contemporaneous relative price, controlling for firm fixed effects (or equivalently, in first differences).[2]

[2]Cross-sectional (between) variation offers an alternative route to the long-run elasticity when permanent price differences across firms are exogenous. If those differences are endogenous (for example through market power), the between estimator can be severely biased. Section 2.4 formalizes this argument and shows when cross-sectional identification fails.

The estimating equation is:

$$x_{it} = \alpha_i + \hat{\sigma} p_{it} + u_{it}$$

Under the true DGP (2.6), the regressor p_t is correlated with the omitted variable x_{t-1} through price persistence. The ordinary least squares (OLS) coefficient $\hat{\sigma}$ therefore combines the true coefficient on p_t with omitted variable bias (OVB).

Omitted variable bias. The auxiliary regression coefficient of x_{t-1} on p_t in the stationary distribution is:[3]

$$\delta := \frac{\mathrm{Cov}(x_{t-1}, p_t)}{\mathrm{Var}(p_t)} = (1-\lambda)\xi\sigma \cdot \frac{\rho}{1-\lambda\rho}$$

By the OVB formula, $\hat{\sigma} = (1-\lambda)\xi\sigma + \lambda\delta$. Substituting and simplifying yields:

$$\frac{\hat{\sigma}}{\sigma} = \frac{(1-\lambda)(1-\beta\lambda)}{(1-\lambda\rho)(1-\beta\lambda\rho)}. \tag{2.11}$$

Proposition 2.2 (Attenuation Decomposition)**.** *The attenuation of the static OLS estimator decomposes into three factors:*

$$\underbrace{\frac{\hat{\sigma}}{\sigma}}_{\textit{attenuation}} = \underbrace{(1-\lambda)}_{\substack{\textit{adjustment}\\ \textit{friction}}} \cdot \underbrace{\xi}_{\substack{\textit{forward-looking}\\ \textit{dampening}}} \cdot \underbrace{\frac{1}{1-\lambda\rho}}_{\substack{\textit{OVB}\\ \textit{amplification}}}$$

The first two factors attenuate. Adjustment friction $(1-\lambda)$ reflects the mechanical partial response; forward-looking dampening (ξ) reflects the firm's anticipation of price mean reversion. The third factor partially offsets: omitting x_{t-1} amplifies the apparent response because x_{t-1} is positively correlated with p_t through price persistence. When $\rho = 0$ (i.i.d. prices), the OVB factor equals one and $\hat{\sigma} = \sigma_{SR}$. When $\rho > 0$, $\hat{\sigma}$ lies

[3] In demeaned form the policy and price process read $\tilde{x}_t = \lambda\tilde{x}_{t-1} + (1-\lambda)\xi\sigma\tilde{p}_t$ and $\tilde{p}_t = \rho\tilde{p}_{t-1} + \eta_t$. Since η_t is independent of $\tilde{x}_{t-1}$, $\mathrm{Cov}(\tilde{x}_{t-1}, \tilde{p}_t) = \rho\,\mathrm{Cov}(\tilde{x}_{t-1}, \tilde{p}_{t-1})$. Multiplying the policy by $\tilde{p}_t$ and taking expectations gives $\mathrm{Cov}(\tilde{x}_t, \tilde{p}_t)(1-\lambda\rho) = (1-\lambda)\xi\sigma\,\mathrm{Var}(\tilde{p}_t)$. Rearranging and dividing by the variance of the price gives δ.

between σ_{SR} and σ_{LR}.

Correction formula. Inverting (2.11) yields the main result:

$$\sigma = \hat{\sigma} \cdot \frac{(1 - \lambda\rho)(1 - \beta\lambda\rho)}{(1 - \lambda)(1 - \beta\lambda)} \tag{2.12}$$

The correction recovers the structural elasticity from the attenuated estimate using only λ, ρ (both estimated), and β (calibrated). The sufficient statistic property of the LQ policy (Corollary 2.1) implies that the persistence parameter λ encodes all relevant information about the deeper cost parameters (κ, ψ) and no structural estimation is necessary.

2.2.8 Connection to ARDL and Error-Correction Models

The optimal policy (2.6) takes the form of an ARDL(1,0) model: an autoregressive distributed-lag model with one lag of the dependent variable and zero lags of the exogenous variable (Pesaran & Shin, 1999):

$$x = \lambda x_{-1} + bp + c, \qquad b := (1-\lambda)\xi\sigma, \quad c := (1-\lambda)\big[\gamma + \sigma\mu(1-\xi)\big]. \tag{2.13}$$

This paper's LQ model provides a microfoundation of the ARDL formulation, which is typically assumed ad hoc (Pesaran, 1997).

Error-correction representation. Reparameterizing (2.13) in error-correction form gives:

$$\Delta x = x - x_{-1} = -(1-\lambda)\big(x_{-1} - \xi\sigma p - m\big), \qquad m := \gamma + \sigma\mu(1-\xi). \tag{2.14}$$

The adjustment speed is $(1 - \lambda)$, the fraction of the gap closed per period. Crucially, the "equilibrium" target in this error-correction model is $\xi\sigma p + m$, which is *not* the frictionless optimum $\sigma p + \gamma$. The firm corrects toward a dampened target because it rationally expects prices to mean-revert.

ARDL long-run multiplier versus structural elasticity. The ARDL long-run multiplier is:

$$\mu_{\text{LR}} \coloneqq \frac{b}{1-\lambda} = \xi\sigma.$$

This object is the behavioral steady-state gain: the response after the dynamic adjustment has played out, under the agent's belief that shocks are transitory. The structural elasticity σ is the response to a known permanent shift in the price level (Definition 2.2). The wedge between the two is the forward-looking dampening factor $\xi = (1-\beta\lambda)/(1-\beta\lambda\rho)$.

Generally, in the ARDL/ECM literature, the long-run multiplier $b/(1-\lambda)$ is *the* long-run object of interest, and ξ is not defined within the model. The structural decomposition provided by the LQ model allows us to separate the ARDL long-run response into ξ and σ. This distinction becomes relevant for counterfactuals that involve permanent changes (for example changes in the price through a tax).

When the two objects coincide. The ARDL multiplier equals the structural elasticity in two limiting cases: (i) With non-stationary prices ($\rho \to 1 \implies \xi \to 1$) the model becomes a standard ECM or co-integration model, and the long-run multiplier equals the structural elasticity (Engle & Granger, 1987). (ii) With myopic agents ($\beta = 0 \implies \xi = 1$) there is no forward-looking dampening (incorporating this forward-looking behavior in adjustment cost models is a contribution of Sargent, 1978).

The correction formula matters in the intermediate case: stationary prices ($\rho < 1$) and forward-looking agents ($\beta > 0$). It decomposes into two logically distinct steps: (i) the ARDL step, recovering $\xi\sigma$ from $\hat{\sigma}$ by accounting for the omitted lagged dependent variable; and (ii) the forward-looking correction, recovering σ from $\xi\sigma$ by inverting the dampening factor ξ. Step (i) is standard in the ARDL literature; step (ii) is this paper's contribution.[4]

[4]Standard ARDL theory delivers consistent, asymptotically normal inference on $\xi\sigma = b/(1-\lambda)$ (Pesaran & Shin, 1999), whether the forcing variable is $I(0)$ or $I(1)$.

2.3 Implementation

This section describes how to compute the corrected elasticity (2.12) in practice and how to obtain standard errors that account for estimation uncertainty in the correction parameters.

2.3.1 Data Requirements

Data format. The core identification strategy is within-unit time-series variation in prices: the estimating equations exploit how the unit responds to changes in prices over time. This is satisfied by two data structures: In a time-series setting, a single unit is observed over T periods and all variation is temporal. In a panel setting, N units are observed over T periods. The within estimator (i.e., including unit fixed effects (FE), or a first-difference specification) exploit the relevant variation in the time dimension. Fixed effects strip out permanent cross-sectional differences and leave only within-unit price variation, replicating the same identification as in the time-series case.[5]

Permissible heterogeneity. The optimal policy function and thus the model-implied DGP is:

$$x_{it} = \lambda x_{i,t-1} + (1-\lambda)\xi\sigma p_{it} + (1-\lambda)\big[\gamma_i + \sigma\mu(1-\xi)\big]. \qquad (2.15)$$

Reading off each component, the intercept $(1-\lambda)[\gamma_i+\sigma\mu(1-\xi)]$ may vary across units through the firm-specific demand shifter γ_i (or difference in the unconditional mean of the relative price, μ). In a panel, this heterogeneity is fully absorbed by unit fixed effects α_i (or the constant in the time-series case). The persistence λ, the elasticity σ, and the price persistence ρ (which determines ξ) parameters must be homogeneous across

Several reparameterizations make $\xi\sigma$ a direct regression coefficient (Bårdsen, 1989; Bewley, 1979; Wickens & Breusch, 1988), but all identify $\xi\sigma$, not σ. This paper's correction adds the step of inverting ξ, which requires the auxiliary estimate $\hat{\rho}$.

[5]The between estimator in panels requires exogenous variation in prices and can be severely biased when price differences are endogenous. Section 2.4 develops the problem.

units, since the dynamics and price-response coefficients are common to all units in the pooled regression.

2.3.2 Estimation Approaches

Approach 1: Correcting the static estimate.

1. Run the static regression: $x_{it} = \alpha_i + \hat{\sigma} p_{it} + u_{it}$.
2. Run the dynamic regression: $x_{it} = \alpha_i + \hat{\lambda} x_{i,t-1} + b p_{it} + u_{it}$.
3. Estimate price persistence: $p_t = (1 - \hat{\rho})\hat{\mu} + \hat{\rho} p_{t-1} + \eta_t$.
4. Apply the correction formula (2.12) with calibrated β.

Approach 2: Direct extraction from the dynamic regression. The coefficient on p_t in the dynamic regression corresponds to the policy coefficient $(1 - \lambda)\xi\sigma$. Solving for σ:

$$\sigma = \frac{b}{(1-\lambda)} \cdot \frac{1 - \beta\lambda\rho}{1 - \beta\lambda} = \frac{b(1 - \beta\lambda\rho)}{(1-\lambda)(1-\beta\lambda)} \tag{2.16}$$

This avoids the two-step correction by extracting σ directly from the dynamic regression coefficients.

Specification diagnostics. Before applying the correction, practitioners should verify that the data are consistent with the LQ model's restrictions on the ARDL structure. The LQ model predicts an ARDL(1,0) specification: additional lags of p_t have zero coefficients when prices follow an AR(1) process, and additional lags of x_t are absent under quadratic adjustment costs. A significant coefficient on p_{t-1} is consistent with higher-order price dynamics; the AR(k) generalization (appendix 2.A.8) then applies. A significant second lag of x_t signals non-quadratic costs or multiple adjustment margins, in which case the correction does not apply. The bounds test of Pesaran et al. (2001) provides a check that a level relationship exists (i.e., $\lambda < 1$), a necessary precondition for a non-trivial correction factor.

2.3.3 Correction under Instrumental Variables

The correction formula (2.12) is derived for $\hat{\sigma}$ from a static levels regression. The attenuation combines the behavioral short-run response $\sigma_{SR} \equiv (1-\lambda)\xi\sigma$ with an omitted-variable amplification $1/(1-\lambda\rho)$. The amplification arises because the lagged state $x_{i,t-1}$ correlates with p_{it} through shared price history, inflating the slope in the static regression.

A subtlety arises when estimating the elasticity of substitution using an IV specification with exogenous variation in the price, as is commonly done in the literature. Let z_{it} be an instrument for p_{it}. The exclusion restriction then has two parts: the standard condition of zero covariance with the regression error term, and additionally uncorrelatedness with the lagged outcome: $\mathrm{Cov}(z_{it}, x_{i,t-1}) = 0$ (which is equivalent to $\mathrm{Cov}(z_{it}, p_{i,t-j}) = 0$ for $j \geq 1$). Intuitively, the instrument must only contain "new" price information, not price information that is already reflected in past choices of $x_{i,t-j}$.

If the instrument satisfies this condition, the OVB channel of the static regression (Proposition 2.2) is absent, which implies a different correction factor. The intuition becomes clear through the MA(∞) representation of the optimal policy,

$$x_t = (1-\lambda)\xi\sigma \sum_{j\geq 0} \lambda^j p_{t-j} + \text{const.}$$

If the identifying variation acts only at $j = 0$, the projection becomes

$$\operatorname{plim} \hat{\sigma}^{\mathrm{IV}} = \frac{\mathrm{Cov}(x_{it}, z_{it})}{\mathrm{Cov}(p_{it}, z_{it})} = (1-\lambda)\xi\sigma \equiv \sigma_{SR}. \tag{2.17}$$

Recovering σ from σ_{SR} uses only the forward-looking dampening factor:

$$\sigma = \hat{\sigma}^{\mathrm{IV}} \frac{1-\beta\lambda\rho}{(1-\lambda)(1-\beta\lambda)}. \tag{2.18}$$

Relative to (2.12), the factor $(1-\lambda\rho)$ drops out: the IV correction factor is larger than its levels counterpart, but applied to a correspondingly smaller estimate, and recovers the same σ under correct specification.[6]

2.3.4 Standard Errors

The corrected estimate inherits estimation error from $\hat{\sigma}$ (or $\hat{b}$), $\hat{\lambda}$, and $\hat{\rho}$. The delta method provides analytical standard errors.

Delta method. For Approach 1, let $\theta = (\hat{\sigma}, \hat{\lambda}, \hat{\rho})'$ with covariance matrix Σ_θ, and define:

$$g(\theta) = \hat{\sigma}\frac{(1-\hat{\lambda}\hat{\rho})(1-\beta\hat{\lambda}\hat{\rho})}{(1-\hat{\lambda})(1-\beta\hat{\lambda})}.$$

The variance of the corrected estimate is:

$$\text{Var}(\hat{\sigma}_{\text{corr}}) \approx \nabla g(\theta)'\Sigma_\theta \nabla g(\theta). \tag{2.19}$$

For the first estimation approach, a bootstrap procedure can be used, since there are no estimates of the covariance between the estimators.

2.4 Cross-Sectional Variation and Long-Run Elasticity

Before applying the method in an empirical setting, I turn to the canonical reduced-form estimation approach for long-run elasticity estimation: the between-estimator. The intuition behind this estimator is that if there are sufficiently persistent cross-sectional price differences *between* units of observation, then the differences in the average outcomes between these units inform the elasticity. Kuh (1959) gives an early argument for between-unit estimation on these grounds; Pirotte (1999) provides the formal result, proving that the between estimator of a static panel model converges to the long-run transfer function of the underlying dynamic

[6] The same argument applies to a first-difference specification of Δx_{it} on Δp_{it} with instrument Δz_{it}.

model. In this section, I describe how this strategy breaks down in the presence of endogenous prices in the form of market power.

A natural source of permanent price differences across firms is market power in input markets. If a firm's purchases are large relative to the supply side, buying more of input 2 drives up P_2 (monopsony), lowering the observed relative price $p = \log(P_1/P_2)$. The observed price then depends on the firm's own input choice, introducing a simultaneity.

To make this precise, consider the symmetric case in which both inputs face the same inverse supply elasticity $\omega \geq 0$: input j's price is $P_{j,i} = P_j^0 \cdot X_{j,i}^{\omega}$, where P_j^0 is the exogenous component. The firm's shadow price for optimization is the marginal expenditure $(1 + \omega P_{j,i})$, so the CES first-order condition still gives a slope of σ on the observed price ratio. In logs, the observed relative price becomes:

$$p_i^{\text{obs}} = p_i^0 - \omega x_i,$$

where $p_i^0 = \log(P_{1,i}^0/P_{2,i}^0)$ is the exogenous price and $-\omega x_i$ is the endogenous feedback: a firm substituting toward input 2 (higher x_i) raises $P_{2,i}$ relative to $P_{1,i}$, lowering p_i^{obs}. Under quantity discounts ($\omega < 0$), the feedback is reversed.[7]

To isolate the implications, consider the steady-state cross-section (averaging observations over time). Using the notation of Section 2.2, the system is:

$$x_i = \sigma p_i + \gamma_i, \qquad p_i = \mu_i - \omega x_i,$$

where γ_i captures unobserved firm-level heterogeneity in the CES share parameter, μ_i collects the exogenous sources of price variation, and $-\omega$ is the price feedback parameter. Under monopsony ($\omega > 0$), a firm using more of input 2 (due to a lower γ_i) raises P_2, which lowers p. The components μ_i and γ_i are assumed independent.

Solving the system jointly gives reduced forms $x_i = (\sigma\mu_i + \gamma_i)/(1 + \sigma\omega)$

[7]For quantity discounts, we need the additional stability condition $\sigma\omega > -1$. The asymmetric case ($\omega_1 \neq \omega_2$) introduces an additional scale channel; see appendix 2.B.1 for the full derivation.

and $p_i = (\mu_i - \omega\gamma_i)/(1+\sigma\omega)$. The between estimator's probability limit is:

$$\begin{aligned} \operatorname{plim}(\hat{\sigma}_{BE}) &= \frac{\operatorname{Cov}(x_i, p_i)}{\operatorname{Var}(p_i)} \\ &= \sigma \frac{\operatorname{Var}(\mu_i)}{\operatorname{Var}(\mu_i) + \omega^2 \operatorname{Var}(\gamma_i)} - \omega \frac{\operatorname{Var}(\gamma_i)}{\operatorname{Var}(\mu_i) + \omega^2 \operatorname{Var}(\gamma_i)}. \end{aligned} \tag{2.20}$$

If there is both price feedback ($\omega \neq 0$), and heterogeneity in demand ($\operatorname{Var}(\gamma_i) > 0$), the estimator is inconsistent. If prices are exogenous ($\omega = 0$) or all firms share the same demand parameters ($\operatorname{Var}(\gamma_i) = 0$), the between estimator recovers σ regardless of market structure. The direction of the inconsistency is determined by the sign of the feedback: the probability limit lies below σ under monopsony ($\omega > 0$) and above σ under quantity discounts ($\omega < 0$).

The severity of the inconsistency is governed by the signal-to-noise ratio $q := \operatorname{Var}(\mu_i)/\operatorname{Var}(\gamma_i)$. Dividing numerator and denominator of (2.20) by $\operatorname{Var}(\gamma_i)$:

$$\operatorname{plim}(\hat{\sigma}_{BE}) = -\frac{1}{\omega} + \frac{\sigma + 1/\omega}{1 + \omega^2/q}. \tag{2.21}$$

The between estimator interpolates between the uninformative limit $-1/\omega$ (at $q \to 0$) and the truth σ (as $q \to \infty$), with weight $w(q) = q/(q+\omega^2)$ on the informative component. The degenerate case is not a knife-edge but the extreme of a continuum: even moderate demand heterogeneity relative to exogenous price variation (q small relative to ω^2) yields a severely inconsistent estimator.

The limiting case of all cross-sectional price variation being endogenous, $\operatorname{Var}(\mu_i) = 0$, makes the failure most transparent: the formula collapses to $\operatorname{plim}(\hat{\sigma}_{BE}) = -1/\omega$. The estimator is completely uninformative about the elasticity. The intuition is straightforward: all cross-sectional variation in (x_i, p_i) is generated by the latent γ_i. A firm with stronger unobserved demand for input 2 (high γ_i) chooses a high x_i; through monopsony, this raises P_2 and lowers its observed p_i. Both variables are driven by the same underlying heterogeneity, with opposite signs. The regression recovers the ratio of γ_i's reduced-form effects on p_i and x_i, and not the structural elasticity. Moreover, no cross-sectional instrument can rescue identification: since $p_i = -\omega\gamma_i/(1+\sigma\omega)$ in the reduced form, any variable correlated with p_i is necessarily correlated with γ_i, making relevance and exogeneity mutually exclusive.

The within estimator sidesteps the endogeneity problem. Firm fixed effects absorb γ_i, eliminating the demand-heterogeneity channel through which endogenous prices create inconsistency. The contemporaneous simultaneity from ω remains but is addressed by instrumenting prices with exogenous shifters; intuitively, the within estimator requires only the changes in prices to be exogenous, not the levels. The correction derived in this paper then recovers σ from the within estimate, requiring only predetermination of the instrument, which is typically a weaker assumption.

2.5 Empirical Application

This section applies the correction framework to elasticity of substitution estimates for intermediate inputs from Peter and Ruane (2025) to test the performance against a setting in which both a short- and long-run elasticity are credibly identified. The paper produces estimates of both a short- and long-run elasticity of substitution based on two complementary identification strategies. The data underlying the paper is publicly available, which allows me to estimate the relevant persistence parameters directly and test the approach.

2.5.1 Setting

Peter and Ruane (2025) study the elasticity of substitution across intermediate inputs in Indian manufacturing using plant-level data from Annual Survey of Industries (ASI). The unexpected and sudden trade liberalization in India during the 1990s lead to substantial changes of the relative prices of intermediate inputs due to the removal of tariffs.

The authors estimate the long-run elasticity of substitution between intermediate input categories as the response to these arguably exogenous price changes after seven years.

In addition to this permanent level shift of the relative prices, the authors also use transitory variation in the global prices of commodities to estimate the short-run elasticity of substitution.

Their identification strategy instruments plant-level input expenditure shares with global commodity prices constructed from Base pour l'Analyse du Commerce International (BACI) bilateral trade data. The short-run IV estimate (1-year horizon) yields $\sigma_{SR} = 0.52$; a separate long-run estimate (7-year horizon) based on tariff variation implies $\sigma_{LR} = 2.47$.

The gap between these estimates is consistent with the attenuation mechanism developed in Section 2.2: if plants face adjustment frictions when reallocating across inputs, short-run responses understate the long-run elasticity. Because σ_{SR} is identified from a price-innovation instrument, the applicable correction is the IV variant $CF^{\text{IV}}(\lambda, \rho, \beta)$ derived in Section 2.3.3, which recovers the long-run elasticity from the attenuated short-run estimate provided the price process is stationary. The required correction factor to bridge the gap is $CF = 2.47/0.52 \approx 4.8$.

2.5.2 Data

ASI

The Annual Survey of Industries provides plant-level microdata on Indian manufacturing. The analysis uses the census sector (plants with 100 or more workers), which is surveyed annually, for the period 1998–2013. Item-level input expenditures, classified at the ASICC 5-digit level, are aggregated to eight one-digit ASICC categories ($k = 1, \ldots, 7, 9$; category 8, Transport, is excluded). The panel unit is a plant–category pair (i, k), observed annually. For details on the ASI, see Peter and Ruane (2025) and Allcott et al. (2016).[8]

[8] The ASI waves are available at https://microdata.gov.in/.

WPI

The Wholesale Price Index provides national commodity price indices, spliced across two base years (1993–94 and 2004–05), covering 1994–2016. A concordance maps WPI commodities to ASICC 5-digit codes via fuzzy description matching, achieving an 85% expenditure-weighted match rate; unmatched items fall back to the WPI sub-group aggregate. WPI is used to construct plant-specific Törnqvist price indices p^T_{ikt} for deflating expenditure shares.[9]

BACI

The BACI dataset provides bilateral trade flows at the HS96 6-digit level for 1996–2024. Global unit values are constructed as $P^C_{kt} = \sum V / \sum Q$ across all bilateral flows excluding India, so that India's own demand does not affect the price measure (the exclusion restriction for the IV estimation in Peter & Ruane, 2025).[10]

BACI rather than domestic WPI prices are used to estimate ρ because the Peter and Ruane (2025) IV strategy identifies σ_{SR} from global commodity price variation. The structurally relevant ρ is the persistence of the identifying price shocks.

Sample Restrictions

The sample restrictions follow Peter and Ruane (2025). Plant–year–categories are restricted to those with at least one Fally and Sayre (2018) commodity item. I apply the cleaning procedure described in Allcott et al. (2016). I remove outliers as the top and bottom 2.5th percentiles of values.

[9] The WPI panels are available at https://eaindustry.nic.in/.

[10] The BACI panels are available at https://www.cepii.fr/DATA_DOWNLOAD/baci/doc/baci_webpage.html.

2.5.3 Estimating Response Persistence λ

The response variable is the deflated log expenditure share:

$$\tilde{s}_{ikt} = \log\left(\frac{E_{ikt}}{P^T_{ikt} \cdot E^m_{it}}\right) \tag{2.22}$$

where E_{ikt} is plant i's expenditure on input category k in year t, P^T_{ikt} is the plant-specific Törnqvist price index, and E^m_{it} is total material expenditure. Deflation removes the mechanical price component from expenditure shares, isolating real expenditure dynamics.

The persistence λ is estimated via an AR(1) in first differences, instrumented with twice-lagged levels (Anderson & Hsiao, 1982):

$$\Delta\tilde{s}_{ikt} = \lambda\Delta\tilde{s}_{ik,t-1} + \delta_{kt} + \Delta\varepsilon_{ikt}, \qquad \text{instrument: } \tilde{s}_{ik,t-2} \tag{2.23}$$

where δ_{kt} are year-by-category fixed effects (year $\times$ ASICC 1-digit category). First-differencing eliminates plant–category fixed effects and, combined with IV, removes the Nickell bias that afflicts within estimators in short panels. The instrument $\tilde{s}_{ik,t-2}$ also corrects measurement error attenuation: under i.i.d. errors, twice-lagged levels are correlated with $\Delta\tilde{s}_{ik,t-1}$ through mean reversion but uncorrelated with $\Delta\varepsilon_{ikt}$.[11]

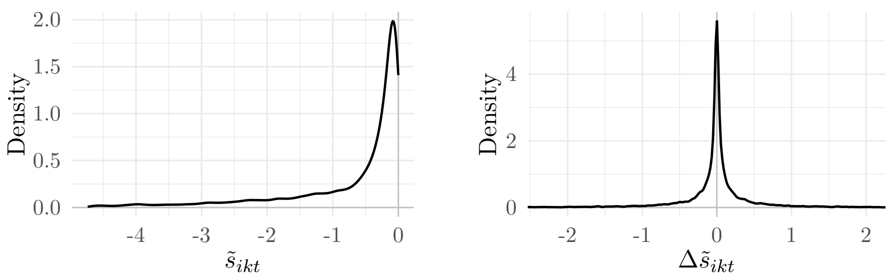
Table 2.2 reports descriptive statistics for the estimation sample; Figure 2.1 shows the distribution of levels and first differences. Table 2.3 reports the estimation results; Figure 2.2 shows the first-stage and reduced-form relationships. The estimate is $\hat{\lambda} = 0.35$ (SE $= 0.25$, two-way clustered at the plant and year level), with a first-stage F-statistic well above conventional thresholds.

[11] The Anderson and Hsiao (1982) estimator is a simple case of the Arellano and Bond (1991) estimator, with only one lag as instrument. It can be implemented as a linear 2SLS regression, and does not require a full non-linear GMM estimation. I choose it here for its simplicity.

Table 2.2: Descriptive statistics: $\hat{\lambda}$ estimation sample.

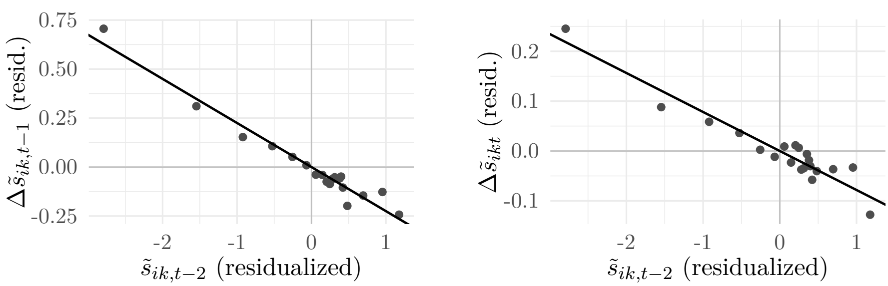
	N	Mean	Median	SD
$\tilde{s}_{ikt}$	16,441	-0.685	-0.241	0.967
$\Delta\tilde{s}_{ikt}$	16,441	-0.014	-0.001	0.605

Notes. Descriptive statistics for the estimation sample used to estimate $\hat{\lambda}$. The sample covers 8,319 plants (9,398 plant–category pairs) in the ASI census sector, 1999–2013, ASICC categories 1–7 and 9, restricted to plant-year-categories containing at least one Fally & Sayre commodity item. A symmetric 2.5% trim on $\tilde{s}_{ikt}$ levels is applied.



(a) Levels: $\tilde{s}_{ikt}$ **(b)** First differences: $\Delta\tilde{s}_{ikt}$

Figure 2.1: Distribution of deflated log expenditure shares ($\hat{\lambda}$ estimation sample). See table 2.2 for summary statistics.



(a) First stage: $\tilde{s}_{ik,t-2} \rightarrow \Delta\tilde{s}_{ik,t-1}$ **(b)** Reduced form: $\tilde{s}_{ik,t-2} \rightarrow \Delta\tilde{s}_{ikt}$

Figure 2.2: Binscatter of first-stage and reduced-form regressions for the estimation of response persistence $\hat{\lambda}$. See table 2.3 for the IV results.

Table 2.3: Estimating response persistence $\hat{\lambda}$.

	$\Delta\tilde{s}_{ikt}$
$\Delta\tilde{s}_{ik,t-1}$	0.3478
	(0.2505)
Year $\times$ ASICC 1-digit fixed effects	Yes
Observations	16,437
First-stage F-statistic	2,181.5

Notes. The table estimates the persistence of deflated log expenditure shares, λ, in an AR(1) model. The regression is Anderson–Hsiao IV: first-differenced AR(1) instrumented with twice-lagged levels $\tilde{s}_{ik,t-2}$ (equation 2.23), with year-by-category fixed effects (year $\times$ ASICC 1-digit category). The year-by-category fixed effects absorb category-level price innovations, serving as a nonparametric control for the relative price variation that would otherwise require a (mismeasured) price regressor. The panel covers plant–category pairs in the ASI census sector, 1999–2013, ASICC categories 1–7 and 9, restricted to plant-year-categories containing at least one Fally & Sayre commodity item. The sample applies a symmetric 2.5% trim on the level of $\tilde{s}_{ikt}$, requiring all three level variables at t, $t-1$, $t-2$ to lie within the trimmed range. Standard errors are two-way clustered at the plant and year level.

2.5.4 Estimating Price Persistence ρ

The price variable is the pairwise log price ratio constructed from BACI global unit values:

$$p_t^{kj} = \log P_{kt}^C - \log P_{jt}^C$$

for all pairs (k, j) of HS6 commodity codes in the Fally and Sayre (2018) list. The pairwise construction differences out common time trends, so that identification comes purely from within-pair variation in global relative prices over time.

The persistence ρ is estimated via an AR(1) in first differences with Anderson and Hsiao (1982) IV:

$$\Delta p_t^{kj} = \rho \Delta p_{t-1}^{kj} + \Delta \eta_t^{kj}, \qquad \text{instrument: } p_{t-2}^{kj} \tag{2.24}$$

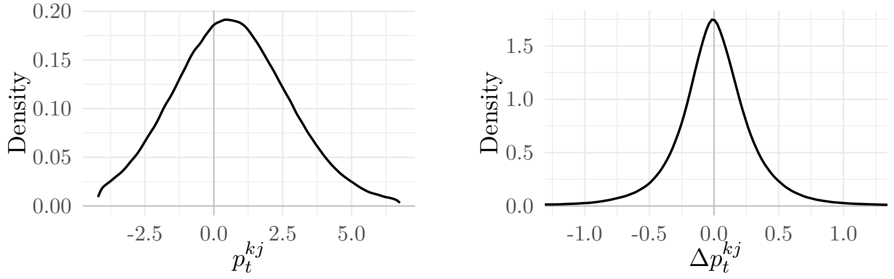
As with λ, first-differencing and IV address the Nickell bias.

Table 2.4 reports descriptive statistics for the price series; Figure 2.3 shows the distribution of levels and first differences. Table 2.5 reports the estimation results; Figure 2.4 shows the first-stage and reduced-form relationships. The estimate is $\hat{\rho} = 0.37$ (SE = 0.24, two-way clustered at the pair and year level), with a first-stage F-statistic well above conventional thresholds. The large sample reflects the combinatorial explosion from forming all pairwise ratios among 492 HS6 codes.

Table 2.4: Descriptive statistics: pairwise log price ratios.

	N	Mean	Median	SD
p_t^{kj}	1,450,365	0.665	0.610	2.034
Δp_t^{kj}	1,450,365	0.004	-0.000	0.404

Notes. Descriptive statistics for the estimation sample used to estimate $\hat{\rho}$. The sample covers 492 HS6 codes (115,164 pairs) over 13 years (1999–2013). Pairwise log price ratios are constructed from BACI global unit values at the HS6 level, restricted to the Fally & Sayre commodity list, with a symmetric 2.5% trim on log price ratio levels. See Figure 2.3 for distribution plots.



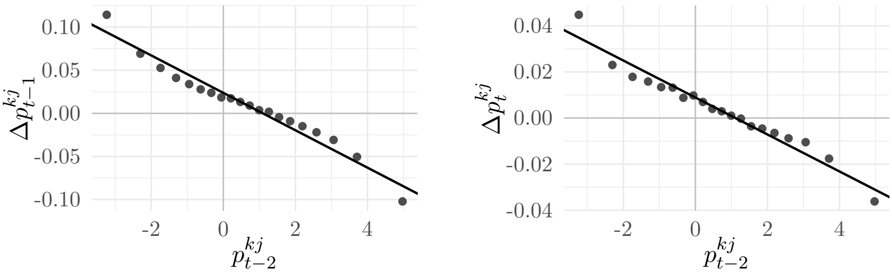
(a) Levels: p_t^{kj} **(b)** First differences: Δp_t^{kj}

Figure 2.3: Distribution of pairwise log price ratios ($\hat{\rho}$ estimation sample). See table 2.4 for summary statistics.

Table 2.5: Estimating price persistence $\hat{\rho}$.

	Δp_t^{kj}
Δp_{t-1}^{kj}	0.3695
	(0.2359)
Observations	1,450,365
First-stage F-statistic	17,152.6

Notes. The table estimates the persistence of pairwise log price ratios, ρ, in an AR(1) model. The regression is Anderson–Hsiao IV: first-differenced AR(1) instrumented with twice-lagged levels p_{t-2}^{kj} (equation 2.24). Pairwise log price ratios are constructed from BACI global unit values at the HS6 level, restricted to the Fally & Sayre commodity list, with a symmetric 2.5% trim on log price ratio levels. Sample: 1999–2013. Standard errors are two-way clustered at the pair and year level.



(a) First stage: $p^{kj}_{t-2} \rightarrow \Delta p^{kj}_{t-1}$ **(b)** Reduced form: $p^{kj}_{t-2} \rightarrow \Delta p^{kj}_{t}$

Figure 2.4: Binscatter of first-stage and reduced-form regressions for the estimation of price persistence $\hat{\rho}$. See table 2.5 for the IV results.

2.5.5 Implied Correction Factor

The IV correction factor (2.18) is computed from the estimates above with a standard annual discount factor:

$$CF^{\text{IV}}(\hat{\lambda}, \hat{\rho}, \beta) = \frac{1 - \beta\hat{\lambda}\hat{\rho}}{(1-\hat{\lambda})(1-\beta\hat{\lambda})}.$$

At $\hat{\lambda} = 0.35$, $\hat{\rho} = 0.37$, and $\beta = 0.95$, this gives $CF^{\text{IV}} \approx 2.01$. The corrected short-run estimate is $\sigma_{SR} \times CF^{\text{IV}} \approx 0.52 \times 2.01 \approx 1.05$. The correction roughly doubles the short-run estimate, but still falls short of the required $CF \approx 4.8$ needed to bridge the full gap to $\sigma_{LR} = 2.47$.

Uncertainty in the corrected estimate propagates from three independent sources: the original short-run estimate σ_{SR}, the estimated response persistence $\hat{\lambda}$, and the estimated price persistence $\hat{\rho}$. Applying the delta method to $\hat{\sigma}_{\text{corr}} = \hat{\sigma}_{SR} \cdot CF^{\text{IV}}(\hat{\lambda}, \hat{\rho}, \beta)$ with $\hat{\sigma}_{SR} = 0.52$ (SE $= 0.28$), $\hat{\lambda} = 0.3478$ (SE $= 0.2505$), and $\hat{\rho} = 0.3695$ (SE $= 0.2359$) at $\beta = 0.95$ yields

$$\hat{\sigma}_{\text{corr}} = 1.05, \qquad \text{SE} = 0.88.$$

The dominant contributions to the standard error are from $\hat{\lambda}$ (0.67) and $\hat{\sigma}_{SR}$ (0.56); price persistence $\hat{\rho}$ contributes only 0.09. The wide confidence interval reflects substantial estimation uncertainty in all three inputs.

The point estimate is insensitive to the choice of $\beta \in \{0.95, 0.97, 0.99\}$.

2.5.6 Lack of Fit and Interpretation

The corrected estimate $\hat{\sigma}_{\text{corr}} \approx 1.05$ is substantially larger than the short-run estimate, confirming that adjustment frictions attenuate the naive estimate. However, it falls short of the long-run estimate of $\sigma_{LR} = 2.47$ reported by Peter and Ruane (2025): a correction factor of $CF^{\text{IV}} \approx 2.01$ explains only a fraction of the gap that would require $CF \approx 4.8$. There is substantial uncertainty around all parameter estimates, including the elasticity estimates from Peter and Ruane (2025). This may point to either statistical noise, or heterogeneity in the parameters, meaning a model misspecification.

A more fundamental misspecification would be if the adjustment cost took the form of a fixed plus convex cost. This implies an "inaction region": agents respond only to sufficiently large shocks. In the data, this generates bunching in the changes of the policy variable around zero. The density for $\Delta\tilde{s}_{ikt}$ in figure 2.1b shows just such concentration around zero, indicating that this may be the problem.

These results fail to verify the method in this setting. Section 2.6 develops general limitations and discusses the model's applicability.

2.6 Discussion

2.6.1 Model (Mis-)Specification and Applicability

The framework requires adjustment costs to be smooth and convex, so that the agent makes an observable adjustment in every period. This limits the set of applications of this method. There exists ample micro-level evidence that the adjustment of two extensively studied factors of production, capital and labor, do not exhibit the required smoothness (Hamermesh & Pfann, 1996). Instead, the adjustment cost is modelled with an additional fixed component, or as irreversible. This introduces inaction bands and so-called (S, s)-policy rules (Caballero, 1999), which fundamentally prevent identification as proposed in this model. Following Chetty (2012), only bounds of the elasticity can be derived with such an adjustment cost structure.

To assess the applicability of this method, the shape of the distribution of the changes to the outcome variable is instructive. A smooth distribution without bunching is consistent with the smooth and convex adjustment cost assumed here. At the micro level, this may hold for variables other than capital and labor, but must be verified.[12] Alternatively, aggregation smooths jumps at the micro level, allowing the method to be applied at sufficiently high levels of aggregation.

2.6.2 Price Process Specification

The correction formula requires the log relative price to follow a stationary process. The baseline case is AR(1) with persistence $\rho < 1$, and it extends to AR(k) for $k > 1$, see appendix 2.A.8.

The model is no longer applicable if prices contain a permanent component, for instance, $p_t = s_t + \tilde{p}_t$, where only $\tilde{p}_t$ is stationary and s_t follows a random walk. This violates stationarity not only for the estimation of ρ, but for the structural model itself: under such a price process, the agent's optimal policy differs from the AR(1) policy derived here, as permanent shocks elicit a full response while only transitory shocks are dampened. The correction formula therefore does not apply, and the direction of the resulting bias depends on the agent's information set and the variance decomposition between the two components. In practice, unit root and KPSS tests on the price series are the natural pre-check before applying the correction. Extending the correction to non-stationary prices requires re-deriving the model under a permanent-transitory price process and is left for future work.

[12]Measurement error complicates this diagnostic, especially for continuous variables.

2.6.3 Parameter Heterogeneity

The correction formula is derived for a representative agent with a common elasticity σ and a common adjustment speed λ. When λ_i varies across units, applying the pooled correction raises two issues.

First, the correction factor is nonlinear in λ, so pooling and correcting do not commute: the pooled corrected estimate does not converge to σ in general. For the range of λ and ρ relevant in practice, this Jensen-type gap points upward, and the pooled correction overshoots.

Second, Pesaran and Smith (1995) show that in heterogeneous dynamic panels the pooled fixed-effect estimator of λ is itself biased upward. Since the correction factor is increasing in λ, this inflates the correction further, compounding the overshoot.

The mean-group estimator—correcting unit by unit and then averaging—is consistent but requires sufficiently long panels (Pesaran & Smith, 1995). In short panels, heterogeneity can be assessed informally by comparing $\hat{\lambda}$ across subgroups such as industries or plant-size quartiles.

2.7 Conclusion

I develop a method to estimate the long-run elasticity of substitution from short-run variation. Using a second-order approximation of a transparent model of input choice subject to an adjustment cost, I show there is a closed-form factor relating the short- and long-run elasticities. The necessary modeling assumptions are (i) an isoelastic relationship, (ii) a stationary relative price, and (iii) a smooth and convex adjustment cost. I show that this factor is a function of three parameters: the persistence of the choice variable, the persistence of the relative price (both of which can be estimated without additional data requirements), and a discount factor (calibrated to standard values). I apply the method in the empirical setting of Peter and Ruane (2025), who estimate both a short- and long-run elasticity of substitution estimate.

The canonical method to estimate long-run elasticities in panel data is the between-estimator. It requires exogenous variation in the level of the relative price between units. I show that the estimator is inconsistent in the presence of endogeneity in prices, for example through monopsony or quantity discounts. In limiting cases, the between-estimator contains no information about the elasticity of interest. An instrument with cross-sectional variation in the relative price can recover the elasticity.

The method proposed in this paper is complementary to the between-estimator. Both aim to recover a long-run elasticity in the absence of exogenous and persistent price shocks, which would directly identify the time-path of the response, and thereby the elasticity. The between-estimator relies on fewer structural assumptions, but requires exogenous cross-sectional variation in prices. The method proposed in this paper requires only exogenous time-series variation in prices, but imposes structure on the behavior of agents.

I derive the method on the setting of substitution elasticities, but it generalizes to other relationships like supply, demand, or trade elasticities, so long as stochastic variables in addition to the price can be controlled for.

References

Allcott, H., Collard-Wexler, A., & O'Connell, S. D. (2016). How do electricity shortages affect industry? Evidence from India. *American Economic Review*, *106*(3), 587–624.

Anderson, T., & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, *18*(1), 47–82.

Apostolakis, B. E. (1990). Energy—capital substitutability/complementarity: The dichotomy. *Energy Economics*, *12*(1), 48–58.

Arellano, M., & Bond, S. (1991). 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.

Atkeson, A., & Kehoe, P. J. (1999). Models of energy use: Putty-putty versus putty-clay. *American Economic Review*, *89*(4), 1028–1043.

Baltagi, B. H., & Griffin, J. M. (1984). Short and long run effects in pooled models. *International Economic Review*, *25*(3), 631–645.

Bårdsen, G. (1989). Estimation of long run coefficients in error correction models. *Oxford Bulletin of Economics & Statistics*, *51*(3).

Benigno, P., & Woodford, M. (2012). Linear-quadratic approximation of optimal policy problems. *Journal of Economic Theory*, *147*(1), 1–42.

Bewley, R. (1979). The direct estimation of the equilibrium response in a linear dynamic model. *Economics Letters*, *3*(4), 357–361.

Boehm, C. E., Levchenko, A. A., & Pandalai-Nayar, N. (2023). The long and short (run) of trade elasticities. *American Economic Review*, *113*(4), 861–905.

Caballero, R. J. (1999). Chapter 12 Aggregate investment. In *Handbook of Macroeconomics* (pp. 813–862, Vol. 1). Elsevier.

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

Chirinko, R. S. (2008). σ: The long and short of it. *Journal of Macroeconomics*, *30*(2), 671–686.

Chirinko, R. S., & Mallick, D. (2017). The substitution elasticity, factor shares, and the low-frequency panel model. *American Economic Journal: Macroeconomics*, *9*(4), 225–253.

Deryugina, T., MacKay, A., & Reif, J. (2020). The long-run dynamics of electricity demand: Evidence from municipal aggregation. *American Economic Journal: Applied Economics*, *12*(1), 86–114.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, *55*(2), 251–276.

Fally, T., & Sayre, J. (2018). *Commodity trade matters* (Working Paper No. 24965). NBER.

Hamermesh, D. S., & Pfann, G. A. (1996). Adjustment costs in factor demand. *Journal of Economic Literature*, *34*(3), 1264–1292.

Kennan, J. (1979). The estimation of partial adjustment models with rational expectations. *Econometrica*, *47*(6), 1441–1455.

Kuh, E. (1959). The validity of cross-sectionally estimated behavior equations in time series applications. *Econometrica*, *27*(2), 197–214.

León-Ledesma, M. A., & Satchi, M. (2019). Appropriate technology and balanced growth. *Review of Economic Studies*, *86*(2), 807–835.

Liu, E., & Tsyvinski, A. (2024). A dynamic model of input–output networks. *The Review of Economic Studies*, *91*(6), 3608–3644.

Ljungqvist, L., & Sargent, T. J. (2018). *Recursive macroeconomic theory* (4th ed.). MIT Press.

Milgrom, P., & Roberts, J. (1996). The LeChatelier principle. *The American Economic Review*, *86*(1), 173–179.

Pesaran, M. H. (1997). The role of economic theory in modelling the long run. *The Economic Journal*, *107*(440), 178–191.

Pesaran, M. H., & Shin, Y. (1999). An autoregressive distributed-lag modelling approach to cointegration analysis. In S. Strøm (Ed.). Cambridge University Press.

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, *16*(3), 289–326.

Pesaran, M., & Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, *68*(1), 79–113.

Peter, A., & Ruane, C. (2025, May). *The aggregate importance of intermediate input substitutability* (Working Paper No. 31233). NBER.

Pindyck, R. S., & Rotemberg, J. J. (1983). Dynamic factor demands and the effects of energy price shocks. *The American Economic Review*, *73*(5), 1066–1079.

Pirotte, A. (1999). Convergence of the static estimation toward the long run effects of dynamic panel data models. *Economics Letters*, *63*(2), 151–158.

Samuelson, P. A. (1947). *Foundations of economic analysis* (1st ed.). Harvard University Press.

Santos, M. S. (2000). Accuracy of numerical solutions using the Euler equation residuals. *Econometrica*, *68*(6), 1377–1402.

Sargent, T. J. (1978). Estimation of dynamic labor demand schedules under rational expectations. *Journal of Political Economy*, *86*(6), 1009–1044.

Stern, D. I. (2012). Interfuel substitution: A meta-analysis. *Journal of Economic Surveys*, *26*(2), 307–331.

Topel, R., & Rosen, S. (1988). Housing investment in the United States. *Journal of Political Economy*, *96*(4), 718–740.

Wickens, M. R., & Breusch, T. S. (1988). Dynamic specification, the long-run and the estimation of transformed regression models. *The Economic Journal*, *98*(390), 189–205.

Appendices

Appendix 2.A Model Derivations

2.A.1 Cost Function in the Scalar Input Ratio

With constant returns to scale, the output constraint $F(X_1, X_2) = \bar{Y}$ and the ratio $X_2/X_1 = \exp(x)$ pin down both input levels. Define the CES aggregator:

$$A(x) := \left(\tilde{\gamma}^{1/\sigma} + (1-\tilde{\gamma})^{1/\sigma} \exp\!\left(\frac{\sigma-1}{\sigma} x\right)\right)^{\frac{\sigma}{\sigma-1}}$$

Then $X_1 = \bar{Y}/A(x)$ and $X_2 = \exp(x)\bar{Y}/A(x)$, and the cost becomes:

$$C(x; p, \bar{Y}) = \frac{\bar{Y}}{A(x)} \left(P_1 + P_2 \exp(x)\right) = \frac{P_2 \bar{Y}}{A(x)} \left(\exp(p) + \exp(x)\right)$$

The second form uses $P_1 = P_2 \exp(p)$ and makes C an explicit function of the log relative price p.

Define the CES weights (component shares of the inner aggregator):

$$\omega_1(x) := \frac{\tilde{\gamma}^{1/\sigma}}{A(x)^{(\sigma-1)/\sigma}}, \qquad \omega_2(x) := \frac{(1-\tilde{\gamma})^{1/\sigma} \exp\!\left(\frac{\sigma-1}{\sigma} x\right)}{A(x)^{(\sigma-1)/\sigma}}$$

and the expenditure shares:

$$s_1(x) := \frac{\exp(p)}{\exp(p) + \exp(x)}, \qquad s_2(x) := \frac{\exp(x)}{\exp(p) + \exp(x)}.$$

Both pairs satisfy $\omega_1 + \omega_2 = 1$ and $s_1 + s_2 = 1$.

Taking logs of $C = \bar{Y}(P_1 + P_2 \exp(x))/A(x)$ gives $\log C = \log \bar{Y} + \log(P_1 + P_2 \exp(x)) - \log A(x)$. Differentiating each component:

$$\frac{d}{dx} \log(P_1 + P_2 \exp(x)) = s_2(x), \qquad \frac{d}{dx} \log A(x) = \omega_2(x)$$

so:

$$C'(x) = C(x)\big[s_2(x) - \omega_2(x)\big].$$

The first derivative is proportional to the gap between the expenditure share and the CES weight. Setting $C'(x^*) = 0$ requires $s_2(x^*) = \omega_2(x^*)$. Cross-multiplying and taking logs recovers $x^* = \sigma p + \gamma$, the static optimum (2.1).

At the optimum, the CES weights equal the cost shares: $\omega_i(x^*) = s_i$. This identity simplifies all higher derivatives of C at x^*.

2.A.2 Cost Function Curvature

Component derivatives. Differentiating the expenditure shares and CES weights:

$$s_2'(x) = s_1(x)s_2(x), \qquad \omega_2'(x) = \frac{\sigma - 1}{\sigma}\omega_1(x)\omega_2(x).$$

Second derivative. Differentiating $C'(x) = C(x)[s_2(x) - \omega_2(x)]$ by the product rule, at x^* the first term vanishes because $C'(x^*) = 0$:

$$C''(x^*) = C(x^*)\left[s_2'(x^*) - \omega_2'(x^*)\right].$$

Using the component derivatives and the identity $\omega_i(x^*) = s_i$:

$$s_2'(x^*) - \omega_2'(x^*) = s_1 s_2 - \frac{\sigma - 1}{\sigma} s_1 s_2 = \frac{s_1 s_2}{\sigma}.$$

Therefore:

$$C''(x^*) = \frac{C^* s_1 s_2}{\sigma}, \tag{2.A.1}$$

where $C^* := C(x^*; p, \bar{Y})$ is the minimized cost.

Unit-cost version. Since $C = P_2 \bar{Y} c$, all derivatives scale by $P_2\bar{Y}$: $C'' = P_2\bar{Y}c''$. Dividing through gives the curvature of the unit cost function:

$$\kappa := c''(x^*) = \frac{c^* s_1 s_2}{\sigma},$$

where $c^* = C^*/(P_2\bar{Y})$ is the minimized unit cost. This is the κ used in the LQ formulation (Section 2.2.3), where the scale factors $P_2\bar{Y}$ cancel from the dynamic problem.

Interpretation. The curvature κ is proportional to $1/\sigma$: higher substitutability flattens the cost function around the optimum, since a deviation from the optimal mix is less costly when close substitutes are available. The factor $s_1 s_2$ peaks at balanced cost shares ($s_1 = s_2 = 1/2$) and approaches zero as either share vanishes—when one input dominates the cost, the ratio of the two inputs matters little. The factor c^* reflects scaling: the percentage curvature $\kappa/c^* = s_1 s_2/\sigma$ is scale-free.

Sensitivity to the input mix. The constant-κ approximation evaluates curvature at the long-run mean $\bar{x}^* = \sigma\bar{p} + \gamma$. The percentage curvature varies along the optimum ridge as:

$$\frac{d}{dx^*} \log \frac{\kappa}{c^*} = \frac{(\sigma - 1)(s_1 - s_2)}{\sigma}.$$

This vanishes at balanced shares and at Cobb-Douglas ($\sigma = 1$, where shares are invariant to the input mix).

2.A.3 LQ Solution via Undetermined Coefficients

Guess a quadratic value function with six coefficients:

$$V(x_{-1}, p) = \frac{1}{2}Ax_{-1}^2 + Bx_{-1}p + \frac{1}{2}Cp^2 + Dx_{-1} + Ep + F.$$

Substitute the guess for the continuation value. Then the first-order condition of the Bellman equation (2.5) with respect to x yields a linear policy:

$$x = \underbrace{\frac{\psi}{\kappa + \psi + \beta A}}_{\lambda} x_{-1} + \underbrace{\frac{\kappa\sigma - \beta B\rho}{\kappa + \psi + \beta A}}_{f} p + \underbrace{\frac{\kappa\gamma - \beta D}{\kappa + \psi + \beta A}}_{g}.$$

Block-triangular structure. The definition of λ implies the zero-sum identity $\kappa\lambda + \psi(\lambda - 1) + \beta A\lambda = 0$, which causes terms proportional to f (or g) to vanish when matching cross-term (or linear) Bellman coefficients. The quadratic coefficients (A, B, C) decouple from the linear and constant terms, allowing sequential solution.

Persistence λ. The Euler equation for x is (combining the first-order condition with the envelope theorem):

$$\kappa(x - x^*(p)) + \psi(x - x_{-1}) - \beta\psi\mathbb{E}[x' - x \mid p] = 0.$$

The homogeneous part has characteristic polynomial $\beta\psi z^2 - (\kappa + \psi + \beta\psi)z + \psi = 0$. The persistence λ is the stable root. Dividing by $\psi\lambda$ and rearranging gives the identity:

$$\frac{\kappa}{\psi} = \frac{(1 - \lambda)(1 - \beta\lambda)}{\lambda}.$$

The factor $(1 - \lambda)/\lambda$ arises from the within-period trade-off between gap-closing (curvature κ) and smoothing (cost ψ). The factor $(1 - \beta\lambda)$ is the forward-looking correction from the envelope condition: the firm anticipates that today's x becomes tomorrow's inherited state, incurring future adjustment costs. In a myopic problem ($\beta = 0$), the identity reduces to $\kappa/\psi = (1 - \lambda)/\lambda$.

Price coefficient f. Matching the $x_{-1}p$ coefficient on both sides of the Bellman equation, the terms proportional to f vanish by the zero-sum identity, leaving:

$$B = \frac{-\kappa\lambda\sigma}{1 - \beta\lambda\rho} \tag{2.A.2}$$

Substituting into the expression for f and using $\kappa/\psi = (1 - \lambda)(1 - \beta\lambda)/\lambda$:

$$f = (1 - \lambda)\xi\sigma, \qquad \xi = \frac{1 - \beta\lambda}{1 - \beta\lambda\rho},$$

confirming (2.6)–(2.7).

Intercept g. Matching the x_{-1} coefficient, the terms proportional to g again vanish. In the zero-mean case ($\mu = 0$):

$$D = \frac{-\kappa\lambda\gamma}{1 - \beta\lambda}. \tag{2.A.3}$$

The equations for B and D have identical algebraic form: B has σ in the numerator and $1 - \beta\lambda\rho$ in the denominator, while D has γ and $1 - \beta\lambda$. The constant γ enters as if it had persistence $\rho_{\text{eff}} = 1$, it is a permanent feature of the technology.

Substituting and simplifying gives $g = (1 - \lambda)\gamma$ at $\mu = 0$. For $\mu \neq 0$, only the intercept changes: the price process $p' = (1 - \rho)\mu + \rho p + \eta'$ introduces additional μ-dependent terms in the linear Bellman coefficients, but the quadratic coefficients—and hence λ and f—are unaffected. The modified intercept is $g = (1 - \lambda)[\gamma + \sigma\mu(1 - \xi)]$, using the identity $\beta\lambda(1 - \rho)/(1 - \beta\lambda\rho) = 1 - \xi$.

Steady-state check. Setting $p = \mu$ and $x = x_{-1} = \bar{x}$ in the optimal policy gives $\bar{x} = \sigma\mu + \gamma = x^*(\mu)$: the firm sits at its frictionless target when nothing is changing.

Remaining coefficients. The coefficients (C, E, F) complete the value function but are not needed for the policy or the correction formula.

2.A.4 LQ Approximation as a Single Second-Order Expansion

The two-step procedure in Section 2.2.3 of a second-order Taylor expansion in x followed by constant-curvature evaluation, is equivalent to a single joint second-order Taylor expansion of the excess cost function $L(x, p) := c(x; p) - c(x^*(p); p)$ around the reference point $(\bar{x}, \bar{p}) = (x^*(\bar{p}), \bar{p})$.

Gradient vanishes along the optimum ridge. The excess cost L equals zero along the entire curve $x = x^*(p)$, not just at the reference point. Consequently, both L and its gradient vanish at $(\bar{x}, \bar{p})$:

$$L(\bar{x}, \bar{p}) = 0, \qquad L_x(\bar{x}, \bar{p}) = 0, \qquad L_p(\bar{x}, \bar{p}) = 0.$$

The first condition holds by definition of excess cost. The second holds because $\bar{x} = x^*(\bar{p})$ minimizes $c(\cdot; \bar{p})$. The third follows by totally differentiating $L(x^*(p), p) = 0$ with respect to p: $L_x(x^*)'(p) + L_p = 0$, and since $L_x = 0$ at $x = x^*(p)$, we get $L_p = 0$.

Rank-one Hessian. Because the value and gradient vanish, only the Hessian contributes to the second-order expansion. Its three entries at $(\bar{x}, \bar{p})$ are:

$$L_{xx} = \bar{\kappa}, \qquad L_{xp} = -\sigma\bar{\kappa}, \qquad L_{pp} = \sigma^2\bar{\kappa},$$

where $\bar{\kappa} := c''(x^*(\bar{p}); \bar{p})$. The cross-derivative $L_{xp} = -\sigma\bar{\kappa}$ follows from differentiating the identity $c_x(x^*(p); p) = 0$ with respect to p (using $x^{*\prime}(p) = \sigma$). The second pure-price derivative $L_{pp} = \sigma^2\bar{\kappa}$ follows from applying the envelope theorem twice to the minimized cost $c^*(p) = c(x^*(p); p)$. The Hessian is rank one:

$$H = \bar{\kappa} \begin{pmatrix} 1 \\ -\sigma \end{pmatrix} \begin{pmatrix} 1 & -\sigma \end{pmatrix}.$$

Because the Hessian is rank one, the joint second-order expansion collapses to:

$$L(x, p) \approx \frac{\bar{\kappa}}{2} (x - \sigma p - \gamma)^2.$$

This is exactly the LQ production-cost term used in the Bellman equation (2.5).

Implications. Two consequences follow. First, the choice of $\bar{p}$ as the reference price is not ad hoc: the gradient of L vanishes at every point on $x = x^*(p)$, so the joint expansion yields the same quadratic form for any reference price. Second, the two-step procedure discards no second-order content: it captures all second-order information about L in (x, p)

jointly. The leading error consists of third-order and higher terms in $(\tilde{x}, \tilde{p}) := (x - \bar{x}, p - \bar{p})$, the same order as the truncation error in the scalar expansion.

2.A.5 LQ Policy as First-Order Euler Equation Linearization

The LQ policy rule (2.6) is identical to the rule obtained by a first-order linearization of the exact Euler equation around the deterministic steady state $(\bar{x}, \bar{p})$.

Exact Euler equation. The exact dynamic problem has the Bellman equation

$$V(x_{-1}; p) = \min_x \left\{ c(x; p) + \frac{\psi}{2}(x - x_{-1})^2 + \beta \mathbb{E}[V(x; p') \mid p] \right\}.$$

The envelope theorem gives $\partial V / \partial x_{-1} = -\psi(x - x_{-1})$ regardless of the cost specification, since x_{-1} enters only through the adjustment cost. Combining the first-order condition with the envelope condition yields the exact Euler equation:

$$c'(x; p) + \psi(x - x_{-1}) - \beta\psi \mathbb{E}[x' - x \mid p] = 0. \tag{2.A.4}$$

Linearization. The adjustment cost terms in (2.A.4) are already linear in x. Linearize the marginal cost $c'(x; p)$ around $(\bar{x}, \bar{p})$, where $\bar{x} = x^*(\bar{p}) = \sigma\bar{p} + \gamma$:

$$c'(x; p) \approx \underbrace{c'(\bar{x}; \bar{p})}_{=0} + \underbrace{c''(\bar{x}; \bar{p})}_{=\kappa}(x - \bar{x}) + c'_p(\bar{x}; \bar{p})(p - \bar{p}).$$

The zero-order term vanishes because $\bar{x}$ minimizes $c(\cdot; \bar{p})$. The cross-derivative follows from differentiating the identity $c'(x^*(p); p) \equiv 0$ with respect to p:

$$c''(x^*(p); p) \cdot \sigma + c'_p(x^*(p); p) = 0 \quad \Longrightarrow \quad c'_p(\bar{x}; \bar{p}) = -\kappa\sigma.$$

Substituting and using $\bar{x} + \sigma(p - \bar{p}) = \sigma p + \gamma = x^*(p)$:

$$c'(x; p) \approx \kappa(x - \bar{x}) - \kappa\sigma(p - \bar{p}) = \kappa\big(x - x^*(p)\big).$$

Equivalence. Substituting into (2.A.4) yields precisely the LQ Euler equation (appendix 2.A.3):

$$\kappa(x_t - x^*(p_t)) + \psi(x_t - x_{t-1}) - \beta\psi\mathbb{E}_t[x_{t+1} - x_t] = 0. \tag{2.A.5}$$

The LQ policy rule therefore solves the linearized first-order conditions of the exact problem.

2.A.6 LQ Policy Approximation Error Order

In this subsection, I quantify the error of the LQ policy rule (2.6) relative to the exact optimal policy of the dynamic problem (2.3). I first evaluate the exact Euler equation (2.A.4) at the LQ policy, decompose the resulting residual into its two approximation-step components, and then bound each component in terms of the innovation volatility σ_η.

Euler residual. Define the Euler residual e_t as the left-hand side of the exact Euler equation (2.A.4) evaluated at the LQ policy x_t^{LQ}, and write $d_t \coloneqq x_t^{LQ} - x^*(p_t)$ for the LQ tracking error:

$$e_t \coloneqq c'(x_t^{LQ}; p_t) + \psi(x_t^{LQ} - x_{t-1}^{LQ}) - \beta\psi\mathbb{E}_t[x_{t+1}^{LQ} - x_t^{LQ}].$$

By construction, the LQ policy satisfies the LQ Euler equation (2.A.5) identically, so evaluating that equation at $(x_{t-1}^{LQ}, x_t^{LQ}, x_{t+1}^{LQ})$ gives zero:

$$0 = \kappa(x_t^{LQ} - x^*(p_t)) + \psi(x_t^{LQ} - x_{t-1}^{LQ}) - \beta\psi\mathbb{E}_t[x_{t+1}^{LQ} - x_t^{LQ}].$$

Subtracting that from e_t gives

$$e_t = c'(x_t^{LQ}; p_t) - \kappa(x_t^{LQ} - x^*(p_t)) = c'(x_t^{LQ}; p_t) - \kappa d_t.$$

Given smoothness and boundedness of c', we immediately get $e_t = O(\sigma_\eta)$: each of $c'(x_t^{LQ}; p_t)$ and κd_t is individually $O(\sigma_\eta)$, since $d_t = O(\sigma_\eta)$ (the LQ policy is linear in p_t and collapses to $\bar{x}$ as $\sigma_\eta \to 0$) and c' inherits the same rate through $c'(\bar{x}; \bar{p}) = 0$. But we can go further, see below.

Second-order Taylor-expand $c'(x; p_t)$ in x around $x^*(p_t)$, using $c'(x^*(p_t); p_t) = 0$ and defining the state-dependent curvature $\kappa(p_t) := c''(x^*(p_t); p_t)$:

$$c'(x_t^{LQ}; p_t) = \kappa(p_t) d_t + \frac{1}{2} c'''(x^*(p_t); p_t) d_t^2 + O(d_t^3).$$

Substituting yields the two-component decomposition, which corresponds to the two steps of the approximation in Section 2.2.3:

$$e_t = \underbrace{[\kappa(p_t) - \kappa] d_t}_{e_t^\kappa} + \underbrace{\frac{1}{2} c'''(x^*(p_t); p_t) d_t^2 + O(d_t^3)}_{e_t^T}. \tag{2.A.6}$$

The first component e_t^κ is the error from freezing the curvature at $\kappa = \kappa(\bar{p})$; the second e_t^T is the error from dropping the nonlinearity of c' in x.

Perturbation. To bound the residual components, I borrow from perturbation theory and treat the stochastic price process as the variable to be perturbed. The innovation standard deviation σ_η indexes the perturbation: as $\sigma_\eta \to 0$ the price process degenerates to the constant $\bar{p}$. Orders $O(\sigma_\eta^k)$ below refer to this limit: there exists a constant C independent of σ_η such that $|X| \leq C\sigma_\eta^k$ when X is deterministic, or $\mathbb{E}[X^2]^{1/2} \leq C\sigma_\eta^k$ when X is a random variable that depends on σ_η only through the innovations $\{\eta_s\}_{s \leq t}$. Under the assumption of $\eta \sim N(0, \sigma_\eta^2)$, every polynomial function of these innovations has bounded moments of all orders, so products compose: $O(\sigma_\eta^a) \cdot O(\sigma_\eta^b) = O(\sigma_\eta^{a+b})$. Since $\rho \in (0, 1)$ is held fixed, the stationary standard deviation $\sigma_p = \sigma_\eta / \sqrt{1 - \rho^2}$ is a fixed multiple of σ_η, so $\sigma_p = O(\sigma_\eta)$ and either may be used interchangeably as the order parameter.

Approximation order. Write $\tilde{p}_t \coloneqq p_t - \bar{p}$ and $\tilde{x}_t \coloneqq x_t^{LQ} - \bar{x}$ for the deviations from steady state. The price deviation satisfies $\mathbb{E}[\tilde{p}_t^2]^{1/2} = \sigma_\eta/\sqrt{1-\rho^2}$ in the stationary distribution, hence $\tilde{p}_t = O(\sigma_\eta)$. Subtracting the steady-state identity $\bar{x} = \lambda\bar{x} + f\bar{p} + g$ from the LQ policy rule (2.6), yields the first-order recursion $\tilde{x}_t = \lambda\tilde{x}_{t-1} + f\tilde{p}_t$. This has the MA($\infty$) representation $\tilde{x}_t = f\sum_{j=0}^{\infty}\lambda^j\tilde{p}_{t-j}$. The coefficients $f\lambda^j$ are independent of σ_η and, since $\lambda \in (0,1)$, the series converges. Then, $\tilde{x}_t$ depends on σ_η only through the innovations $\{\eta_s\}_{s\le t}$. Together with $\tilde{p}_t = O(\sigma_\eta)$ this gives $\tilde{x}_t = O(\sigma_\eta)$. Combined with $x^*(p_t) - \bar{x} = \sigma\tilde{p}_t = O(\sigma_\eta)$, the tracking error is $d_t = \tilde{x}_t - \sigma\tilde{p}_t = O(\sigma_\eta)$. Taylor-expanding $\kappa(\cdot)$ around $\bar{p}$ gives $\kappa(p_t) - \kappa = \kappa'(\bar{p})\tilde{p}_t + O(\sigma_\eta^2) = O(\sigma_\eta)$. Both residual components in (2.A.6) are products of $O(\sigma_\eta)$ factors, the order of the Euler residual is therefore $O(\sigma_\eta^2)$. The policy error inherits that order: $x_t^{LQ} - x_t^{\text{exact}} = O(\sigma_\eta^2)$ (Santos, 2000).

2.A.7 Within-Model Steady-State Gain

A subtlety arises if one feeds a permanent level shift through the existing policy rule without updating the intercept: The firm observes a higher p_t but does not know that μ has changed, treating the shift as a transitory shock. Its steady-state response is then not σ but the *within-model gain*:

$$\xi\sigma = \frac{1-\beta\lambda}{1-\beta\lambda\rho}\sigma.$$

This is the object the between estimator identifies under stationary idiosyncratic prices. It exceeds the short-run elasticity $\sigma_{SR} = (1-\lambda)\xi\sigma$ but falls short of σ whenever $\rho < 1$, because the firm never fully adjusts to what it expects to be a transitory shock.

The gap $\sigma - \xi\sigma$ vanishes only when $\rho \to 1$, which makes the firm's belief and the shock's true persistence consistent. The contrast with the long-run elasticity $\sigma_{LR} = \sigma$ (Definition 2.2) is one of *information*: when the firm knows μ has shifted, it updates the intercept and adjusts fully; when it does not, it relies only on the price coefficient $\xi\sigma$ and under-adjusts. The within-model gain $\xi\sigma$ equals the cumulative impulse response (cumulative impulse response) ratio, which is the correct long-run concept for transitory shocks, while σ governs permanent environment shifts.

This is the distinguishing feature of the error-correction model and co-integration models (Engle & Granger, 1987). Those models assume a unit root process, i.e., $\rho \to 1$, and thus do not feature this dampening.

2.A.8 AR(k) Generalization

This subsection sketches the two results stated in Section 2.2.4: that the persistence parameter λ is unchanged when prices follow an AR(k) process, and that the correction formula generalizes to involve the companion matrix A of the price process.

Setup. Write the AR(k) price process in companion form $\mathbf{p}_t = A\mathbf{p}_{t-1} + \eta_t \mathbf{e}_1$, where $\mathbf{p}_t = (p_t, p_{t-1}, \ldots, p_{t-k+1})'$, A is the $k \times k$ companion matrix, $\mathbf{e}_1$ is the first standard basis vector, and all eigenvalues of A lie strictly inside the unit circle. The j-step-ahead forecast is $\mathbb{E}_t[p_{t+j}] = \mathbf{e}_1' A^j \mathbf{p}_t$. The frictionless target remains $x^*(p) = \sigma p + \gamma = \sigma \mathbf{e}_1' \mathbf{p} + \gamma$.

λ is price-process independent. The Euler equation of the dynamic problem is:

$$\kappa(x - x^*(p)) + \psi(x - x_{-1}) - \beta\psi\mathbb{E}[x' - x \mid p] = 0.$$

This equation involves only the cost parameters (κ, ψ, β) and the conditional expectation of the policy, not the parametric form of the price process. The homogeneous characteristic polynomial,

$$\beta\psi z^2 - (\kappa + \psi + \beta\psi)z + \psi = 0,$$

is therefore identical to the AR(1) case (appendix 2.A.3). The stable root $\lambda < 1$ and the identity $\kappa/\psi = (1 - \lambda)(1 - \beta\lambda)/\lambda$ carry over unchanged.

Policy under AR(k). Iterating the Euler equation forward to eliminate the unstable root $1/(\beta\lambda)$ (as in Sargent, 1978) and substituting the j-step-ahead expected target $\sigma\mathbf{e}_1' A^j \mathbf{p} + \gamma$ gives:

$$x = \lambda x_{-1} + \frac{\lambda\kappa}{\psi} \sum_{j=0}^{\infty} (\beta\lambda)^j \left(\sigma\mathbf{e}_1' A^j \mathbf{p} + \gamma\right).$$

Summing the matrix geometric series (which converges because $|\beta\lambda\lambda_i^A| < 1$ for all eigenvalues λ_i^A of A):

$$x = \lambda x_{-1} + \mathbf{f}'\mathbf{p} + (1-\lambda)\gamma, \qquad \mathbf{f}' = (1-\lambda)(1-\beta\lambda)\sigma\mathbf{e}_1'(I - \beta\lambda A)^{-1}, \tag{2.A.7}$$

where the prefactor uses the identity $\lambda\kappa/\psi = (1-\lambda)(1-\beta\lambda)$. For $k = 1$, $(I-\beta\lambda A)^{-1} = (1-\beta\lambda\rho)^{-1}$, and $\mathbf{f}' = (1-\lambda)\xi\sigma$ with $\xi = (1-\beta\lambda)/(1-\beta\lambda\rho)$, recovering (2.6).

Attenuation and correction under AR(k). The static regression $x_t = a + \hat{\sigma} p_t + u_t$ omits x_{t-1} and the lagged prices $p_{t-1}, \ldots, p_{t-k+1}$, inducing omitted variable bias. Let $\boldsymbol{\gamma}_x \coloneqq \operatorname{Cov}(x_t, \mathbf{p}_t)$ and $\Gamma_{\mathbf{p}} \coloneqq \operatorname{Var}(\mathbf{p}_t)$. Expanding $\boldsymbol{\gamma}_x$ from the DGP (2.A.7) and using stationarity to relate $\operatorname{Cov}(x_{t-1}, \mathbf{p}_t) = A\boldsymbol{\gamma}_x$ yields the cross-covariance system:

$$\boldsymbol{\gamma}_x = (I - \lambda A)^{-1}\Gamma_{\mathbf{p}}\mathbf{f}.$$

Since $\hat{\sigma} = \mathbf{e}_1'\boldsymbol{\gamma}_x / \operatorname{Var}(p_t)$, substituting the expression for $\mathbf{f}$ gives the attenuation ratio:

$$\frac{\hat{\sigma}}{\sigma} = (1-\lambda)(1-\beta\lambda)\mathbf{e}_1'(I-\lambda A)^{-1} R_{\mathbf{p}} (I-\beta\lambda A')^{-1}\mathbf{e}_1 \coloneqq \phi(\lambda, \beta, A), \tag{2.A.8}$$

where $R_{\mathbf{p}} \coloneqq \Gamma_{\mathbf{p}} / \operatorname{Var}(p_t)$ is the autocorrelation matrix of $\mathbf{p}_t$, determined by A via the Yule-Walker equations. The correction is:

$$\sigma = \hat{\sigma} \cdot [\phi(\lambda, \beta, A)]^{-1}.$$

For $k = 1$, both resolvents are scalars: $(1 - \lambda A)^{-1} = (1 - \lambda\rho)^{-1}$, $(1 - \beta\lambda A')^{-1} = (1-\beta\lambda\rho)^{-1}$, and $R_{\mathbf{p}} = 1$, so $\phi = (1-\lambda)(1-\beta\lambda)/[(1-\lambda\rho)(1-\beta\lambda\rho)]$, recovering (2.12).

Appendix 2.B Cross-Sectional Identification

2.B.1 Micro-Foundation: Endogenous Prices from Market Power

This appendix provides the micro-foundation for the endogenous price equation used in Section 2.4 and develops the econometric identification argument for IV estimation of long-run elasticities in panel settings.

Suppose each input $j \in \{1, 2\}$ is supplied to the firm along an upward-sloping schedule:

$$P_{j,it} = P_{j,t}^0 \cdot X_{j,it}^{\omega_j}, \qquad \omega_j \geq 0,$$

where $P_{j,t}^0$ is the exogenous component of input j's price and ω_j is the inverse supply elasticity. The parameter $\omega_j = 0$ corresponds to competitive input markets (price-taking); $\omega_j > 0$ means the firm has monopsony power, pushing up the price as it buys more. ($\omega_j < 0$ corresponds to quantity discounts.)

The firm's total expenditure on input j is $E_j = P_j^0 X_j^{1+\omega_j}$. The relevant price for optimization is the marginal expenditure (shadow price):

$$w_j = \frac{\partial E_j}{\partial X_j} = (1 + \omega_j) P_j^0 X_j^{\omega_j} = (1 + \omega_j) P_j.$$

The CES first-order condition equates the marginal rate of technical substitution to the ratio of shadow prices. In the notation of Section 2.2.1:

$$x_i^* = \sigma \left[p_i^{\text{obs}} + \log \frac{1 + \omega_1}{1 + \omega_2} \right] + \gamma = \sigma p_i^{\text{obs}} + \gamma + \sigma \log \frac{1 + \omega_1}{1 + \omega_2},$$

where $p_i^{\text{obs}} = \log(P_{1,i}/P_{2,i})$ is the observed (average) price ratio. The slope on the observed price ratio is still σ: market power shifts the intercept but preserves the structural elasticity.

In logs, the observed price ratio is $p_i^{\text{obs}} = p_i^0 + \omega_1 \log X_{1,i} - \omega_2 \log X_{2,i}$. Define $s_i = \frac{1}{2}(\log X_{1,i} + \log X_{2,i})$ as the log scale of input use. Since $\log X_{1,i} = s_i - x_i/2$ and $\log X_{2,i} = s_i + x_i/2$:

$$\omega_1 \log X_{1,i} - \omega_2 \log X_{2,i} = (\omega_1 - \omega_2)s_i - \frac{\omega_1 + \omega_2}{2} x_i.$$

With $\alpha := \omega_1 - \omega_2$ (differential market power) and $\omega := (\omega_1 + \omega_2)/2$ (average market power):

$$p_i^{\text{obs}} = p_i^0 + \alpha s_i - \omega x_i. \tag{2.B.1}$$

Under monopsony ($\omega_1, \omega_2 > 0$), $\omega \geq 0$: a firm with a high input ratio x_i uses relatively more X_2, pushing up P_2 and lowering the relative price p_i. Under quantity discounts, $\omega < 0$: the feedback is reversed. Under constant returns to scale, the optimal input ratio x^* depends only on relative prices, not on output level; scale s_i is therefore independent of the structural equation's error, providing a source of exogenous price variation through the αs_i channel.

Chapter 3

Financial frictions and aggregate risk exposure

I thank my advisors Per Krusell and Joshua Weiss for helping shape this paper over time. I am also grateful for comments and feedback from John Hassler, Timo Boppart, Kieran Larkin, Yimei Zou, Kurt Mitman, and other participants of the IIES Macro Group, as well as the Yale Macro Breakfast.

3.1 Introduction

Recent supply chain disruptions have prompted calls for preemptive industrial policy: government intervention before shocks occur to reduce exposure to economic risk. This paper studies the question in a model with two key features. If the government has limited ability to react after a shock, it can only try to shape the composition of production preemptively by redistributing resources toward less-exposed technologies before adverse events occur. But if the government cannot commit to a redistribution rule before firms choose their technology, firms anticipate the government's response to their choices and adjust accordingly, potentially undoing the very benefits the policy was meant to achieve. The core question this paper asks is whether such intervention improves welfare when the government lacks commitment power.

Supply chain vulnerabilities have become a central concern for policymakers. China's 2010 restrictions on rare earth exports to Japan demonstrated how supply dominance can be weaponized: the restrictions were unexpected, targeted goods for which substitution was difficult, and caused significant disruption (Gholz & Hughes, 2021; Morrison & Tang, 2012). China restricted graphite exports in 2023 in response to U.S. semiconductor controls, another episode of politically motivated supply chain disruption.[1] Energy dependence creates similar vulnerabilities: the 1973–74 oil crisis precipitated a major recession in the U.S. (Alpanda & Peralta-Alva, 2010), while Russia's 2022 gas cutoff to Europe led to output contractions in energy-intensive industries (Moll et al., 2023). Considerable fractions of world crude oil and natural gas production can not pass the Straight of Hormuz as I write this paper, the impacts of which are still unfolding.[2] These episodes can be characterized by their disruption severity: the extent to which affected firms lose access to their inputs. Most episodes involve a partial disruption, a price spike or cost

[1] The Economist, "Why China is restricting exports of graphite" (2023).

[2] The Economist, "The third Gulf war will scar energy markets for a long time yet" (2026).

increase that reduces but does not eliminate productivity. A complete supply cutoff is the extreme case. The model predicts that the policy implications differ sharply across this spectrum, and the results suggest that the extreme scenario of near-complete supply cutoff is the only case in which preemptive intervention improves welfare.

I study this question in a model in which entrepreneurs choose between a safe technology with stable productivity and a risky technology with higher expected productivity but exposure to adverse shocks. The only friction is a collateral constraint that limits firms' ability to borrow capital relative to their assets, capturing the limited ex post adjustment documented in empirical episodes. A planner redistributes assets across technology types before production occurs, relaxing borrowing constraints for some firms at the cost of tightening them for others. Crucially, the planner lacks commitment power: it chooses redistribution after observing the composition of technology choices, and agents anticipate this when making their decisions.

The model maps directly to the empirical episodes. The risky technology's productivity in the bad state parameterizes disruption severity. A moderate productivity loss captures partial disruptions such as the rare earth and graphite restrictions: productivity falls, but production continues. A near-total productivity collapse captures near-complete supply cutoff, as faced by energy-intensive industries during Russia's 2022 gas cutoff. The results characterize welfare outcomes across this spectrum.

The analysis yields three main results. First, I characterize three regimes based on how borrowing constraints bind across technology types. When constraints are slack, redistribution cannot affect output. When they bind broadly, redistribution involves a trade-off: relaxing one type's constraint tightens another's. The paper focuses on this trade-off regime, the relevant case in which the planner faces a genuine reallocation problem.

Second, comparing no-commitment intervention to laissez-faire—which is unique and constrained efficient—the welfare effect decomposes into two opposing forces: (i) a redistribution gain from optimally reallocating assets holding technology composition fixed, and (ii) a distortion cost from the shift in technology composition induced by anticipated redistribution. The distortion arises because agents respond to anticipated redistribution through two channels: higher expected profits from relaxed constraints (the profit channel) and a direct wealth transfer (the wealth channel). The planner internalizes only the first when choosing redistribution. This wedge causes agents to overshoot: they shift toward the favored technology beyond what is socially optimal, and the resulting shift may be welfare-reducing.

Third, numerical analysis across a wide parameter grid establishes the main quantitative finding: the distortion cost exceeds the redistribution gain for the majority of empirically relevant parameterizations. The exception arises only near complete supply cutoff: only when the risky technology can produce almost nothing does intervention marginally improve welfare.

Mapping back to the empirical episodes, the policy implication is sobering. For partial disruptions (equivalent to moderate price spikes) the distortion cost dominates, and a government lacking commitment is better off not intervening. Only for near-complete supply cutoff (substantial price spikes), an extreme and empirically rare scenario, does intervention improve welfare. The conclusion is not that redistribution is always harmful, but that a government without the ability to commit to a redistribution rule before agents choose their technology will typically do more harm than good.

The remainder of the paper proceeds as follows. Section 3.2 presents the model environment, including the technology choice problem, collateral constraints, and the planner's redistribution instrument. Section 3.3 characterizes the constraint regimes, derives the welfare decomposition, and presents the numerical comparison of laissez-faire to no-commitment intervention. Section 3.4 discusses policy implications and extensions.

Literature A growing literature studies supply chain disruptions and their consequences. Baldwin and Freeman (2022) propose a risk-versus-reward framework for evaluating policy interventions. Barrot and Sauvagnat (2016) document empirically that idiosyncratic shocks propagate through production networks, with propagation amplified by input specificity, or the degree of ex post substitutability. Whether supply chains are generally efficiently structured remains debated: Kopytov et al. (2024) find that equilibrium networks balance productivity against reliability efficiently, whereas Capponi et al. (2024) argue that market power leads to underinvestment in resilience. Grossman et al. (2023) study the optimal policy for diversification versus "reshoring". My contribution to this literature is to study technology choice under supply risk when financial frictions limit ex post adjustment and when the government lacks commitment.

Geopolitical tensions have renewed attention to deliberate supply disruptions. Gholz and Hughes (2021) document Japan's adjustment to China's 2010 rare earth restrictions through stockpiling and substitution, emphasizing that geopolitical leverage is strongest when restrictions are unexpected. Morrison and Tang (2012) analyze the impacts on the U.S. of the same episode. Energy dependence offers related examples: Alpanda and Peralta-Alva (2010) and Wei (2003) show that capital obsolescence after the 1973–74 oil crisis explains much of the decline in U.S. firm valuations. Moll et al. (2023) analyze Germany's output contraction following Russia's 2022 gas shock. The model in this paper is designed to speak to the severity dimension that distinguishes these episodes.

The core mechanism in my model, collateral constraints limiting firms' ability to scale production, builds on Moll (2014), who shows that such frictions generate substantial aggregate productivity losses through misallocation. Midrigan and Xu (2014) find that financial frictions distort technology adoption in particular. Guntin and Kochen (2024) provide a closely related parameterization of the borrowing constraint.

A key result is that the government's redistribution policy is time inconsistent: without commitment power, the planner cannot credibly promise ex ante what it will do ex post. Chari and Kehoe (2016) show that benevolent governments without commitment introduce inefficiencies through bailouts; my no-commitment equilibrium exhibits the same logic applied to preemptive industrial policy. Traiberman and Rotemberg (2023) study precautionary trade policy when adjustment is slow; my paper shares the feature that intervention is only effective ex ante, once technology choices are sunk. The novel finding is that even this ex ante effectiveness is undone by the anticipation problem.

3.2 Model

3.2.1 Setup

Agents choose between two production technologies before observing the state of the economy. The safe technology delivers a stable productivity; the risky technology has higher expected productivity but is exposed to an aggregate bad-state shock. Producers borrow capital against their assets, subject to a collateral constraint that limits ex post adjustment. After the state realizes, capital is allocated and goods are produced. A planner can intervene preemptively by redistributing assets across technology types but cannot react to the state itself.

Environment The model features a static economy, populated by a unit mass of consumer-entrepreneur agents. They are endowed with assets $a \geq 0$.

The economy faces aggregate uncertainty. The state of the economy S can be low or high, $S \in \{\ell, h\}$. With probability $\theta \in (0, 1)$ it is low, and with probability $1 - \theta$ it is high.

There are two stages: First, a decision stage, in which agents choose a production technology, and second the production and consumption stage. The aggregate state realizes between the decisions in the first stage, and production and consumption in the second stage.

Preferences Agents have linear utility over consumption $c > 0$, $u(c) = c$. They either consume their endowment a directly, or become producers.

Technology They produce the output good using capital $k \geq 0$, with a decreasing returns to scale production function. Output y is produced according to

$$y = f(k, z) = z^{1-\alpha} k^{\alpha}, \tag{3.1}$$

with returns to scale parameter $0 < \alpha < 1$ and productivity z. The value of z depends on the chosen technology T and the state of the economy S. Agents have access to two production technologies T, call them safe ($T = d$) and risky ($T = v$). The safe technology delivers a constant productivity, while the risky technology has a high productivity in the good state and a low productivity in the bad state.[3] Formally, the realizations of z depending on an agent's technology choice T and aggregate state S are

$$z(S, T) = \begin{cases} z_d & \text{if } T = d, \\ z_h & \text{if } T = v \text{ and } S = h, \\ z_\ell & \text{if } T = v \text{ and } S = \ell, \end{cases} \tag{3.2}$$

with $z_h > z_d > z_\ell > 0$.

Agents borrow production capital k against their assets a, subject to the collateral constraint

$$k \leq \lambda a, \tag{3.3}$$

with $\lambda \geq 1$ describing the tightness of the collateral constraint.

[3] Appendix 3.A.1 provides a micro-foundation for this productivity process based on intermediate inputs as a complementary input to capital. The productivity term z then depends on the elasticity of substitution, and the price of the intermediate input, which is stochastic for the risky technology and constant for the safe technology. The "risky" technology corresponds to supply chains exposed to input price volatility (e.g., imported materials), while the "safe" technology uses inputs with stable prices.

Markets There are two markets: for the output good and for capital. The price of the output good is the numéraire and normalized to 1. The price of capital, the interest rate r, is determined in equilibrium.

The timing of the model is as follows:

1. A unit mass of agents with asset endowment a populates the economy.

2. Agents choose whether to produce, and a production technology $T \in \{d, v\}$.

3. The aggregate state S realizes, and agents observe it.

4. Agents trade their assets on the capital market, subject to the collateral constraint, and produce output using their chosen technology. They trade the output good, and consume.

Equilibrium I consider a rational expectations equilibrium. An equilibrium consists of state-contingent interest rates, $r^\star(S)$ for $S \in \{\ell, h\}$, state- and technology-contingent capital choices, $k^\star(S,T)$, an entry decision, $e^\star \in \{0,1\}$, and a technology choice, $T^\star \in \{d, v\}$, such that:

- Given the interest rates $r^\star(S)$, the production decision $e^\star$, and the technology choice $T^\star$, agents maximize their expected utility by choosing capital $k^\star(S,T)$ in each state S and for each technology T.

- The markets for capital and the output good clear in each state S.

- The production decision $e^\star$ and technology choice $T^\star$ are optimal given the expected payoffs from production, and given the choices of other agents.

Given the ex ante homogeneity of agents, the equilibrium can be characterized by one variable: the share of agents choosing the safe technology, denoted by q.

3.2.2 The agents' problem

Agents maximize their expected utility of consumption. If they do not produce, they consume their initial endowment a. If they produce, they choose a technology T, and generate profits from production. They will then consume their initial endowment plus the profits from production.

To describe the agents' problem, start from the second stage. The agent has chosen a technology T, the aggregate state S has realized, and the agent observes q, the share of agents that use the safe technology. Consumption is given by

$$c = \pi + (1 + r)a, \tag{3.4}$$

where π are the profits from production, and $(1 + r)a$ is the gross return on the agent's assets. Maximizing utility is equivalent to maximizing profits from production, given by

$$\pi(r, z, a) = \max_{k \geq 0} z^{1-\alpha} k^{\alpha} - rk, \quad \text{s.t. } k \leq \lambda a. \tag{3.5}$$

The optimal capital choice as a function of r, z, and a is then

$$k^{\star}(r, z, a) = \min \left\{ \lambda a; z \left(\frac{\alpha}{r} \right)^{\frac{1}{1-\alpha}} \right\}. \tag{3.6}$$

If the optimal scale is below the collateral constraint, the agent operates at the first-best scale. If it is above, the agent is constrained and chooses the maximum capital allowed by the collateral constraint.

The agent's expected utility from entering and choosing technology T in the first stage is then

$$U(T, q, a) = \mathbb{E} \left[\pi(r(S, q), z(S, T), a) + (1 + r(S, q))a \right], \tag{3.7}$$

where the expectation is taken over the aggregate state S. They will choose the technology that maximizes their expected utility from production, and enter if this expected utility exceeds their outside option of not producing, which is a:

$$T^\star = \arg\max_{T\in\{d,v\}} U(T), \tag{3.8}$$

$$e^\star = \begin{cases} 1 & \text{if } U(T^\star) > \mathbb{E}[a(1+r(S,q))], \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

Profits are weakly positive, so entry is always optimal.

3.2.3 Decentralized equilibrium

Walras' law holds. The goods market clears if the capital market clears, and I will solve for the equilibrium through the capital market.

Capital market

An equilibrium is characterized by the share of agents choosing the safe technology, q. Given q, the state-contingent interest rates $r(S,q)$ must clear the capital market in each state S. The capital market clearing condition in state S is given by

$$\bar{a} = qk^\star(r(S,q), z(S,d), a) + (1-q)k^\star(r(S,q), z(S,v), a), \tag{3.10}$$

total assets $\bar{a} = qa + (1-q)a = a$ must equal the total demand for capital from both types of agents. There are three cases for the interest rate: (i) both types are unconstrained, (ii) the safe technology type is constrained and the risky unconstrained, and (iii) the risky technology type is constrained and the safe unconstrained. Note that both types can not be constrained at the same time: that would imply both types want to borrow more than their assets, which is incompatible with the market clearing condition.

I derive the equilibrium interest rates for each case by substituting the optimal capital choices into the market clearing condition.

(i) Both types unconstrained:

$$r(S,q) = \alpha \left(\frac{qz(S,d) + (1-q)z(S,v)}{a} \right)^{1-\alpha}$$

(ii) Safe type constrained, risky unconstrained:

$$r(S,q) = \alpha \left(\frac{(1-q)z(S,v)}{a(1-q\lambda)} \right)^{1-\alpha} \quad \text{if } \lambda \leq \frac{z(S,d)}{qz(S,d) + (1-q)z(S,v)}$$

(iii) Risky type constrained, safe unconstrained:

$$r(S,q) = \alpha \left(\frac{qz(S,d)}{a(1-(1-q)\lambda)} \right)^{1-\alpha}$$

if

$$\lambda \leq \frac{z(S,v)}{(1-q)z(S,v) + qz(S,d)}$$

Note that in the unconstrained case, the price of capital is a function of the population-weighted average productivity. In the constrained cases, the interest rate depends on the productivity only of the unconstrained type and the assets available to them (after subtracting the capital borrowed by the constrained type). The agents of the unconstrained type are the marginal borrowers, so their marginal product determines the interest rate.

For the interest rate to be well-defined in both constrained cases, we must have $1 - q\lambda > 0$ and $1 - (1-q)\lambda > 0$ respectively. These conditions will always hold: they are implied by the respectively conditions on λ above. Consider the case of safe type being constrained, and risky unconstrained: The interest rate is well-defined if $q\lambda < 1$, and we are in this case if $\lambda \leq \frac{z(S,d)}{qz(S,d)+(1-q)z(S,v)}$. But the latter condition implies the former, since

$$
\begin{aligned}
\lambda &\le \frac{z(S,d)}{qz(S,d)+(1-q)z(S,v)} \\
\iff q\lambda &\le \frac{qz(S,d)}{qz(S,d)+(1-q)z(S,v)} < 1 \quad \forall q \in (0,1),
\end{aligned}
$$

and symmetrically for the other case. So the interest rates are well-defined in all cases of the constraints binding.

Definition 3.1 (Feasibility set)**.** *Call* $\mathcal{F} := (1-1/\lambda, 1/\lambda)$ *the* feasibility set*. The conditions* $q\lambda < 1$ *and* $(1-q)\lambda < 1$*, derived above, are equivalent to* $q \in \mathcal{F}$*.*

Remark 3.1 (Feasibility requires $\lambda < 2$)**.** *The solution being in* $\mathcal{F}$ *is not an additional restriction imposed on the model. It is an arithmetic consequence of three features already present: a fixed aggregate capital endowment* a*, a collateral constraint that pins one type's capital at* λa*, and the requirement that the other type's residual capital allocation be non-negative. Violating the bounds of* $\mathcal{F}$ *makes the allocation arithmetically infeasible. The width of* $\mathcal{F}$ *is* $1/\lambda - (1-1/\lambda) = 2/\lambda - 1$*, which is positive if and only if* $\lambda < 2$*. For* $\lambda \ge 2$ *the feasibility set is empty: no* $q \in (0,1)$ *allows both types to coexist with at most one constrained. The model therefore implicitly requires* $\lambda \in (1,2)$ *for an interior equilibrium with mixed technology adoption to exist.*

Technology choice

Given the homogeneity of agents, the equilibrium condition for technology choice is an indifference condition: the expected utility of consumption of both technologies must be equal in equilibrium,

$$
U(d, q^\star, a) = U(v, q^\star, a). \tag{3.11}
$$

Corner solutions with all agents choosing one technology can also arise, depending on parameters. They follow when one technology strictly dominates the other in expected utility.

3.2.4 Industrial policy

A planner can engage in industrial policy, modeled as direct taxes and subsidies on the asset holdings conditional on an entrepreneur's technology choice. This is a reduced-form modeling device, which can represent for example taxes and subsidies to entry costs conditional on the technology choice. Importantly, I impose that the planner can not react to the aggregate state, but must set its policy ex ante. This is preemptive industrial policy in the sense that the planner aims to influence the composition of the economy in expectation of a shock, while facing constraints in the speed and ability to react ex post.

Formally, the planner chooses technology-contingent taxes or subsidies $\tau(T)$ for $T \in \{d, v\}$, which change the agents' asset holdings to

$$\tilde{a}(T) = a(1 + \tau(T)), \tag{3.12}$$

depending on the chosen technology. The planner is subject to a balanced budget,

$$q\tau(d) + (1 - q)\tau(v) = 0, \tag{3.13}$$

where q is the share of agents choosing the safe technology.

The timing with the planner is as follows:

1. The planner chooses a schedule of technology-contingent taxes or subsidies, $\tau(T; q)$ for $T \in \{d, v\}$, depending on the share of agents choosing the safe technology q (to ensure the budget constraint).

2. Agents observe the tax schedule, and choose whether to produce and which technology to use, taking into account the modified asset holdings $\tilde{a}(T)$.

3. The planner implements the policy, conditional on the agents' technology choices described by q.

4. The aggregate state S realizes, and agents observe it.

5. Agents trade their assets on the capital market, subject to the collateral constraint, and produce output using their chosen technology. They trade the output good, and consume.

Industrial policy has potential implications for the collateral constraint, and thus the capital market. For the analysis it is useful to define the share of total assets held by the safe technology type, denoted by w:

$$w = \frac{q\tilde{a}(d)}{q\tilde{a}(d) + (1-q)\tilde{a}(v)} = \frac{qa(1+\tau(d))}{qa(1+\tau(d)) + (1-q)a(1+\tau(v))}.$$

Imposing the balanced budget in equation (3.13), and simplifying yields

$$w = q(1+\tau(d)). \tag{3.14}$$

The share of wealth held by the safe technology type equals its population share, adjusted by the subsidy or tax on assets. Note that conditional on q, if the balanced budget holds, w is a sufficient statistic for both taxes $\tau(d)$ and $\tau(v)$. For the subsequent analysis, I will thus work with w as the policy instrument, from which the taxes can be recovered. Industrial policy acts to redistribute assets across types of agents, which directly affects their borrowing capacity: since the collateral constraint is $k \leq \lambda\tilde{a}(T)$, a higher wealth share w relaxes the constraint for the safe type while tightening it for the risky type.

Welfare

Welfare is given by the population-weighted expected utility of consumption,

$$W(q,w) = \mathbb{E}\left[qU(d,q,\tilde{a}(d)) + (1-q)U(v,q,\tilde{a}(v))\right]. \tag{3.15}$$

Given that interest payments on assets stay within the economy, and given the linear utility, welfare simplifies to expected total production (and thus consumption). First, define production output of an agent with technology T in state S, and in an environment with share q of safe agents holding share w of total assets as

$$y^{\star}(T,S,q,w) = f\left(k^{\star}(r(S,q),z(S,T),\tilde{a}(T)),z(S,T)\right).$$

Then, welfare is given by

$$W(q,w) = \mathbb{E}\left[qy^{\star}(d,S,q,w)+(1-q)y^{\star}(v,S,q,w)\right]+a.$$

Since a is constant, welfare maximization is equivalent to maximizing expected total production Y:

$$Y(q,w) = \mathbb{E}\left[qy^{\star}(d,S,q,w)+(1-q)y^{\star}(v,S,q,w)\right]. \tag{3.16}$$

Capital market after redistribution

The capital market clearing condition is the same as before, except for the modified asset holdings:

$$\bar{a} = qk^{\star}(r(S,q),z(S,d),\tilde{a}(d))+(1-q)k^{\star}(r(S,q),z(S,v),\tilde{a}(v)).$$

There are again three cases for the interest rate: (i) both types are unconstrained, (ii) the safe technology type is constrained and the risky unconstrained, and (iii) the risky technology type is constrained and the safe unconstrained, as in the baseline model without policy. The equilibrium interest rates for each case are derived by substituting the optimal capital choices into the market clearing condition.

(i) Both types unconstrained: The interest rate is the same as in the baseline model.

(ii) Safe type constrained, risky unconstrained:

$$r(S,q) = \alpha\left(\frac{(1-q)z(S,v)}{a(1-w\lambda)}\right)^{1-\alpha}$$

if

$$\lambda \frac{w}{q} \leq \frac{z(S,d)}{qz(S,d) + (1-q)z(S,v)}$$

(iii) Risky type constrained, safe unconstrained:

$$r(S,q) = \alpha \left(\frac{qz(S,d)}{a(1-(1-w)\lambda)} \right)^{1-\alpha}$$

if

$$\lambda \frac{1-w}{1-q} \leq \frac{z(S,v)}{(1-q)z(S,v) + qz(S,d)}$$

In cases (ii) and (iii), the interest rate again depends on the productivity only of the unconstrained type and the assets available to them after redistribution. Note that from equation (3.14), $w/q = 1 + \tau(d)$ represents the asset holdings of a safe technology type, and thus $\lambda w/q$ represents their borrowing capacity after redistribution.

For the interest rate to be well-defined in both constrained cases, we must have $1 - w\lambda > 0$ and $1 - (1-w)\lambda > 0$ respectively. As in the case without redistribution in section 3.2.3, these conditions are implied by the conditions for being in the respective case. Consider the case of the safe type being constrained, and risky unconstrained: The interest rate is well-defined if $w\lambda < 1$, and we are in this case if $\lambda \frac{w}{q} \leq \frac{z(S,d)}{qz(S,d)+(1-q)z(S,v)}$. But the latter condition implies the former, since

$$\begin{aligned} & \lambda \frac{w}{q} \leq \frac{z(S,d)}{qz(S,d) + (1-q)z(S,v)} \\ \iff & w\lambda \leq \frac{qz(S,d)}{qz(S,d) + (1-q)z(S,v)} < 1 \quad \forall q \in (0,1), \end{aligned}$$

and symmetrically for the other case. So the interest rates are well-defined in all cases of the constraints binding.

Remark 3.2 (Feasibility of redistribution)**.** *The same arithmetic applies to the wealth share* w*. The conditions* $w\lambda < 1$ *and* $(1-w)\lambda < 1$

derived above are equivalent to $w \in \mathcal{F}$ (Definition 3.1). The planner's instrument is therefore bounded by the same feasibility set as the equilibrium technology share: $w \in \mathcal{F}$. Consequently, the planner optimizes over $(q, w) \in \mathcal{F} \times \mathcal{F}$. This constraint is arithmetic; no tax schedule can push w outside $\mathcal{F}$ while keeping the capital market well-defined.

Planner problem

The planner problem is then

$$\begin{aligned} \max_{\tau(d),\tau(v)} & qU(d, q, \tilde{a}(d)) + (1-q)U(v, q, \tilde{a}(v)) \\ \text{s.t.} \quad & q\tau(d) + (1-q)\tau(v) \geq 0 \\ & U(d, q, \tilde{a}(d)) = U(v, q, \tilde{a}(v)), \end{aligned}$$

where the last constraint is the equilibrium condition for technology choice. Through the redistribution of assets between technologies, the planner necessarily also influences the utility levels of each type through the available assets. The planner's problem above is subject to a time-inconsistency: the tension at the core of the analysis. Through the redistribution of assets it necessarily also affects agents' technology choice, because agents anticipate the policy when they choose. The next section characterizes four regimes (constrained first-best, commitment, no-commitment, and laissez-faire) and asks whether this anticipation effect is large enough to undo the direct welfare benefit of redistribution.

3.3 Results

Does a government without commitment power improve welfare by redistributing assets across technology types, or does the anticipation of redistribution distort technology choice enough to undo the benefit? To answer this question, I decompose the welfare difference between no-commitment intervention and laissez-faire into a redistribution gain and distortion cost from agents' anticipatory technology-choice response.

The main result is that the distortion cost typically dominates the redistribution gain by a wide margin: across the empirically relevant parameter space, laissez-faire dominates no-commitment intervention in approximately 93% of parameterizations, with a median distortion-to-gain ratio of approximately four.

I first define the comparison objects: four policy regimes ranked by the planner's degree of control (Section 3.3.1). I then decompose and sign ΔY (Section 3.3.2), characterize magnitudes numerically (Section 3.3.3), and discuss exceptions and interpretation (Section 3.3.4).

3.3.1 Policy regimes and welfare ranking

I define four policy regimes as a descending hierarchy of planner power. Each regime adds a constraint on the planner relative to the previous one. The welfare ranking follows immediately from this nesting.

Planner's objective

The planner redistributes assets across types to maximize expected output. As shown in Section 3.2.4, with linear utility welfare equals expected output plus initial assets; since assets are constant, maximizing output is equivalent to maximizing welfare. The planner's objective is

$$Y(q, w) = \mathbb{E}\left[qy^{\star}(d, S, q, w) + (1-q)y^{\star}(v, S, q, w)\right]. \tag{3.17}$$

Whether redistribution affects output depends on whether the collateral constraints bind. Three configurations arise: (i) both types unconstrained—redistribution has no effect on output; (ii) one type constrained in one state, and redistribution can fully alleviate the constraint; (iii) both types constrained (safe in state ℓ, risky in state h, following from the productivity ordering $z_\ell < z_d < z_h$), and redistribution relaxes one type's constraint but tightens the other's. Across the numerical grid, almost all interior laissez-faire equilibria are in configuration (ii): only the safe type's constraint binds at $w = q$. Once the planner redistributes optimally, however, 20.7% of these cases enter configuration (iii), while the remaining cases either lift all constraints (62.2%) or stay in configuration (ii) (17.2%). The analysis focuses on configuration (iii) because it is the binding case in which the planner faces a genuine trade-off: any redistribution relaxes one type's constraint and tightens the other's.

The marginal product of assets is zero for unconstrained types and strictly positive for constrained types:

$$\frac{\partial y^\star}{\partial \tilde{a}} = \begin{cases} 0 & \text{if unconstrained,} \\ \alpha \lambda^\alpha z^{1-\alpha} \tilde{a}^{\alpha-1} & \text{if constrained.} \end{cases}$$

In configuration (iii), redistributing assets toward the safe type (increasing w) relaxes the safe type's constraint in state ℓ but tightens the risky type's constraint in state h, and vice versa. The planner cannot simultaneously relax both constraints.

Constrained first-best

The strongest benchmark is the constrained first-best (CFB): a planner that is subject only to the borrowing constraint but can freely choose both q and w. The planner solves $\max_{q,w} Y(q, w)$. The first-order conditions are

$$\frac{\partial Y}{\partial w} = \mathbb{E}\left[\frac{\partial y_d^\star}{\partial \tilde{a}_d} \times \frac{a}{q} - \frac{\partial y_v^\star}{\partial \tilde{a}_v} \times \frac{a}{1-q}\right] = 0, \tag{3.18}$$

$$\frac{\partial Y}{\partial q} = (1-\alpha)\mathbb{E}\left[y_d^\star - y_v^\star\right] = 0, \tag{3.19}$$

where $\tilde{a}_d = wa/q$ and $\tilde{a}_v = (1-w)a/(1-q)$ are post-redistribution asset holdings. The FOC for w equalizes expected marginal products of assets across types. The FOC for q equalizes expected per-agent output across technologies.[4] These conditions show what the planner would equalize

[4] The formula $\partial Y/\partial q = (1-\alpha)\mathbb{E}[y_d^\star - y_v^\star]$ holds in all constraint configurations; see Lemma 3.2 in the appendix.

if unconstrained in its choice of both instruments. Each subsequent regime shows what happens when one of these equalization conditions is disrupted.

Commitment

Under commitment, the planner announces a redistribution rule w before agents choose technology. Agents best-respond with $q^\star(w)$, the technology share that restores indifference given the announced redistribution. The planner solves

$$\begin{aligned} Y^C &= \max_w Y(q^\star(w), w) \\ &\quad \text{s.t.} \quad U(d, q^\star(w), \tilde{a}(d, w)) = U(v, q^\star(w), \tilde{a}(v, w)). \end{aligned}$$

Relative to the CFB, the planner loses direct control over technology composition: it must work through agents' incentive-compatible response. This is a hypothetical policy for comparison only; the planner has no commitment power in the model.

No-commitment equilibrium

Given the timing of policy, the planner chooses w after agents have chosen their technology. The ex post redistribution problem is

$$w^\star(q) = \arg\max_w Y(q, w). \tag{3.20}$$

The FOC is identical to (3.18), but now q is taken as given.

Without commitment, agents anticipate the planner's optimal redistribution $w^\star(q)$ when choosing technology. The no-commitment (NC) equilibrium is characterized by:

$$U(d, q^{NC}, \tilde{a}(d, w^\star(q^{NC}))) = U(v, q^{NC}, \tilde{a}(v, w^\star(q^{NC}))). \tag{3.21}$$

Agents are indifferent given the anticipated redistribution.[5] Relative to

[5] In all specifications considered numerically, the NC equilibrium is unique: the

commitment, the planner loses the ability to bind its future actions: agents' technology choice internalizes the anticipated redistribution.

The time inconsistency arises because agents care about both profits and wealth, while the planner maximizes only output. Redistribution affects utility through:

1. *Profit channel:* relaxing/tightening borrowing constraints changes expected profits.

2. *Wealth channel:* the direct asset transfer changes expected gross return on wealth.

The planner values only the first channel; agents value both. This divergence distorts technology choice away from the constrained-efficient level.

Laissez-faire

Under laissez-faire (no redistribution, $w = q$), agents choose technology to maximize expected utility. Due to ex ante homogeneity of agents, the equilibrium is characterized by indifference:

$$U(d, q^{LF}, a) = U(v, q^{LF}, a), \tag{3.22}$$

where $U(T, q, a) = \mathbb{E}[\pi(T, S, q, a) + (1 + r(S, q))a]$.

Proposition 3.1 (Unique equilibrium)**.** *There exists a unique decentralized equilibrium $q^{LF} \in [0, 1]$.*

The proof shows that the utility differential $U(d, q, a) - U(v, q, a)$ is strictly monotonic in q (appendix 3.B.1).

Proposition 3.2 (Constrained efficiency)**.** *The laissez-faire technology share q^{LF} is constrained efficient: it maximizes $Y(q, q)$ subject to $w = q$.*

indifference function is strictly monotone-decreasing throughout the feasible domain in every case examined.

This result establishes that laissez-faire achieves the best outcome among policies that maintain $w = q$ (no redistribution). It does not imply that q^{LF} is optimal when the planner can redistribute: a different q paired with optimal $w^\star(q)$ may yield higher welfare. The proof is in appendix 3.B.1.

Welfare ranking

The nesting of feasible sets across the four regimes immediately yields a welfare ranking. Define:

- $Y^{CFB} = \max_{q,w} Y(q, w)$: Constrained first-best.
- $Y^C = \max_w Y(q^\star(w), w)$ s.t. indifference: Commitment.
- $Y^{NC} = Y(q^{NC}, w^\star(q^{NC}))$: No-commitment equilibrium.
- $Y^{LF} = Y(q^{LF}, q^{LF})$: Laissez-faire.

Proposition 3.3 (Welfare ranking)**.**

$$Y^{CFB} \geq Y^C \geq \max\{Y^{NC}, Y^{LF}\}. \tag{3.23}$$

Proof. The ranking $Y^{CFB} \geq Y^C$ holds because CFB optimizes over (q, w) jointly, while commitment optimizes over w subject to agents' best response $q^\star(w)$. Under CFB, the planner could replicate the commitment outcome, so CFB weakly improves on it.

The ranking $Y^C \geq Y^{LF}$ holds because the commitment planner can replicate laissez-faire by setting $w = q^\star(w)$, which induces q^{LF}.

The ranking $Y^C \geq Y^{NC}$ holds because the commitment planner can replicate the no-commitment outcome. The NC equilibrium is characterized by $w = w^\star(q^{NC})$ and indifference at q^{NC}. The commitment planner can achieve this by committing to $w = w^\star(q^{NC})$, which induces the same q^{NC}. Since commitment allows choosing any w, it weakly improves on this outcome. □

The key ambiguity is the comparison between Y^{NC} and Y^{LF}: does intervention without commitment improve or worsen welfare relative to laissez-faire? This is the central question of the following section.

3.3.2 Welfare decomposition

I decompose the welfare difference $\Delta Y := Y^{NC} - Y^{LF}$ into two opposing forces, pin down their signs, and characterize their magnitudes numerically.

Decomposition

Definition 3.2 (Risky output dominance)**.** Risky output dominance *holds at an allocation* (q, w) *if the risky technology yields weakly higher expected output per agent than the safe technology at that allocation,*

$$\mathbb{E}[y_v^\star(S)] \geq \mathbb{E}[y_d^\star(S)],$$

where the expectation is over the aggregate state S *and* $y_\tau^\star(S)$ *denotes the equilibrium output per agent of type* $\tau \in \{d, v\}$ *at* (q, w)*.*

Proposition 3.4 (Welfare decomposition)**.** *Suppose* q^{LF} *is interior in* $\mathcal{F}$*, risky output dominance (Definition 3.2) holds at* $(q^{LF}, w^\star(q^{LF}))$*, and* $w^\star(q^{LF}) \neq q^{LF}$*. Then the welfare difference decomposes as* $\Delta Y = R + D$*, where:*

$$\begin{aligned} R &:= Y(q^{LF}, w^\star(q^{LF})) - Y(q^{LF}, q^{LF}) \geq 0, \\ D &:= Y(q^{NC}, w^\star(q^{NC})) - Y(q^{LF}, w^\star(q^{LF})) \leq 0. \end{aligned} \tag{3.24}$$

The redistribution gain R *measures the benefit of optimal redistribution holding technology composition fixed. The distortion cost* D *measures the welfare change due to the shift in equilibrium technology composition induced by anticipated redistribution.*

Proof. The decomposition follows by adding and subtracting $Y(q^{LF}, w^\star(q^{LF}))$. $R \geq 0$ because $w^\star(q)$ maximises $Y(q, \cdot)$ by definition, so $Y(q, w^\star) \geq Y(q, q)$. $D \leq 0$ is established heuristically in Section 3.3.2 for the empirically relevant case $w^\star(q^{LF}) > q^{LF}$; the formal proof is in appendix 3.B.4. □

The redistribution gain is strictly positive ($R > 0$) whenever collateral constraints bind, because redistribution can relax binding constraints and raise output. $R = 0$ only when no constraints bind and redistribution has no effect on output.

Overshooting

The distortion cost arises because agents respond to anticipated redistribution by shifting their technology choice in the direction of the redistribution. At laissez-faire, $w = q$ and agents are indifferent between technologies: $\Phi(q^{LF}, q^{LF}) = 0$, where $\Phi(q, w) := U_d(q, w) - U_v(q, w)$ denotes the utility differential. The planner's redistribution shifts w away from q; agents respond through two channels:

1. *Profit channel:* redistribution relaxes the favored type's borrowing constraint, raising its expected profits.

2. *Wealth channel:* the direct asset transfer raises the favored type's expected gross return on wealth.

The planner's redistribution decision accounts only for the profit channel (which determines output), but agents value both. The utility differential is strictly monotone in w ($\partial\Phi/\partial w > 0$), so any redistribution away from $w = q$ breaks indifference in the direction of the favored type.[6] Agents shift technology choice until indifference is restored at a new q^{NC} displaced from q^{LF} in the same direction. The following lemma formalizes this displacement.

Lemma 3.1 (Overshooting). *The no-commitment equilibrium technology composition is distorted in the direction of the planner's redistribution:* $(q^{NC} - q^{LF})$ *and* $(w^\star(q^{LF}) - q^{LF})$ *have the same sign.*

[6] The monotonicity of Φ in w can not be shown analytically, but is verified numerically across the parameter grid.

The formal proof is in appendix 3.B.4.

At the planner's optimum $w^\star$, the aggregate output gain from further redistribution is zero ($\partial Y/\partial w|_{w^\star} = 0$), but agents still perceive a strictly positive gain ($\partial \Phi/\partial w|_{w^\star} > 0$): the per-capita weighting in the utility differential prevents cancellation of profit effects, and the wealth channel adds a positive term with no counterpart in $\partial Y/\partial w$. Any output-optimal redistribution therefore necessarily distorts technology choice.

Distortion cost

Does the overshoot reduce welfare? It does, under risky output dominance (Definition 3.2). Define $V(q) := Y(q, w^\star(q))$, welfare along the planner's best response. Risky output dominance implies $V'(q^{LF}) \leq 0$: marginally raising q from the laissez-faire level lowers welfare along V, because the planner is already optimizing w for any given q, and the residual effect on output runs through the technology mix alone. In the empirically relevant case where the planner redistributes toward the safe type, $w^\star(q^{LF}) > q^{LF}$ and overshooting implies $q^{NC} > q^{LF}$—moving q in exactly the welfare-reducing direction. This holds whenever the safe type is the only constrained type, which is the generic pattern in the parameterization of Section 3.3.3.

$Y(q, w)$ is jointly concave in (q, w) on the feasibility set (appendix, Theorem 3.1); by the partial-maximum theorem, V is concave. By the envelope theorem, $V'(q) = \partial Y/\partial q|_{w=w^\star}$. The output factorization (appendix, Lemma 3.2) gives $\partial Y/\partial q = (1-\alpha)\mathbb{E}[y_d^\star - y_v^\star]$, so $V'(q^{LF}) \leq 0$ whenever risky output dominance (Definition 3.2) holds at $(q^{LF}, w^\star(q^{LF}))$.

Risky output dominance holds because the risky technology carries a productivity premium ($\bar{z}_v > z_d$) that is not overcome by the constraint-induced disadvantage at moderate leverage. This condition is not proved in closed form; it is verified numerically across ∼22,000 interior parameterizations spanning the empirically relevant parameter range ($\theta \leq 0.20$, $\lambda \in [1.20, 1.95]$). Near the feasibility boundary $q^{LF} \to 1/\lambda$, risky output dominance fails and $D > 0$; Section 3.3.3 characterizes this exception.

Since V is concave and $V'(q^{LF}) \leq 0$, V is non-increasing on $[q^{LF}, \infty)$. Combined with overshooting ($q^{NC} > q^{LF}$), this gives $D = V(q^{NC}) - V(q^{LF}) \leq 0$. The formal proof is in appendix 3.B.4.

Magnitude comparison. The analysis establishes $R \geq 0$ and $D \leq 0$, so $\Delta Y > 0$ if and only if $R > |D|$. The analytical framework pins down the structure of the welfare comparison; the magnitude comparison is characterized numerically.

3.3.3 Numerical characterization

Parameterization. I use a mean-preserving spread parameterization for the risky technology's productivity. Let $\bar{z}_v$ denote the mean productivity and $s \geq 0$ the spread parameter, so that:

$$\begin{aligned} z_h &= \bar{z}_v + \theta s, \\ z_\ell &= \bar{z}_v - (1-\theta)s. \end{aligned} \tag{3.25}$$

This preserves $\mathbb{E}[z_v] = \bar{z}_v$ for any s. The baseline parameters are $\alpha = 0.6$ (corresponding to a demand elasticity of $1/(1-\alpha) = 2.5$ when interpreting the decreasing returns as arising from a monopolistic competition demand system), $\theta = 0.05$, $z_d = 1$, $\bar{z}_v = 1.07$, $a = 1$. The baseline lies in constraint configuration (iii) (safe type constrained in state ℓ, risky type constrained in state h). The assumption $\bar{z}_v > z_d$ ensures a productivity premium for the risky technology. The feasibility set $\mathcal{F}$ (Definition 3.1) requires $\lambda < 2$; the analysis covers $\lambda \in [1.20, 1.95]$. In a related model, Guntin and Kochen (2024) calibrate $\lambda \approx 1.7$ for U.S. firms, while Moll (2014) uses $\lambda = 1.2$ for India.

Grid. Table 3.1 summarises the parameter grid. Of the $30 \times 25 \times 4 \times 4 \times 3 = 36{,}000$ main-grid combinations, 14,062 yield interior parameterizations (at least one collateral constraint binds at the laissez-faire allocation). Together with an additional boundary-focused oversample concentrated near the feasibility boundary ($q^{LF} \to 1/\lambda$), the combined verification set contains 22,396 interior parameterizations.

Table 3.1: Parameter grid for the numerical welfare analysis. "Interior" excludes parameterizations in which no collateral constraint binds at the laissez-faire allocation (configuration (i)), for which $R = D = \Delta Y = 0$ trivially.

Parameter	Description	Grid	Values / range
λ	Leverage limit	Fine (30 pts)	$[1.20, 1.95]$
s	Spread	Fine (25 pts)	$[0.10, s_{\max}(\theta, \bar{z}_v)]^a$
θ	Disaster probability	Discrete (4 pts)	$\{0.02, 0.05, 0.10, 0.20\}$
α	Output elasticity	Discrete (4 pts)	$\{0.4, 0.5, 0.6, 0.7\}$
$\bar{z}_v$	Mean risky productivity	Discrete (3 pts)	$\{1.03, 1.07, 1.15\}$

[a] $s_{\max}(\theta, \bar{z}_v) = (\bar{z}_v - 0.005)/(1 - \theta)$, the largest spread consistent with $z_\ell > 0$.

Results. Of the 22,396 interior parameterizations, 92.87% have $\Delta Y < 0$: laissez-faire dominates. At the median parameterization, the distortion cost $|D|$ exceeds the redistribution gain R by a factor of approximately four (median $R/|D| = 0.258$).

Figure 3.1 shows the welfare decomposition as a function of λ at baseline ($s = 1.121$, $z_\ell = 0.005$, $z_h = 1.126$). It plots R, D, and $\Delta Y = R + D$: $D \leq 0$ throughout, and $\Delta Y < 0$ for λ below approximately 1.83. Near the upper feasibility boundary the redistribution gain R and distortion cost $|D|$ converge, and ΔY turns slightly positive. Risky output dominance holds across this range ($V'(q^{LF}) \leq 0$, verified numerically), confirming that $D \leq 0$ follows from the concavity and overshooting arguments.

Figure 3.2 maps $\Delta Y/|Y^{LF}|$ (in percent) across the (λ, s) parameter space at fixed $\theta = 0.05$. Red shading indicates $\Delta Y < 0$ (laissez-faire dominates). The white region at high λ and low s is configuration (i), where no collateral constraint binds and redistribution is inert ($\Delta Y = 0$ exactly). The exception ($\Delta Y > 0$) appears only at high s ($z_\ell \approx 0$), where risky output dominance fails and D turns positive.

Figure 3.3 shows the decomposition as a function of θ at $\lambda = 1.40$ (interior) and $\lambda = 1.65$ (near exception). At baseline spread $s = 1.121$,

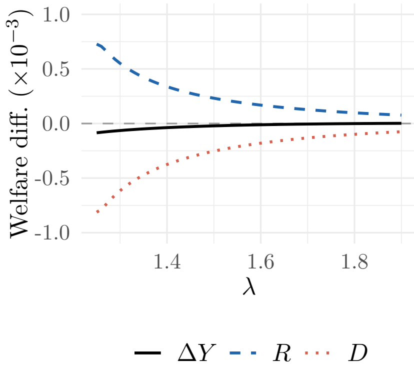


Figure 3.1: Welfare decomposition as a function of λ for baseline parameters ($\theta = 0.05$, $\bar{z}_v = 1.07$, $s = 1.121$). Redistribution gain $R \geq 0$ (blue dashed), distortion cost $D \leq 0$ (red dotted), and net welfare difference $\Delta Y = R + D$ (black solid), scaled by 10^3.

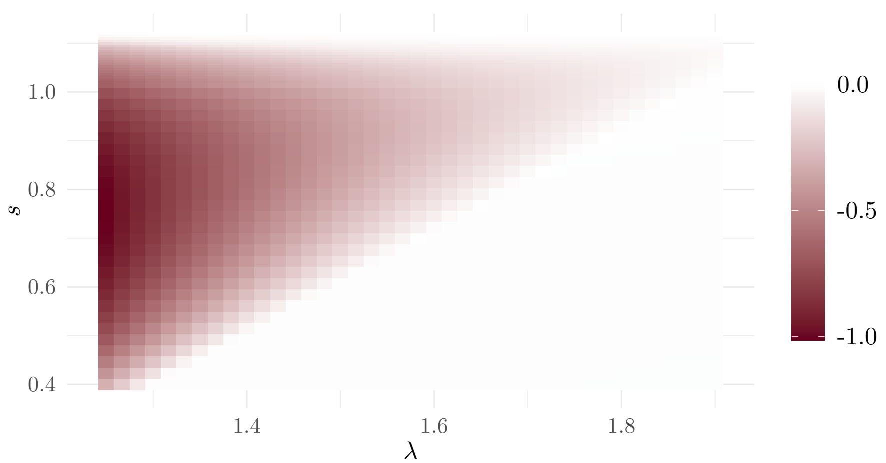
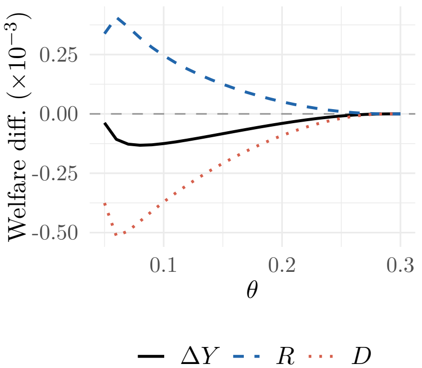
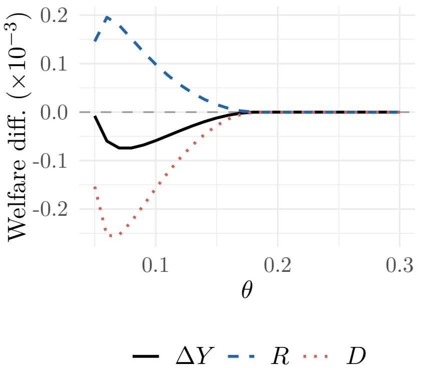


Figure 3.2: Welfare difference $\Delta Y/|Y^{LF}|$ (in percent) over the (λ, s) parameter space ($\theta = 0.05$, $\bar{z}_v = 1.07$). Red shading: $\Delta Y < 0$ (laissez-faire dominates). The white region at high λ and low s is configuration (i), where no collateral constraint binds and redistribution has no effect ($\Delta Y = 0$ exactly). The exception ($\Delta Y > 0$) is concentrated at high s, where $z_\ell \approx 0$ and risky output dominance fails.

the exception does not appear as θ varies: $\Delta Y \leq 0$ throughout. For $\lambda = 1.65$, ΔY approaches zero from below as θ increases and equals zero at high θ, where the collateral constraint is slack and the NC equilibrium coincides with laissez-faire. This confirms that the boundary exception ($\Delta Y > 0$) requires severe downside risk (high s, so $z_\ell \approx 0$); varying θ at baseline spread is insufficient to trigger it.



(a) $\lambda = 1.40$



(b) $\lambda = 1.65$

Figure 3.3: Welfare decomposition as a function of θ for baseline $s = 1.121$ and $\bar{z}_v = 1.07$. $\Delta Y \leq 0$ throughout: the boundary exception ($\Delta Y > 0$) does not appear at baseline spread. For $\lambda = 1.65$, ΔY reaches zero at high θ where the collateral constraint is slack and the NC equilibrium coincides with laissez-faire.

Exceptions. The small share of interior parameterizations with $\Delta Y > 0$ arise through two distinct mechanisms.

Boundary exception. Consider the limiting case $z_\ell \to 0$ (equivalently, $s \to \bar{z}_v/(1-\theta)$). As the low-state productivity of the risky technology vanishes, risky output $y_v^\star(\ell) \to 0$ while safe output $y_d^\star(\ell)$ remains bounded. Agents flee to safety, pushing q^{LF} to the upper feasibility boundary $1/\lambda$. At this corner, the interior first-order condition for constrained efficiency becomes a strict KKT inequality: more safe production would be beneficial, but the collateral constraint prevents it. Risky output dominance fails ($V'(q^{LF}) > 0$), so the arguments of

Section 3.3.2 do not apply. The planner has no room to redistribute ($w^\star \approx q^{LF} \approx 1/\lambda$, so $R \to 0$), but the no-commitment overshoot—agents anticipating redistribution choose more safety—accidentally pushes q^{NC} in the welfare-improving direction, yielding $D > 0$ and $\Delta Y > 0$.

Regime-switching exception (undershooting). When the safe-type collateral constraint is barely binding at the LF equilibrium ($w^\star(q^{LF})$ lies only marginally above q^{LF}), the planner's best response lifts the constraint entirely, shifting the capital market from the partially constrained configuration into the unconstrained regime. In the unconstrained regime both types face the same interest rate, so expected utility differentials are governed by productivities alone. Since $\bar{z}_v > z_d$ by assumption, the risky sector has the higher expected productivity and $\mathbb{E}[y_v^\star] > \mathbb{E}[y_d^\star]$: agents prefer the risky technology. The NC equilibrium therefore settles at $q^{NC} < q^{LF}$ (undershooting). Although unintuitive, this is welfare-improving: since $V'(q^{LF}) \leq 0$ the LF equilibrium has too many safe agents relative to the constrained optimum, and the undershooting corrects rather than amplifies that distortion, giving $D = V(q^{NC}) - V(q^{LF}) > 0$.

The incidence of undershooting is monotone in two primitives. It rises with the probability of the bad state θ: a higher θ amplifies the low-state interest-rate feedback and expands the region in which the constraint is barely binding at the LF allocation, so that a small redistribution suffices to shift regimes. It also rises with mean risky productivity $\bar{z}_v$: a higher $\bar{z}_v$ strengthens the productivity advantage the risky sector enjoys once the constraint is lifted, amplifying agents' migration away from the safe technology. Both conditions are required: without a barely-binding constraint the regime shift does not occur; without a sufficiently high $\bar{z}_v$ the productivity differential in the unconstrained regime is too small to reverse agents' technology choice.

3.3.4 Interpretation

The generic result and its magnitude. Across the empirically relevant parameter range, laissez-faire dominates no-commitment intervention. The distortion cost is not marginal: it exceeds the redistribution gain by approximately four-fold at the median parameterization. The overshooting mechanism is systematic. Any output-improving redistribution triggers a technology-composition response whose welfare cost exceeds the redistribution gain, and it operates in all three constraint configurations where redistribution is non-trivial. The overall welfare losses are meaningful but not large: Figure 3.2 shows that the relative welfare loss is almost always smaller than 1%.

Which primitives drive the result. Laissez-faire dominance is strongest at low λ (tight constraints), moderate s (interior spread), and low θ (rare disasters). In this region the wealth channel is large relative to the profit channel: tight constraints amplify the direct asset-transfer effect on agents' utility, while the output-improving effect of redistribution is limited by the planner's inability to simultaneously relax both types' constraints.

What the exceptions reveal. The boundary exception ($z_\ell \approx 0$) shows that when the economy is jammed against the feasibility boundary, intervention helps because there is little room for the distortion to operate ($R \approx 0$, and the overshoot accidentally goes in the welfare-improving direction). The undershooting exception shows that when the constraint is barely binding, a small redistribution can lift it entirely, and the resulting regime switch reverses agents' technology response. Both are edge cases, but they sharpen the characterization of when intervention helps: precisely when the standard overshooting mechanism is disrupted.

3.4 Discussion and Conclusion

In this paper, I study technology choice under uncertainty, where agents face a trade-off between a safe low-return technology and a risky high-return technology. I embed this into a setting with a limitation to the ex post adjustment of production factors in the form of a financial friction, or collateral constraint. The limitation creates an opportunity for industrial policy in the form of an ex ante tax or subsidy on assets, conditional on agents' technology choices. This redistribution of assets between types of agents allows the planner to influence the allocation of capital across technologies, potentially alleviating borrowing constraints and improving overall welfare. Anticipated industrial policy changes agents' incentives though, and I show that in general, this time-inconsistency problem and lack of commitment leads to lower welfare than in the laissez-faire case. Numerical analysis shows that the no-commitment policy dominates laissez-faire only for a narrow range of parameters, in particular when the risky technology has almost zero productivity in the bad state of the world.

The time-inconsistency problem arises since the planner wants to change two margins (the distribution of entrepreneurs over technologies, and the distribution of assets across types), but has only one policy instrument. An additional instrument, like ex post redistribution of consumption could alleviate the problem. Such policy may be infeasible though, since the planner is indifferent to such a policy ex post. Any positive cost of the policy would lead to no redistribution ex post, and thus no commitment power ex ante.

The paper presents a sobering result, in that for relevant parameter values, and under the relevant assumption of no-commitment, industrial policy may reduce welfare compared to laissez-faire. The planner has no commitment power both in engaging in redistribution, but also in refraining from it.

In a rules-based setting, the simple rule of "Do not intervene" would be optimal for the planner.

Mapping to real-world cases. The model maps to different real-world episodes of supply disruption. China's 2010 rare earth restrictions caused price spikes but not complete supply cutoff: Japan and other importers faced higher costs but continued production, corresponding to moderate z_ℓ in the model. The 2022 Russian gas cutoff to Europe represents a more severe case: while not a complete cessation (some gas continued to flow through alternative routes), energy-intensive industries faced dramatic cost increases, corresponding to lower z_ℓ. Recent graphite export restrictions by China similarly raise costs for battery manufacturers without eliminating supply entirely.

The numerical result that no-commitment intervention dominates laissez-faire only when $z_\ell \approx 0$ has a sharp interpretation: government redistribution improves welfare only when disruptions approach complete supply cutoff. For the more common case of price increases or partial restrictions, the distortion from anticipated intervention outweighs the direct efficiency gain. This suggests that preemptive industrial policy is most justified for inputs where complete cutoff is a realistic threat—but these are precisely the cases where governments may lack commitment power, as the political pressure to intervene would be overwhelming.

Extensions. The model presented above is static, and a dynamic extension is natural. Two main aspects would be interesting to explore. First, asset accumulation over time might exacerbate the downside of the bad state realizing: if risky-type agents accumulated assets in good states, the safe type may be even more constrained in bad states. Second, repeated interactions between the planner and agents may allow for more sophisticated policy designs that can mitigate the time-inconsistency problem. This is left to future work.

References

Alpanda, S., & Peralta-Alva, A. (2010). Oil crisis, energy-saving technological change and the stock market crash of 1973–74. *Review of Economic Dynamics*, *13*(4), 824–842.

Baldwin, R., & Freeman, R. (2022). Risks and global supply chains: What we know and what we need to know. *Annual Review of Economics*, *14*(1), 153–180.

Barrot, J.-N., & Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics*, *131*(3), 1543–1592.

Capponi, A., Du, C., & Stiglitz, J. E. (2024, March). *Are supply networks efficiently resilient?* (Working Paper No. 32221). NBER.

Chari, V. V., & Kehoe, P. J. (2016). Bailouts, time inconsistency, and optimal regulation: A macroeconomic view. *American Economic Review*, *106*(9), 2458–2493.

Gholz, E., & Hughes, L. (2021). Market structure and economic sanctions: The 2010 rare earth elements episode as a pathway case of market adjustment. *Review of International Political Economy*, *28*(3), 611–634.

Grossman, G. M., Helpman, E., & Lhuillier, H. (2023). Supply chain resilience: Should policy promote international diversification or reshoring? *Journal of Political Economy*, *131*(12), 3462–3496.

Guntin, R., & Kochen, F. (2024, August). *Financial frictions and the market for firms* (Working Paper).

Kopytov, A., Mishra, B., Nimark, K., & Taschereau-Dumouchel, M. (2024). Endogenous production networks under supply chain uncertainty. *Econometrica*, *92*(5), 1621–1659.

Midrigan, V., & Xu, D. Y. (2014). Finance and misallocation: Evidence from plant-level data. *American Economic Review*, *104*(2), 422–458.

Moll, B. (2014). Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review*, *104*(10), 3186–3221.

Moll, B., Schularick, M., & Zachmann, G. The power of substitution: The great german gas debate in retrospect. In: In *Brookings papers on economic activity*. 2023, September.

Morrison, W. M., & Tang, R. (2012, April). *China's rare earth industry and export regime: Economic and trade implications for the United States* (CRS Report for Congress No. R42510). Congressional Research Service.

The third Gulf war will scar energy markets for a long time yet. (2026). *The Economist*. Retrieved April 11, 2026, from https://www.economist.com/finance-and-economics/2026/04/08/the-third-gulf-war-will-scar-energy-markets-for-a-long-time-yet

Traiberman, S., & Rotemberg, M. (2023). Precautionary protectionism. *Journal of International Economics*, *145*, 103836.

Wei, C. (2003). Energy, the stock market, and the putty-clay investment model. *American Economic Review*, *93*(1), 311–323.

Why China is restricting exports of graphite. (2023). *The Economist*. Retrieved January 21, 2026, from https://www.economist.com/business/2023/10/25/why-china-is-restricting-exports-of-graphite

Appendices

Appendix 3.A Model

3.A.1 Micro-foundation: Materials prices and productivity

This section provides a micro-foundation for the reduced-form productivity process in the main text, showing that variation in materials prices maps into variation in effective productivity.

3.A.1.1 Setup

Consider a production function that is CES in capital k and materials m:

$$y = f(k,m) = \left(k^{\frac{\sigma-1}{\sigma}} + m^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \tag{3.A.1}$$

where $\sigma \in (0,1)$ is the elasticity of substitution. The restriction $\sigma < 1$ means capital and materials are complements—a natural assumption for many production processes where materials are essential inputs.

3.A.1.2 Optimal materials choice

Suppose the firm takes the price of materials b as given (with output price normalized to $p = 1$) and chooses materials to maximize profits for given capital k:

$$\max_m \; y - bm.$$

The first-order condition yields $m^\star = (1/b)^\sigma y$. Substituting into the production function gives the reduced-form output as a function of capital alone:

$$y = z(b)k, \quad \text{where} \quad z(b) = \left(1 - b^{1-\sigma}\right)^{\frac{\sigma}{1-\sigma}}. \tag{3.A.2}$$

We restrict attention to $b \in (0,1)$ (which is equivalent to positive profits) so that $z(b)$ is well-defined.

3.A.1.3 Application to the model

Consider two technologies:

- **Safe technology**: Materials always available at price $\bar{b}$, giving constant productivity $z_d = z(\bar{b})$.
- **Risky technology**: Materials price varies with the aggregate state. In state h, materials are cheap ($b^h < \bar{b}$), giving high productivity $z_h = z(b^h) > z_d$. In state ℓ, materials are expensive ($b^\ell > \bar{b}$), giving low productivity $z_\ell = z(b^\ell) < z_d$.

This maps directly to the productivity process in the main text: the "risky" technology corresponds to supply chains exposed to price volatility (e.g., imported materials subject to trade disruptions), while the "safe" technology corresponds to stable domestic supply.

Remark on functional form. The reduced-form production function (3.A.2) is linear in capital, while the main text uses $y = z^{1-\alpha}k^\alpha$ with decreasing returns. The micro-foundation establishes the key economic mechanism of materials price variability translating into productivity variability. The decreasing returns specification in the main text can be interpreted as arising from monopolistic competition with CES demand.

Appendix 3.B Proofs and Derivations

3.B.1 Proofs for decentralized equilibrium

Proposition 3.1 (Unique equilibrium)**.** *There exists a unique decentralized equilibrium* $q^{LF} \in [0, 1]$.

Proof. If the solution is a corner, it is necessarily unique. Consider the interior case. Define $F(q) = U(d, q, a) - U(v, q, a)$. By the intermediate value theorem, there exists a unique root $q^{LF} \in (0, 1)$ if F is strictly monotonic.

Since gross returns to assets are equal for both types:

$$F(q) = \mathbb{E}\left[\pi(r(S,q), z(S,d), a) - \pi(r(S,q), z(S,v), a)\right].$$

Taking the derivative:

$$F'(q) = \mathbb{E}\left[\left(-k^\star(r(S,q), z(S,d), a) + k^\star(r(S,q), z(S,v), a)\right)\frac{\partial r(S,q)}{\partial q}\right].$$

Using the implicit function theorem on the capital market clearing condition:

$$\frac{\partial r(S,q)}{\partial q} = \left[k_v^\star(r(q)) - k_d^\star(r(q))\right]\left[q\frac{\partial k_d^\star}{\partial r} + (1-q)\frac{\partial k_v^\star}{\partial r}\right]^{-1}.$$

Since $\frac{\partial k_T^\star}{\partial r} \leq 0$ and at least one type is unconstrained (else no market clearing), the denominator is strictly negative. Thus, $\operatorname{sign}\frac{\partial r}{\partial q} = (-1)\operatorname{sign}[k_v^\star - k_d^\star]$.

Then:

$$F'(q) = \mathbb{E}\left[\left(k_v^\star - k_d^\star\right)\frac{\partial r}{\partial q}\right] < 0,$$

since both factors always have opposite sign. This proves strict monotonicity and uniqueness. □

Proposition 3.2 (Constrained efficiency)**.** *The laissez-faire technology share q^{LF} is constrained efficient: it maximizes $Y(q,q)$ subject to $w = q$.*

Proof. On the diagonal $w = q$, capital-market clearing $qk_d^\star + (1-q)k_v^\star = a$ and zero redistribution give $qU(d,q,a) + (1-q)U(v,q,a) = Y(q,q) + a$, so stationary points of $Y(q,q)$ coincide with those of the agent-welfare objective $qU(d,q,a) + (1-q)U(v,q,a)$. The FOC of the latter is

$$0 = U(d,q,a) - U(v,q,a) + q\frac{\partial U(d,q,a)}{\partial q} + (1-q)\frac{\partial U(v,q,a)}{\partial q}.$$

This equals the decentralized condition if the last two terms sum to zero.

Computing:

$$\frac{\partial U(T,q,a)}{\partial q} = \mathbb{E}\left[(-k^\star(r(S,q),z(S,T),a)+a)\,\frac{\partial r(S,q)}{\partial q}\right].$$

Then:

$$\begin{aligned} & q\frac{\partial U(d,q,a)}{\partial q} + (1-q)\frac{\partial U(v,q,a)}{\partial q} \\ &= \mathbb{E}\left[\underbrace{(-qk_d^\star - (1-q)k_v^\star + a)}_{=0 \text{ by market clearing}}\frac{\partial r}{\partial q}\right] = 0. \end{aligned}$$

By Theorem 3.1, $Y(q,q)$ is concave in q, so the stationary point is a global maximum. ☐

3.B.2 Unconstrained benchmark and CFB comparison

Unconstrained benchmark. Without the financial friction ($\lambda \to \infty$), the planner solves $\max_q \mathbb{E}[(qz_d + (1-q)z(S,v))^{1-\alpha}]$. The objective is globally concave. The interior solution is

$$q^{\text{uc}} = \frac{z_h}{z_h - z_d + \frac{(z_h - z_\ell)z_d}{z_h\left(\frac{z_d - z_\ell}{z_h - z_d}\frac{\theta}{1-\theta}\right)^{\frac{1}{\alpha}} - z_\ell}}. \tag{3.B.1}$$

CFB comparison.

Proposition 3.5 (CFB versus unconstrained benchmark)**.**

1. *In configuration (ii) (redistribution can fully alleviate constraints),* $q^{CFB} = q^{uc}$*.*

2. *In configuration (iii) (both types constrained),* $q^{CFB} \neq q^{uc}$ *generically, with* $q^{CFB} > q^{uc}$ *if and only if* $w^{CFB} < q^{CFB}$ *(redistribution toward the risky type).*

Proof. Part 1. In configuration (ii), there exists $w^\star$ such that no constraint binds. At this $w^\star$, the CFB first-order condition for q reduces to $\mathbb{E}[y_d^\star -$

$y_v^\star] = 0$, because the marginal output with respect to assets is zero for both types in all states. When no constraints bind, per-agent output is $y_T^\star = z(S,T)^{1-\alpha} a^\alpha / (qz_d + (1-q)z(S,v))^\alpha$, and the FOC is equivalent to the unconstrained benchmark FOC.

Part 2. In configuration (iii), $w^\star(q)$ keeps at least one constraint binding for every feasible q. When a type is constrained, its per-agent output $y_T^\star = z(S,T)^{1-\alpha}(\lambda \tilde{a}_T)^\alpha$ depends on w through $\tilde{a}_d = wa/q$ and $\tilde{a}_v = (1-w)a/(1-q)$. The expected output gap $\mathbb{E}[y_d^\star - y_v^\star]$ is therefore strictly monotone in w, so the constrained gap evaluated along $w^\star(q)$ differs from the unconstrained gap. Since the unconstrained benchmark satisfies $\mathbb{E}[y_d^{uc} - y_v^{uc}] = 0$ at q^{uc}, and the constrained gap at q^{uc} is generically nonzero, $q^{CFB} \neq q^{uc}$.

For the direction: if $w^{CFB} < q^{CFB}$, the safe type holds fewer per-capita assets than the risky type, shifting $\mathbb{E}[y_d^\star - y_v^\star]$ downward relative to equal per-capita assets ($w = q$). Since the output gap is also strictly decreasing in q, the zero of the constrained gap lies above q^{uc}, giving $q^{CFB} > q^{uc}$. The converse follows symmetrically. □

3.B.3 Concavity of the welfare function

Theorem 3.1 (Joint concavity of Y)**.** *$Y(q,w)$ is jointly concave in (q,w) on the feasibility set $\mathcal{F}$.*

Proof. Since $Y = (1-\theta)Y_h + \theta Y_\ell$ is a non-negative linear combination of state-level outputs, it suffices to show each Y_S is jointly concave in (q,w), as concavity is preserved under non-negative linear combinations. The proof proceeds case by case over the constraint configurations. In each case we compute the Hessian H_S of Y_S in (q,w) and verify that it is negative semi-definite, which for symmetric 2×2 matrices reduces to checking that both diagonal entries are non-positive and that $\det(H_S) \geq 0$.

Case 1: Both types unconstrained. $Y_S = \bar{z}_S^{1-\alpha}$ with $\bar{z}_S = qz_d + (1-q)z_{v,S}$, which is independent of w. The Hessian is therefore $H_S = \operatorname{diag}\big(-\alpha(1-\alpha)\bar{z}_S^{-\alpha-1}(z_d - z_{v,S})^2,\ 0\big)$, which is negative semi-definite.

Case 2: Safe type (d) constrained, risky type (v) unconstrained. Capital allocations are $k_d = \lambda w/q$ and $k_v = (1-\lambda w)/(1-q)$ (with $a = 1$), giving

$$Y_S = z_d^{1-\alpha}(\lambda w)^\alpha q^{1-\alpha} + z_{v,S}^{1-\alpha}(1-\lambda w)^\alpha (1-q)^{1-\alpha}.$$

It is convenient to abbreviate the per-type output flows by

$$y_d^\star = z_d^{1-\alpha}\left(\frac{\lambda w}{q}\right)^\alpha, \qquad y_v^\star = z_{v,S}^{1-\alpha}\left(\frac{1-\lambda w}{1-q}\right)^\alpha,$$

so that $Y_S = q y_d^\star + (1-q) y_v^\star$.

We compute the three independent second-order partial derivatives. The two diagonal terms are

$$\frac{\partial^2 Y_S}{\partial q^2} = -\alpha(1-\alpha)\left[\frac{y_d^\star}{q} + \frac{y_v^\star}{1-q}\right] < 0,$$
$$\frac{\partial^2 Y_S}{\partial w^2} = -\alpha(1-\alpha)\lambda^2\left[\frac{y_d^\star \cdot q}{(\lambda w)^2} + \frac{y_v^\star(1-q)}{(1-\lambda w)^2}\right] < 0.$$

For the cross-partial, write $A_d = z_d^{1-\alpha}(\lambda w)^\alpha$ and $A_v = z_{v,S}^{1-\alpha}(1-\lambda w)^\alpha$, so that $Y_S = A_d q^{1-\alpha} + A_v(1-q)^{1-\alpha}$ and

$$\frac{\partial Y_S}{\partial q} = (1-\alpha)\left[A_d q^{-\alpha} - A_v(1-q)^{-\alpha}\right].$$

Differentiating with respect to w acts only on A_d and A_v, with $\partial A_d/\partial w = \alpha\lambda z_d^{1-\alpha}(\lambda w)^{\alpha-1}$ and $\partial A_v/\partial w = -\alpha\lambda z_{v,S}^{1-\alpha}(1-\lambda w)^{\alpha-1}$, yielding

$$\frac{\partial^2 Y_S}{\partial q \partial w} = \alpha(1-\alpha)\lambda\left[z_d^{1-\alpha}(\lambda w)^{\alpha-1}q^{-\alpha} + z_{v,S}^{1-\alpha}(1-\lambda w)^{\alpha-1}(1-q)^{-\alpha}\right].$$

Recognizing $z_d^{1-\alpha}(\lambda w)^{\alpha-1}q^{-\alpha} = y_d^\star/(\lambda w)$ and $z_{v,S}^{1-\alpha}(1-\lambda w)^{\alpha-1}(1-q)^{-\alpha} = y_v^\star/(1-\lambda w)$, this simplifies to

$$\frac{\partial^2 Y_S}{\partial q \partial w} = \alpha(1-\alpha)\lambda\left[\frac{y_d^\star}{\lambda w} + \frac{y_v^\star}{1-\lambda w}\right] > 0.$$

The cross-partial is strictly positive, so the Hessian is not diagonal and the sign of $\det(H_S)$ is not immediately obvious.

Factoring the common scalar $[\alpha(1-\alpha)\lambda]^2$ out of the determinant yields

$$\det(H_S) = [\alpha(1-\alpha)\lambda]^2 \times \left\{ \left[\frac{y_d^\star}{q} + \frac{y_v^\star}{1-q}\right]\left[\frac{y_d^\star q}{(\lambda w)^2} + \frac{y_v^\star(1-q)}{(1-\lambda w)^2}\right] - \left[\frac{y_d^\star}{\lambda w} + \frac{y_v^\star}{1-\lambda w}\right]^2 \right\}.$$

The bracketed expression has the form $|\mathbf{u}|^2|\mathbf{v}|^2 - (\mathbf{u}\cdot\mathbf{v})^2$ for a suitable choice of vectors. Specifically, define

$$\mathbf{u} = \left(\frac{\sqrt{y_d^\star}}{\sqrt{q}}, \frac{\sqrt{y_v^\star}}{\sqrt{1-q}}\right), \qquad \mathbf{v} = \left(\frac{\sqrt{y_d^\star}\sqrt{q}}{\lambda w}, \frac{\sqrt{y_v^\star}\sqrt{1-q}}{1-\lambda w}\right).$$

A direct calculation gives

$$\begin{aligned}
|\mathbf{u}|^2 &= \frac{y_d^\star}{q} + \frac{y_v^\star}{1-q}, \\
|\mathbf{v}|^2 &= \frac{y_d^\star q}{(\lambda w)^2} + \frac{y_v^\star(1-q)}{(1-\lambda w)^2}, \\
\mathbf{u}\cdot\mathbf{v} &= \frac{\sqrt{y_d^\star}}{\sqrt{q}} \cdot \frac{\sqrt{y_d^\star}\sqrt{q}}{\lambda w} + \frac{\sqrt{y_v^\star}}{\sqrt{1-q}} \cdot \frac{\sqrt{y_v^\star}\sqrt{1-q}}{1-\lambda w} = \frac{y_d^\star}{\lambda w} + \frac{y_v^\star}{1-\lambda w},
\end{aligned}$$

so the bracketed expression equals exactly $|\mathbf{u}|^2|\mathbf{v}|^2 - (\mathbf{u}\cdot\mathbf{v})^2$. The Cauchy–Schwarz inequality $|\mathbf{u}|^2|\mathbf{v}|^2 \geq (\mathbf{u}\cdot\mathbf{v})^2$ therefore gives $\det(H_S) \geq 0$, and combined with the strictly negative diagonal entries, H_S is negative semi-definite.

Case 3: Risky type (v) constrained, safe type (d) unconstrained. The substitution $\tilde{w} = 1-w$, $\tilde{q} = 1-q$ maps this case to Case 2 with types swapped. The Hessian determinant has the same form with λw replaced by $\lambda(1-w)$, so $\det(H_S) \geq 0$ by the same argument.

Cross-configuration boundaries. At points where the binding constraint changes, Y may be non-differentiable. Since each piece is jointly concave on its convex domain, the overall function Y is jointly concave on

$\mathcal{F}$. This is verified numerically across 6,400 parameter combinations spanning α, λ, s, and θ: all evaluated Hessians are negative semi-definite. $\square$

3.B.4 Proof of the distortion cost sign

Lemma 3.2 (Formula F)**.** *For (q, w) in the interior of $\mathcal{F} \times \mathcal{F}$ and any constraint configuration in state S,*

$$\frac{\partial Y}{\partial q} = (1-\alpha)\mathbb{E}_S[y_d^\star(S) - y_v^\star(S)]\,.$$

Proof. We differentiate $Y_S(q, w)$ case by case and take expectations; in each constraint configuration Y_S has a closed form in (q, w) from which the derivative follows by the chain rule.

Case 1: Both types unconstrained. Capital market clearing pins down a common interest rate r, so $k_T^\star = z_T(\alpha/r)^{1/(1-\alpha)}$ and market clearing $\sum q_T k_T^\star = 1$ gives $k_T^\star = z_T/\bar{z}_S$ with $\bar{z}_S = qz_d + (1-q)z_{v,S}$. Hence $y_T^\star = z_T(1/\bar{z}_S)^\alpha$ and $Y_S = \bar{z}_S^{1-\alpha}$. Differentiating, $\partial Y_S/\partial q = (1-\alpha)\bar{z}_S^{-\alpha}(z_d - z_{v,S}) = (1-\alpha)(y_d^\star - y_v^\star)$.

Case 2: Type $c \in \{d, v\}$ constrained, type $u \neq c$ unconstrained. Case 2 of Theorem 3.1 (Case 3 obtained by the substitution $w \leftrightarrow 1-w$, $q \leftrightarrow 1-q$) gives

$$\frac{\partial Y_S}{\partial q} = (1-\alpha)\big[A_d q^{-\alpha} - A_v(1-q)^{-\alpha}\big]\,,$$

with $A_d = z_d^{1-\alpha}(\lambda w^{(c)})^\alpha$ if $c = d$ and $A_d = z_d^{1-\alpha}(1-\lambda w^{(c)})^\alpha$ if $c = v$, and analogously for A_v (and $w^{(d)} = w$, $w^{(v)} = 1-w$). Recognizing $A_d q^{-\alpha} = y_d^\star$ and $A_v(1-q)^{-\alpha} = y_v^\star$ yields $\partial Y_S/\partial q = (1-\alpha)(y_d^\star - y_v^\star)$.

Taking expectations over S in Cases 1 and 2 gives the result. $\square$

Lemma 3.1 (Overshooting)**.** *The no-commitment equilibrium technology composition is distorted in the direction of the planner's redistribution: $(q^{NC} - q^{LF})$ and $(w^\star(q^{LF}) - q^{LF})$ have the same sign.*

Proof. Define the utility differential $\Phi(q, w) \coloneqq U_d(q, w) - U_v(q, w)$. In laissez-faire, no redistribution occurs ($w = q$), so the equilibrium condition is $\Phi(q^{LF}, q^{LF}) = 0$. In the no-commitment equilibrium, the planner redistributes optimally given q, so the equilibrium condition is $\Phi(q^{NC}, w^{\star}(q^{NC})) = 0$.

The utility differential is strictly increasing in the safe-type wealth share: $\partial\Phi/\partial w > 0$. This follows from the wealth channel: a higher w relaxes the collateral constraint of safe-type agents and tightens that of risky-type agents, raising their respective investment levels $\tilde{a}_d$ and lowering $\tilde{a}_v$, which increases U_d relative to U_v.

Combined with $\partial\Phi/\partial w > 0$ and $\Phi(q^{LF}, q^{LF}) = 0$:

$$\operatorname{sign}\big(\Phi(q^{LF}, w^{\star}(q^{LF}))\big) = \operatorname{sign}\big(w^{\star}(q^{LF}) - q^{LF}\big).$$

That is, at $q = q^{LF}$ under the planner's redistribution, the technology favored by the redistribution is strictly preferred whenever $w^{\star}(q^{LF}) \neq q^{LF}$.

By uniqueness of the laissez-faire equilibrium, $\Phi(\cdot, w^{\star}(\cdot))$ is strictly decreasing in q. If $\Phi(q^{LF}, w^{\star}(q^{LF})) > 0$ (planner redistributes toward safe), there exists a unique $q^{NC} > q^{LF}$ satisfying $\Phi(q^{NC}, w^{\star}(q^{NC})) = 0$. If $\Phi(q^{LF}, w^{\star}(q^{LF})) < 0$ (planner redistributes toward risky), there exists a unique $q^{NC} < q^{LF}$. In both cases, $(q^{NC} - q^{LF})$ and $(w^{\star}(q^{LF}) - q^{LF})$ have the same sign. □

Proposition 3.4 (Welfare decomposition)**.** *Suppose q^{LF} is interior in $\mathcal{F}$, risky output dominance (Definition 3.2) holds at $(q^{LF}, w^{\star}(q^{LF}))$, and $w^{\star}(q^{LF}) \neq q^{LF}$. Then the welfare difference decomposes as $\Delta Y = R + D$, where:*

$$\begin{aligned} R &:= Y(q^{LF}, w^{\star}(q^{LF})) - Y(q^{LF}, q^{LF}) \geq 0, \\ D &:= Y(q^{NC}, w^{\star}(q^{NC})) - Y(q^{LF}, w^{\star}(q^{LF})) \leq 0. \end{aligned} \tag{3.24}$$

The redistribution gain R measures the benefit of optimal redistribution holding technology composition fixed. The distortion cost D measures the welfare change due to the shift in equilibrium technology composition induced by anticipated redistribution.

Proof of Proposition 3.4: $D \leq 0$. By the envelope theorem, $V'(q) = \partial Y/\partial q\big|_{w=w^\star(q)}$, since $w^\star(q)$ maximizes $Y(q,\cdot)$. Lemma 3.2 gives $\partial Y/\partial q = (1-\alpha)\mathbb{E}[y_d^\star(S) - y_v^\star(S)]$ at any fixed w, so

$$V'(q^{LF}) = (1-\alpha)\mathbb{E}\left[y_d^\star(S) - y_v^\star(S)\right]\Big|_{(q^{LF},w^\star(q^{LF}))}.$$

This expression is non-positive if and only if $\mathbb{E}[y_v^\star(S)] \geq \mathbb{E}[y_d^\star(S)]$ at $(q^{LF}, w^\star(q^{LF}))$.[7] This is risky output dominance (Definition 3.2). The no-commitment equilibrium satisfies $q^{NC} > q^{LF}$ (Lemma 3.1). Since $V'(q^{LF}) \leq 0$ and V is concave (by the partial-maximum theorem applied to Theorem 3.1), V is non-increasing on $[q^{LF}, \infty)$, so

$$D = V(q^{NC}) - V(q^{LF}) \leq 0.$$

□

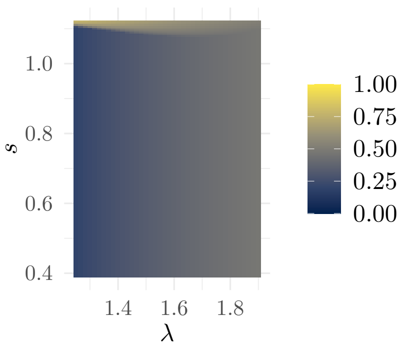
3.B.5 Equilibrium allocations over the (λ, s) parameter space

Figure 3.B.1 maps the equilibrium technology shares q^{LF}, q^{NC} and the no-commitment wealth share w^{NC} across the same (λ, s) grid as figure 3.2, together with the overshoot gap $q^{NC} - q^{LF}$ and the redistribution margin $w^{NC} - q^{NC}$.

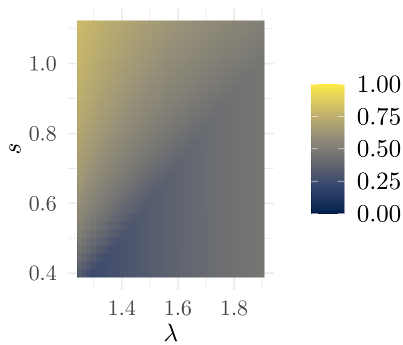
Three features stand out. First, the laissez-faire share q^{LF} varies considerably across the parameter space: it is low at low s and high λ, where the risky technology dominates, and rises steeply toward the feasibility ceiling $1/\lambda$ as s increases and $z_\ell \to 0$ drives agents toward safety. Second, the no-commitment share q^{NC} is much more uniform across most of the parameter space and only rises to match q^{LF} near $s_{\max}$. Third, the redistribution margin $w^{NC} - q^{NC}$ is everywhere small (order 10^{-3}), confirming that the planner barely deviates from $w = q$ at the NC equilibrium.

The near-equality $w^{NC} \approx q^{NC}$ is a consequence of incentive

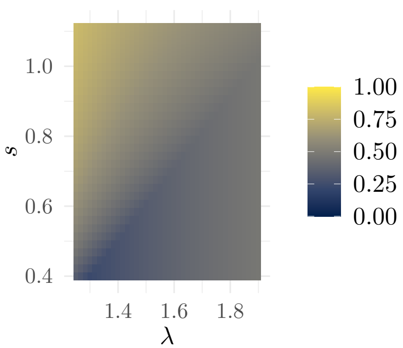
[7] This condition holds throughout the empirically relevant parameter range ($\theta \leq 0.1$, $\lambda \in [1.25, 1.90]$), verified numerically across ~40,000 interior parameterizations.



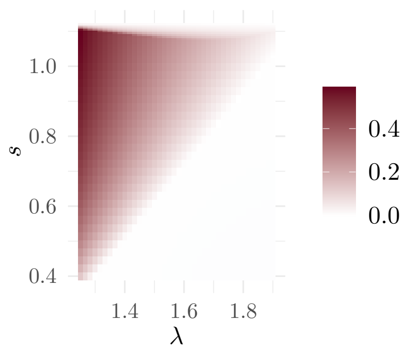
(a) q^{LF}



(b) q^{NC}



(c) w^{NC}



(d) $q^{NC} - q^{LF}$

Appendix Figure 3.B.1: Equilibrium allocations over the (λ, s) parameter space ($\theta = 0.05$, $\bar{z}_v = 1.07$). Panels (a)–(c): technology shares q^{LF}, q^{NC}, and wealth share w^{NC}, all on a common $[0, 1]$ color scale. Panel (d): overshoot gap $q^{NC} - q^{LF}$.

compatibility. The no-commitment equilibrium requires indifference: $\Phi(q^{NC}, w^{NC}) = 0$, where $\Phi = U_d - U_v$. Since $\partial\Phi/\partial w > 0$ (the wealth channel, more assets for the safe type make the safe technology strictly more attractive), any redistribution $w > q$ must be offset by a higher q to restore indifference. The planner *wants* to set $w > q$ to relax the collateral constraint, but agents respond by shifting into the safe technology until the utility differential is eliminated. Every unit of redistribution is absorbed by the technology-choice response, and the equilibrium slides up along the approximate locus $w \approx q$ to a higher (q, w) pair. The result is that the planner achieves $w^{NC} \approx q^{NC} \gg q^{LF}$: substantially more safe technology, but negligible effective redistribution.

This is the core of the time-inconsistency problem: the planner's instrument (w) and agents' choice (q) are tied together by incentive compatibility. The redistribution gain R that the planner *could* achieve at the laissez-faire technology share q^{LF} is large, but at the actual NC equilibrium the redistribution margin has collapsed to near zero. The overshoot gap $q^{NC} - q^{LF}$ is the technology distortion required to absorb the anticipated redistribution back into incentive compatibility.

These allocations also imply that the overshoot gap opens primarily because q^{LF} moves, not because q^{NC} moves. As s rises, q^{LF} climbs toward $1/\lambda$ while q^{NC} remains nearly flat. The anticipated redistribution insulates the NC equilibrium from the productivity shift, since agents who expect the planner to reallocate wealth toward the safe type find the safe technology attractive even when z_ℓ is relatively high. The gap therefore grows monotonically until both shares converge near $s_{\max}$.

3.B.6 Non-monotonicity of ΔY in s

The welfare loss $|\Delta Y|$ vanishes at both endpoints of s, for different reasons, producing the interior peak in Figure 3.2. At $s = 0$, the safe and risky technologies deliver identical state-contingent productivity, so all values of q yield the same welfare and $|\Delta Y| = 0$ trivially. At $s \to s_{\max}$, $z_\ell \to 0$ pins q^{LF} against the feasibility ceiling $1/\lambda$, and q^{NC}, bounded above by the same ceiling, is forced to converge to it; the overshoot gap closes and $|\Delta Y|$ collapses with it.

The decomposition makes both forces explicit. By concavity of V (Theorem 3.1, applied via the partial-maximum theorem) and $q^{NC} \geq q^{LF}$ in the relevant region (Lemma 3.1), a first-order expansion around q^{LF} gives

$$|\Delta Y| \approx |V'(q^{LF})| \cdot (q^{NC} - q^{LF}).$$

The slope $|V'(q^{LF})|$ vanishes as $s \to 0$, where the productivity wedge collapses and all technology choices are equivalent. The gap $q^{NC} - q^{LF}$ vanishes as $s \to s_{\max}$, where the feasibility ceiling binds. Both factors are strictly positive at intermediate s, and the product peaks where their joint contribution is largest. The mechanism behind the turnaround is the ceiling: it converts what would otherwise be a monotonically widening welfare loss into a function with an interior maximum.

Sammanfattning

Substitution mellan insatsvaror är en central del av produktionsteorin. John Hicks (i *The Theory of Wages*, 1932) och Joan Robinson (i *The Economics of Imperfect Competition*, 1933) introducerade substitution-selasticiteten för att beskriva i vilken utsträckning en insatsvara kan ersätta en annan vid produktion av en given produkt. Begreppet är en egenskap hos tekniken: det visar hur en ekonomi eller ett företag kommer att reagera på förändringar i de relativa kostnaderna för sina insatsvaror. Nästan ett sekel senare står samma parameter i centrum för en av vår tids största utmaningar: övergången till en ekonomi som är mindre beroende av koldioxid. Framgången för en övergång bort från fossila bränslen beror på om el från förnybara källor kan ersätta dem i produktionen.

Substitutionen sker dock inte omedelbart. Utrustningen har lång livslängd, leveranskedjorna är etablerade och teknikvalen är låsta. Det finns en klyfta mellan den reaktion som kan observeras på kort sikt och den som är möjlig på lång sikt, och storleken på denna klyfta är viktig av två skäl. Hur vi mäter elasticiteten beror på tidshorisonten för de data som finns tillgängliga. Och hur en omställning utvecklas—dess kostnad och hastighet—beror inte bara på det långsiktiga värdet utan också på hur snabbt ekonomin kan röra sig mot den. När substitution på kort sikt är svår, skiftar frågan från hur mycket ekonomin kan omorganisera sig till vad politiken kan göra medan den inte kan det.

Denna avhandling består av tre fristående kapitel, som vart och ett behandlar en aspekt av denna uppsättning frågor. Kapitel 1 undersöker hur stor substitutionsresponsen är mellan fossila bränslen och el inom den tyska tillverkningsindustrin. Kapitel 2 undersöker om långsiktiga elasticiteter kan härledas från kortvariga variationer när anpassning är kostsam. Kapitel 3 undersöker om förebyggande politik kan ersätta omfördelning efter en chock när sådan omfördelning är omöjlig.

I det första kapitlet, *Can the plants turn green?*, studerar jag elektrifieringspotentialen inom den tyska tillverkningssektorn. Det centrala resultatet är en elasticitet för substitution mellan fossila bränslen och el på 5 inom en enskild fabrik, vilket är betydligt större än befintliga skattningar på mikronivå på omkring 1,5.

Med hjälp av mikrodata på anläggningsnivå från den tyska tillverkningsstatistiken visar jag att fossila bränslen och el reagerar olika på förändringar i produktionsvolymen. När efterfrågechocker används som instrumentvariabler för produktionsförändingar, anpassas elanvändningen nästan tre gånger så mycket som användningen av fossila bränslen, och användningen av fossila bränslen uppvisar betydligt högre beständighet över tid. Denna asymmetri stämmer inte överens med standardantagandet om ett statiskt val av energiinput, men stämmer överens med förekomsten av anpassningsfriktioner som är specifika för användningen av fossila bränslen.

Jag utvecklar en strukturell produktionsfunktionsmodell med anpassningskostnader för användningen av fossila bränslen för att studera implikationerna av detta resultat. I en sådan modell är kanoniska skattningar av substitutionselasticiteten underskattade: för lokala chocker och korta tidshorisonter är en låg elasticitet och en anpassningskostnad observationellt likartade. Den strukturella modellen inför ytterligare momentrestriktioner på insatsernas tidsseriebeteende, vilket möjliggör separat identifiering av elasticiteten och friktionen. Strukturell skattning ger en elasticitet inom anläggningen på 5.

Jag illustrerar konsekvenserna för klimatpolitiken med hjälp av Tysklands sektorsvisa utsläppsmål. En minskning av användningen av fossila bränslen med 40% kan uppnås med en skatt på 11% på priserna på fossila bränslen, vilket orsakar en produktionsminskning på 7%. Med den lägre elasticiteten enligt befintliga skattningar kräver samma minskning en skatt på 23% och medför en produktionsminskning på 14%.

Slutligen dokumenterar jag en, i tvärsnittet, stor heterogenitet i energimixen mellan anläggningar, även inom snävt definierade produktgrupper. Denna heterogenitet innebär en ytterligare omfattande anpassningsmarginal genom omfördelning mellan anläggningar, vilket förstärker den intensiva marginalen inom anläggningarna. Elasticiteten inom anläggningarna på 5 är således en nedre gräns för den aggregerade potentialen för substitution.

Upptäckten att anpassningsfriktioner i användningen av fossila bränslen leder till en underskattning av elasticiteten med kanoniska metoder väcker en allmän metodologisk fråga. Anpassningsfriktioner är inte unika för energi: den empiriska litteraturen dokumenterar elasticitetsuppskattningar som ökar med identifieringens tidshorisont i många sammanhang, vilket stämmer överens med förekomsten av anpassningsfriktioner. Kortsiktiga skattningar som erhålls med kanoniska metoder kan därför systematiskt underskatta motsvarande långsiktiga elasticiteter i många tillämpningar.

I det andra kapitlet, *Long-run elasticities from short-run variation*, bygger jag vidare på denna punkt och utvecklar ett linjärt skattningsramverk för att härleda den långsiktiga substitutionselasticiteten mellan insatsvaror från kortsiktiga variationer i insatsanvändningen, när valet av insatsvaror är föremål för anpassningskostnader.

Ett vanligt fynd i den empiriska litteraturen är att elasticitetsskattningarna ökar med tidshorisonten för prisförändringen, och att det finns en skillnad mellan kort- och långsiktig elasticitet. Båda värdena är relevanta för politiken: den kortsiktiga elasticiteten styr övergångens hastighet, medan det långsiktiga värdet beskriver dess slutpunkt. Den kortsiktiga elasticiteten är lättare att uppskatta, eftersom exogena övergående prisvariationer är lättare att få tillgång till än exogena och bestående variationer.

Jag visar att i vissa fall ger den kortsiktiga dynamiken i insatsanvändningen ändå information om den långsiktiga elasticiteten, och att förhållandet mellan de två kan uttryckas i en korrigeringsfaktor som kan uttryckas i sluten form. För att härleda detta resultat modellerar jag valet av insatsförhållandet som föremål för en kvadratisk anpassningskostnad. Under en andra ordningens approximation av en CES-produktionsfunktion är kortfristig elasticitet produkten av en dämpningsfaktor och långfristig elasticitet. Den centrala insikten är att insatsförhållandets persistens är en tillräcklig statistik för anpassningsfriktionen, vilket möjliggör separat identifiering av elasticiteten och friktionen. Dämpningsfaktorn har ett slutligt uttryck i tre parametrar: insatsförhållandets persistens, det relativa prisets persistens och diskonteringsfaktorn. De två första kan estimeras från data; den tredje kan kalibreras till standardvärden. När anpassningskostnader förekommer är långsiktig elasticitet något större än kortsiktig elasticitet.

Jag visar att den kanoniska inomestimatorn i paneldata riktar in sig på kortsiktig elasticitet. Den kanoniska långsiktiga estimatorn, mellanestimatorn, kan vara inkonsekvent om inte prisnivån är strikt exogen mellan enheterna. Korrigeringsmetoden är komplementär: den riktar in sig på samma långsiktiga parameter under en annan uppsättning antaganden, och förlitar sig på exogenitet i förändringar snarare än prisnivåer, men lägger större struktur på aktörernas beteende.

Jag tillämpar metoden på substitution av mellanprodukter inom den indiska tillverkningsindustrin, med hjälp av data från Peter och Ruane (2025). De uppskattar en kortsiktig elasticitet på cirka 0,5 och en långsiktig elasticitet på cirka 2,5 över sju år, med den indiska handelsliberaliseringen som ett naturligt experiment. Min korrigeringsfaktor innebär en långsiktig elasticitet som är 1,75 gånger den kortsiktiga uppskattningen, vilket är mindre än deras direkt uppskattade kvot på cirka 5. Möjliga orsaker till avvikelsen är bland annat olika identifieringsantaganden, modellstruktur och den tidshorisont över vilken långsiktigheten definieras.

De två första kapitlen undersöker hur anpassningsfriktioner dämpar ekonomins kortsiktiga respons på chocker, och vad detta innebär för mätning och politik. De undersöker hur stor responsen på en chock är, och finner att den är större än vad kanoniska skattningar antyder.

En kompletterande fråga uppstår när reaktionen är kraftigt begränsad: kan politiska ingripanden kompensera för detta? Om det är svårt att omfördela produktionen efter en chock kan en regering försöka forma dess sammansättning innan negativa händelser inträffar, snarare än att framkalla en omfördelning i efterhand. Men sådana förebyggande ingripanden medför en ny utmaning: rationella aktörer förutser regeringens reaktion på deras val, vilket potentiellt kan upphäva de avsedda fördelarna.

Det tredje kapitlet, *Financial frictions and aggregate risk exposure*, övergår till en situation med en annan typ av friktion: en finansiell begränsning. Denna begränsning hämmar omfördelningen av produktionen mellan olika tekniker efter en chock, och kapitlet undersöker om förebyggande politiska ingripanden kan förbättra välfärden.

Entreprenörer väljer mellan en säker teknik med stabil produktivitet och en riskfylld teknik med högre förväntad produktivitet men utsatthet för negativa chocker. En säkerhetsbegränsning begränsar företagens förmåga att låna kapital, vilket fångar upp det begränsade utrymmet för omfördelning efter en chock. En planerare omfördelar tillgångar mellan olika typer av teknik innan produktionen sker, vilket lättar på lånebegränsningarna för vissa företag på bekostnad av att skärpa dem för andra. Avgörande är att planeraren saknar åtagande: den väljer omfördelning efter att ha observerat sammansättningen av teknikvalen, och aktörerna förutser detta när de fattar sina beslut.

Den laissez-faire-baserade allokeringen är begränsat effektiv. Om man jämför en intervention utan åtagande med laissez-faire kan välfärdseffekten delas upp i två motsatta krafter: en omfördelningsvinst från en optimal omfördelning av tillgångar med fast tekniksammansättning, och en snedvridningskostnad från den förändring i tekniksammansättningen som orsakas av den förväntade omfördelningen. Snedvridningen uppstår eftersom aktörerna reagerar på den förväntade omfördelningen genom två kanaler: högre förväntade vinster från lättade begränsningar, och en direkt förmögenhetsöverföring. Planeraren internaliserar endast den första när han väljer omfördelning, vilket skapar en klyfta som får aktörerna att överskrida målet mot den gynnad tekniken.

Snedvridningskostnaden överstiger omfördelningsvinsten för den stora majoriteten av parametriseringarna. Endast när den riskfyllda tekniken producerar nästan ingenting i det dåliga läget förbättrar icke-bindande ingripanden välfärden marginellt. Detta är ett empiriskt extremt scenario som motsvarar en nästan fullständig avstängning av utbudet. I det empiriskt mer relevanta fallet med stora pristoppar men utan fullständiga avbrott dominerar snedvridningskostnaden, och förebyggande ingripande minskar välfärden jämfört med en planerare som inte ingriper alls.

Kapitel 1 och 2 visar att kanoniska metoder kan missa elasticiteten när friktioner förekommer, och att skillnaden går i en specifik riktning: kortsiktiga variationer underskattar långsiktig flexibilitet. Att korrigera för detta förändrar både siffrorna och den politiska slutsats som följer av dem. Kapitel 3 tar upp gränsfallet – en friktion så allvarlig att omfördelning efter chocken är praktiskt taget omöjlig – och frågar om politiken före chocken kan kompensera för detta. Svaret beror på åtagande: utan det snedvrider förväntningarna på politiken just de val som politiken var avsedd att forma.

Friktioner är inte olägenheter som ska jämnas ut. Deras struktur – anpassningskostnader, beständighet, finansiella begränsningar – avgör vad vi kan mäta och vad en regering trovärdigt kan göra. Att ta friktionen på allvar, snarare än att absorbera den i en aggregerad respons, är det som förändrar den politiska lösningen i varje kapitel.

IIES Monograph Series

1. Michaely, Michael, *The Theory of Commercial Policy: Trade and Protection*, 1973
2. Söderström, Hans Tson, *Studies in the Microdynamics of Production and Productivity Change*, 1974
3. Hamilton, Carl B., *Project Analysis in the Rural Sector with Special Reference to the Evaluation of Labour Cost*, 1974
4. Nyberg, Lars and Staffan Viotti, *A Control Systems Approach to Macroeconomic Theory and Policy in the Open Economy*, 1975
5. Myhrman, Johan, *Monetary Policy in Open Economies*, 1975
6. Krauss, Melvyn, *International Trade and Economic Welfare,* 1975
7. Wihlborg, Clas, *Capital Market Integration and Monetary Policy under Different Exchange Rate Regimes,* 1976
8. Svensson, Lars E.O., *On Competitive Markets and Intertemporal Resources Allocation*, 1976
9. Yeats, Alexander J., *Trade Barriers Facing Developing Countries*, 1978
10. Calmfors, Lars, *Prices, Wages and Employment in the Open Economy*, 1978
11. Kornai, Jànos, *Economics of Shortage, Vols I and II*, 1979
12. Flam, Harry, *Growth, Allocation and Trade in Sweden. An Empirical Application of the Heckscher-Ohlin Theory*, 1981
13. Persson, Torsten, *Studies of Alternative Exchange Rate Systems. An Intertemporal General Equilibrium Approach*, 1982

14. Erzan, Refik, *Turkey's Comparative Advantage, Production and Trade Patterns in Manufactures. An Application of the Factor Proportions Hypothesis with Some Qualifications*, 1983
15. Horn af Rantzien, Henrik, *Imperfect Competition in Models of Wage Formation and International Trade*, 1983
16. Nandakumar, Parameswar, *Macroeconomic Effects of Supply Side Policies and Disturbances in Open Economies*, 1985
17. Sellin, Peter, *Asset Pricing and Portfolio Choice with International Investment Barriers*, 1990
18. Werner, Ingrid, *International Capital Markets: Controls, Taxes and Resources Allocation*, 1990
19. Svedberg, Peter, *Poverty and Undernutrition in Sub-Saharan Africa: Theory, Evidence, Policy*, 1991
20. Nordström, Håkan, *Studies in Trade Policy and Economic Growth*, 1992
21. Hassler, John, Petter Lundvik, Torsten Persson, and Paul Söderlind, *The Swedish Business Cycle: Stylized facts over 130 years,* 1992
22. Lundvik, Petter, *Business Cycles and Growth*, 1992
23. Söderlind, Paul, *Essays in Exchange Rates, Business Cycles and Growth*, 1993
24. Hassler, John A.A., *Effects of Variations in Risk on Demand and Measures of Business Cycle Comovements*, 1994
25. Daltung, Sonja, *Risk, Efficiency, and Regulation of Banks*, 1994
26. Lindberg, Hans, *Exchange Rates: Target Zones, Interventions and Regime Collapses*, 1994
27. Stennek, Johan, *Essays on Information-Processing and Competition*, 1994
28. Jonsson, Gunnar, *Institutions and Incentives in Monetary and Fiscal Policy*, 1995
29. Dahlquist, Magnus, *Essays on the Term Structure of Interest Rates and Monetary Policy*, 1995
30. Svensson, Jakob, *Political Economy and Macroeconomics: On Foreign Aid and Development,* 1996

31. Blix, Mårten, *Rational Expectations and Regime Shifts in Macroeconometrics*, 1997
32. Lagerlöf, Nils-Petter, *Intergenerational Transfers and Altruism*, 1997
33. Klein, Paul, *Papers on the Macroeconomics of Fiscal Policy*, 1997
34. Jonsson, Magnus, *Studies in Business Cycles*, 1997
35. Persson, Lars, *Asset Ownership in Imperfectly Competitive Markets*, 1998
36. Persson, Joakim, *Essays on Economic Growth*, 1998
37. Domeij, David, *Essays on Optimal Taxation and Indeterminacy*, 1998
38. Flodén, Martin, *Essays on Dynamic Macroeconomics*, 1999
39. Tangerås, Thomas, *Essays in Economics and Politics: Regulation, Elections and International Conflict*, 2000
40. Lidbom, Per Pettersson, *Elections, Party Politics and Economic Policy*, 2000
41. Vestin, David, *Essays on Monetary Policy*, 2001
42. Olofsgård, Anders, *Essays on Interregional and International Political Economics*, 2001
43. Johansson, Åsa, *Essays on Macroeconomic Fluctuations and Nominal Wage Rigidity*, 2002
44. Groth, Charlotta, *Topics on Monetary Policy*, 2002
45. Gancia, Gino A., *Essays on Growth, Trade and Inequality*, 2003
46. Harstad, Bård, *Organizing Cooperation: Bargaining, Voting and Control*, 2003
47. Kohlscheen, Emanuel, *Essays on Debts and Constitutions*, 2004
48. Olovsson, Conny, *Essays on Dynamic Macroeconomics*, 2004
49. Stavlöt, Ulrika, *Essays on Culture and Trade*, 2005
50. Herzing, Mathias, *Essays on Uncertainty and Escape in Trade Agreements*, 2005
51. Bonfiglioli, Alessandra, *Essays on Financial Markets and Macroeconomics*, 2005

52. Pienaar, Natalie, *Economic Applications of Product Quality Regulation in WTO Trade Agreements*, 2005
53. Song, Zheng, *Essays on Dynamic Political Economy*, 2005
54. Eisensee, Thomas, *Essays on Public Finance: Retirement Behavior and Disaster Relief*, 2005
55. Favara, Giovanni, *Credit and Finance in the Macroeconomy*, 2006
56. Björkman, Martina, *Essays on Empirical Development Economics: Education, Health and Gender*, 2006
57. Larsson, Anna, *Real Effects of Monetary Regimes*, 2007
58. Prado, Jr., Jose Mauricio, *Essays on Public Macroeconomic Policy*, 2007
59. Tonin, Mirco, *Essays on Labor Market Structures and Policies*, 2007
60. Queijo von Heideken, Virginia, *Essays on Monetary Policy and Asset Markets*, 2007
61. Finocchiaro, Daria, *Essays on Macroeconomics*, 2007
62. Waisman, Gisela, *Essays on Discrimination and Corruption*, 2008
63. Holte, Martin Bech, *Essays on Incentives and Leadership*, 2008
64. Damsgaard, Erika Färnstrand, *Essays on Technological Choice and Spillovers*, 2008
65. Fredriksson, Anders, *Bureaucracy, Informality and Taxation: Essays in Development Economics and Public Finance*, 2009
66. Folke, Olle, *Parties, Power and Patronage*, 2010
67. Drott, David Yanagizawa, *Information, Markets and Conflict*, 2010
68. Meyersson, Erik, *Religion, Politics, and Development*, 2010
69. Klingelhofer, Jan, *Models of Electoral Competition*, 2010
70. Perrotta, Maria Carmela, *Aid, Education and Development*, 2010
71. Caldara, Dario, *Essays on Empirical Macroeconomic*s, 2011
72. Mueller, Andreas, *Business Cycles, Unemployment and Job Search*, 2011
73. von Below, David, *Essays in Climate and Labour Economics*, 2011

74. Gars, Johan, *Essays on the Microeconomics of Climate Change*, 2012
75. Spiro, Daniel, *Some Aspects of Resource and Behavioral Economics*, 2012
76. Ge, Jinfeng, *Essays on Macroeconomics and Political Economy*, 2012
77. Li, Yinan, *Institutions, Political Cycles and Corruption*, 2012
78. Håkanson, Christina, *Changes in Workplaces and Careers*, 2013
79. Qin, Bei, *Essays on Empirical Development and Political Economics*, 2013
80. Jia, Ruixue, *Essays on the Political Economics of China's Development*, 2013
81. Campa, Pamela, *Media Influence on Pollution and Gender Equality*, 2013
82. Seim, David, *Essays on Public, Political and Labor Economics*, 2013
83. Shifa, Abdulaziz B., *Essays on Growth, Political Economy and Development*, 2013
84. Panetti, Ettore, *Essays on the Economics of Banks and Markets*, 2013
85. Schmitt, Alex, *Beyond Pigou: Climate Change Mitigation, Policy Making and Decisions*, 2014
86. Rogall, Thorsten, *The Economics of Genocide and War*, 2015
87. Baltrunaite, Audinga, *Political Economics of Special Interests and Gender*, 2016
88. Harbo Hansen, Niels-Jakob, *Jobs, Unemployment, and Macroeconomic Transmission*, 2016
89. Stryjan, Miri, *Essays on Development Policy and the Political Economy of Conflict*, 2016
90. Karadja, Mounir, *On the Economics and Politics of Mobility*, 2016
91. Kitamura, Shuhei, *Land, Power and Technology, Essays on Political Economy and Historical Development*, 2016

92. Malmberg, Hannes, *Human Capital in Development Accounting and Other Essays in Economics*, 2017
93. Öberg, Erik, *On Money and Consumption*, 2017
94. Lane, Nathaniel, *States of Development: Essays on the Political Economy of Development in Asia*, 2017
95. Prawitz, Erik, *On the Move: Essays on the Economic and Political Development of Sweden*, 2017
96. Iwanowsky, Mathias, *Essays in Development and Political Economics*, 2018
97. Dehdari, Sirus Håfström, *Radical Right, Identity, and Retaliation*, 2018
98. Harmenberg, Karl, *Essays on Income Risk and Inequality*, 2018
99. Kilström, Matilda, *Households' Responses to Policy in Labor and Credit Markets*, 2018
100. Meriläinen, Jaakko, *Essays in Political Economics*, 2019
101. Sigurdsson, Jósef , *Essays on Labor Supply and Adjustment Frictions*, 2019
102. Cocciolo, Serena, *Participatory Governance and Public Service Provision* , 2019
103. Mitrunen, Matti, *Essays on the Political Economy of Development*, 2019
104. Olsson, Jonna, *Work, wealth, and well-being: Essays in macroeconomics*, 2019
105. Darougheh, Saman, *Search and Mismatch*, 2019
106. van Vlokhoven, Has, *On the Cost of Capital, Profits and the Diffusion of Ideas*, 2020
107. Foltyn, Richard, *Essays in Macroeconomics and Household Finance*, 2020
108. Lerva, Benedetta, *Worms, Farms and Schools: Three Essays on Investment Choices in Uganda*, 2020
109. Kragh-Sørensen, Kasper, *On Housing, Mortgages, and Taxation*, 2020

110. Karlman, Markus, *Essays on Housing: tax treatment, prices, and macroeconomic implications*, 2020
111. Kinnerud, Karin, *Financial Choice and Public Policy*, 2020
112. Ghisolfi, Selene, *Fairness, technology adoption, water sanitation and pandemic control: Six essays on four topics in Development Economics*, 2020
113. Åhl, Magnus, *Essays on Individual-Level Wage Stickiness and Forward Guidance*, 2020
114. Dev, Divya, *Essays on Gender, Development and Political Economy*, 2021
115. Gallegos Dago, José Elias, *Essays on Macroeconomics*, 2022
116. Sun, Xueping, *Essays on China's Economic Development: Innovation, Public Debt and Social Connections*, 2022
117. Kramer, John Vincent, *Essays on Macroeconomics, Monetary Policy and Mobility*, 2022
118. Almgren, Mattias, *Essays on Home Production, Mobility, and Monetary Policy*, 2023
119. Thürwächter, Claire, *Essays on Macroeconomics, Monetary Policy and Firm Heterogeneity*, 2023
120. von Carnap, Tillmann, *Markets and marketplaces: Essays on access and transformation in remote rural economies*, 2023
121. Peters, Markus, *Essays on Savings Behavior, Inflation Measurement, and Growth*, 2023
122. Loiacono, Francesco, *Firms and Labor Markets: Essays in Development Economics*, 2023
123. Tebbe, Sebastian, *Externalities and Coordination Failures*, 2023
124. Azzalini, Gualtiero, *Essays on Income Risk, Portfolio Choices and the Macroeconomy*, 2023
125. Kondziella, Markus, *Essays on Economic Growth, Inflation and Inequality*, 2023
126. Paues, Fredrik, *Essays on Housing Deregulation and Investment Behavior*, 2024

127. Hochmuth, Philipp, *Essays on Hours Worked and Cost-of-Living Inequality*, 2024
128. Hinkelmann, Stefan, *On the Macroeconomics of the Energy Transition*, 2024
129. Sen, Sreyashi, *Essays on Markets and Institutions in Developing Countries*, 2024
130. Linnros, Evelina, *Essays on Fertility and Health*, 2024
131. Sinn, Fabian, *Essays on Labor Economics and Social Dynamics*, 2024
132. Farajpour, Mohammadreza, *On the Economics of Energy and Climate Change*, 2024
133. Biesalski, Dominik, *Essays on Working Conditions, Occupational Choices, and Discrimination*, 2025
134. Berge, Agneta, *On the Formation of Wages and Wealth*, 2025
135. Massner, Patrizia, *Essays on Elder Care, Intergenerational Mobility, and Firms in the Labor Market*, 2025
136. Bianchi, Iacopo, *Productivity, Career Choice, and Colonial Legacies: Essays in Development Economics*, 2025
137. Bernadino, Tiago, *Consumption, Prices, and Policy: Essays in Macroeconomics and Public Economics*, 2026
138. Liu, Jinci, *Managers, Workers, and Teams: Essays in Organizational and Labor Economics*, 2026

This thesis comprises three self-contained chapters.

Can the plants turn green? studies how electricity and fossil fuels are used in manufacturing, and describes implications for the substitution between the two.

Long-run elasticities from short-run variation shows that the substitution potential between two production factors can be underestimated in the presence of adjustment frictions. It develops a closed-form correction factor for a particular case.

Financial frictions and aggregate risk exposure studies a model of technology choice under a financial friction, and shows the implications for optimal policy.

Thore Petersen
holds a B.Sc. in Economics and a M.Sc. in Quantitative Economics from Kiel University.

ISBN 978-91-8107-632-5
ISSN 0346-6892

Department of Economics

Stockholm University