LIETUVOS BANKAS
EUROSISTEMA

Optimal Firm Entry with Returns to Scale

Working Paper Series

No 129 / 2025

Working Paper Series
No 129 / 2025

ISSN 2029-0446 (online)

Optimal Firm Entry with Returns to Scale

Anup Mulay*

Anthony Savagar†

Disclaimer: The views expressed herein are solely those of the authors and do not necessarily reflect the views of Lietuvos Bankas or of the Eurosystem.

* Lietuvos Bankas. E-mail: AMulay@lb.lt

† University of Kent. E-mail: a.savagar@kent.ac.uk

Gedimino pr. 6, LT-01103 Vilnius

www.lb.lt

Working papers describe research in progress by the author(s) and are published to stimulate discussion and critical comments.

The series is managed by the Applied Macroeconomic Research Division of the Economics Department and the Center for Excellence in Finance and Economic Research.

All papers in the series are refereed by internal and external experts.

ABSTRACT

We study the welfare implications of distortions, such as markups and returns to scale, when firm entry is slow to adjust, allowing quasi-rents to persist for longer. First, we present evidence on differences in speed of firm entry adjustment across US industries. In some industries, such as hospitality, firms respond rapidly to profit opportunities, arbitraging quasi-rent quickly. Whereas, in other industries, such as construction, entrants respond slowly, sustaining incumbents' quasi-rents for longer. We develop a model of sluggish firm adjustment, which shows that the sluggishness of firm adjustment magnifies the welfare costs of distortions. We study a model with a fixed cost and increasing marginal cost such that a perfectly competitive equilibrium exists, and in the absence of distortions market and planner equilibrium coincide with firms operating at minimum efficient scale. We contrast outcomes when there is curvature on the demand-side of the economy from markups and curvature on the supply-side of the economy from returns to scale, adding counter-evidence to the perception that the setups are isomorphic.

Keywords: Markups, Firm Entry, Returns to Scale, Welfare

JEL codes: E32, D21, D43, L13, C62

1 Introduction

There is a growing perception that markets in many advanced economies have become less competitive. However, recent episodes like the advent of large language models contradict this intuition of less competitive markets. During this episode, entry was rapid as challenger LLMs emerged. The markets seem contestable, even if markups are high. In this paper we ask whether the ability for firms to enter quickly (contestability) can yield welfare gains similar to the welfare costs of static distortions.

The speed at which firms can adjust to profit opportunities differs across markets, sectors and countries. There are many frictions that can affect this mechanism such as regulations, startup costs, infrastructure, and labour availability. In our framework a faster speed of adjustment means that quasi-profits are arbitraged quicker. We consider this to represent dynamic competition or contestability.

Our main research questions are: How costly are distortions when firms are slow to adjust? Alternatively, what are the welfare gains of fast firm adjustment? We show that slow firm adjustment significantly magnifies the welfare losses from static distortions, whereas rapid firm adjustment mitigates these losses. We compare slow-adjustment economies to economies where the distortion can be eliminated instantaneously, and the number of firms adjusts immediately, as in standard static analyses. Consequently, policy should not only focus on reducing static distortions, but it should reduce barriers to entry that inhibit fast adjustment.

We study a perfect foresight, deterministic, DGE model. Firms and households are representative, and there are endogenous entry costs which are determined by a network effect. This is formulated as a quadratic queuing cost to enter a market leading to congestion.[1] We consider this to represent dynamic barriers to entry or contestability. It gives households – who invest in firm creation – an incentive to delay entry to a later date when endogenous entry costs have declined. Our framework includes firms with output-denominated fixed costs and arbitrarily sloping marginal costs, as well as external scale returns, and monopolistic competition, leading to constant markups.[2] In our framework, increasing returns at the firm level (downward-sloping marginal cost) is admissible either through markups if there are constant external returns or increasing external returns if there are no markups. Either approach yields downward-sloping marginal-revenue curves, such that firms do not expand boundlessly given the increasing returns.

Our theoretical results begin with a static analysis. We compare the steady-state

[1]Our entry model and analysis has close analogs with investment adjustment cost models, where investment in our setup is in firms rather than physical capital. This allows us to acquire analytical solutions to policy functions which allows us to specify the adjustment cost parameter precisely.

[2]In the literature, external returns to scale are referred to by various names depending on the context, for example love of variety, returns to specialization, agglomeration effects, thick markets, and Ethier effects. They are mathematically similar, representing aggregations of generalized means.

market outcome to the steady-state planner outcome, which illustrates the distance to the efficient frontier. We show how both markups and external returns to scale cause the market outcome to differ from the planner outcome. The market outcome achieves the planner outcome when there are constant external returns to scale and no markups.[3] If there are external returns to scale or if there are markups, the planner can always achieve a higher utility level for the representative household than the decentralised outcome will deliver. The most costly case in terms of welfare is when there is a markup and no external returns to scale – this leads to excessive entry (too many small firms at the micro level) without the benefit from external returns at the aggregate level. With increasing external returns, raising markups is welfare enhancing because it encourages further entry and greater exploitation of external increasing returns. Hence, eliminating markups is not optimal in this scenario. In other words, it is better to have more firms producing below their minimum efficient scale, than fewer firms all producing at their minimum efficient scale. This is because the gain in efficiency from aggregation offsets the loss in efficiency at the firm level.

Our dynamic analysis compares two scenarios. We compare eliminating the markup when there is instantaneous adjustment to eliminating the markup when adjustment is sluggish.[4] In both scenarios the markup is eliminated, but in the first-scenario the new, no markup, steady state is achieved instantaneously, whereas in the second scenario there is a transition period. We find 4% lower welfare in the slow-adjusting case compared to the instantaneous adjustment case.[5]

Our results have important policy implications. Often policymakers focus on static measures of competition and reducing static distortions. We explain that dynamic competition can enhance the welfare gains from removing static distortions. This justifies existing policy efforts such as the World Bank's Doing Business Indicators, which measure regulatory barriers to entry such as number of days to setup a firm. In general, it supports policies that reduce regulation allowing for fast firm adjustment alongside efforts to limit static markups by regulating mergers and anti-competitive behaviour.

Related literature

There is a long-established literature on optimal firm entry under imperfect competition. Traditionally, this focuses on individual markets, but recent work attempts to aggregate welfare implications in macroeconomic settings. Morrow and Dhingra

[3] In this case, *internal* returns to scale must be decreasing (upward-sloping marginal cost) due to the fixed cost, leading to U-shaped average costs curves and optimality where all firms operate at the minimum of their average cost, i.e. minimum efficient scale (MES).

[4] Our parameterization for slow adjusment, makes all firms in the economy adjust as slowly as in the US construction sector.

[5] In our model notation $\psi = 10$, AR(1) coefficient of 0.7, half-life 2 quarters.

(2014) focus on the role of heterogeneity in static models with love-of-variety, whilst Bilbiie, Ghironi, and Melitz (2019) focus on representative firms in a dynamic framework with endogenous markups, and Edmond, Midrigan, and Xu (2023) focus on a dynamic model with firm heterogeneity, and roles for both love-of-variety and endogenous markups. Baqaee and Farhi (2020) layout non-parametric results for aggregating economies with distortions in a static framework that includes returns to scale. Barro (2024) presents a framework that is reminiscent of contestable markets with markups depending on distances between stores. The parallel with our work is that our congestion effects setup could be microfounded from a Salop circle type model, similar to Ershov (2024). Bergin and Bernhardt (2008) show that negative demand shocks can increase welfare because they increase exit of low productivity firms. Low demand narrows the difference between the social and private opportunity cost of keeping a firm alive. Raising the private opportunity cost closer to the social opportunity cost increases exit. Consequently, by forcing inefficient firms to close ('Darwinian cleansing'), leading to a more productive economy in the future, the long-term gain can outweigh the short-term losses.

Our value-added relative to existing literature is to focus on how the speed of firm adjustment affects the welfare gains of moving from distorted to undistorted equilibrium. Additionally, we study returns to scale in detail both at the firm level (internal returns to scale) and at the industry level (external returns to scale). Including increasing external returns to scale allows us to study scenarios where there is perfect competition on the demand side and increasing returns at the firm level. And, crucially, scenarios where perfect competition in conjunction with constant external returns to scale, and decreasing internal returns, leads the decentralised equilibrium to coincide with the planner equilibrium. Unlike studies that use a second-best benchmark where markups persist, our study employs a first-best benchmark where markups are eliminated, resulting in a Pareto efficient equilibrium.

2 Empirical Motivation

In this section we present empirical evidence of slow firm adjustment. To measure the sluggishness of firm adjustment, we estimate an AR(1) model on quarterly number of firms data N_t for US sectors from 2001 Q1 to 2019 Q1. There are $t = 73$ time periods and eight industries. A full description of our data is in Appendix A. We transform the AR(1) coefficient into a half-life which represents the number of quarters it takes for the number of firms to revert halfway to trend following a shock. Therefore, we estimate the following relationship, where hat notation represents linearly detrended data and Δ implies it is differenced:[6]

$$\Delta \hat{N}_{t+1} = \varphi \Delta \hat{N}_t + \varepsilon_t,$$

The half-life is a transformation of the AR(1) coefficient:

$$\text{Half-life} = -\frac{\ln(2)}{\ln(|\varphi|)}.$$

Table 1 presents our results and orders industries from slow-adjusting to fast-adjusting. The standard errors show that the coefficients are significant and also statistically significantly different to eachother in most cases. Industries with slower adjustment are able to sustain economic profits for longer and competition effects from entry can be abated for longer.[7] The ranking suggests a plausible correlation between industries that have high fixed costs and greater regulation, such as construction, taking longer to adjust compared to industries with lower fixed costs and regulation adjusting more quickly. The results show that it takes the construction industry two quarters to revert halfway to trend, whereas it takes the leisure and hospitality industry one-third of a quarter to revert halfway to trend.

Industry	φ	**S.E.**	**Half-life**
Construction	0.71	0.09	1.98
Financial Activities	0.67	0.09	1.73
Natural Resources and Mining	0.49	0.10	0.97
Trade, Transportation, Utilities	0.37	0.11	0.69
Manufacturing	0.36	0.11	0.69
Information	0.30	0.12	0.58
Professional and Business	0.27	0.12	0.52
Leisure and Hospitality	0.13	0.12	0.34

Table 1: AR(1) Results for Establishment Speed of Adjustment, US Industries 2001 Q1-2019 Q1, (Source: QCEW, BLS)

3 Model

We develop a model where a representative household can invest in setting-up firms subject to an adjustment cost. Labour is endogenously determined by household sup-

[6]We remove seasonality, linearly detrend and difference our data.

[7]The temporary profits that occur as adjustment takes place are traditionally referred to as quasi-rents, quasi-profits or Marshallian rents. They are profits that arise in the short run due to short-run entry frictions, barriers to entry, patents and so on. In the long run these profits are competed away.

ply and firm demand. Firms are representative and compete under monopolistic competition, leading to a fixed price markup that depends on product substitutability. Individual firms have output-denominated fixed costs and a variable production function with arbitrary returns to scale (i.e. increasing, decreasing or constant marginal cost). We present a generalized CES aggregator that explicitly parameterises the love of variety effect that is implicit in the standard CES aggregator. Since we present output aggregators rather than consumption aggregators, we refer to this love of variety parameter as external returns to scale. Increasing external returns to scale allows a determinate equilibrium with internal increasing returns but without markups. Similarly including markups allows a determinate equilibrium with internal increasing returns.[8]

3.1 Household

Households maximise utility by solving the following problem following

$$\max_{C_t,L_t,E_t,N_{t+1}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma}-1}{1-\sigma} - \chi \frac{L_t^{1+\eta}}{1+\eta} \right)$$

subject to

$$C_t + E_t + \mathcal{C}(E_t, N_t) = w_t L_t + \pi_t N_t \tag{1}$$

$$N_{t+1} = E_t + (1-\delta) N_t. \tag{2}$$

In the budget constraint, prices are in real terms. π_t denotes real operating profits, w_t the real wages.[9] N_t is the total number of firms held in period t. E_t is new entrants or investment in new firms, which can be positive or negative.

We assume an adjustment cost to creating new firms given by $\frac{\psi}{2}\left(\frac{E_t}{N_t}-\delta\right)^2$, where ψ is the adjustment cost parameter that determines the speed of adjustment of firms. Firms adjust more slowly when ψ is higher. The cost of setting-up a firm is higher when the entry rate E/N is higher.[10] This is a congestion effect. Therefore the total adjustment cost is:

$$\mathcal{C}(E_t, N_t) \equiv \frac{\psi}{2}\left(\frac{E_t}{N_t}-\delta\right)^2 N_t.$$

[8] Many existing papers have used increasing returns aggregators to allow for internal increasing returns, without the need to introduce downward-sloping demand for example Matsuyama (1991).

[9] We refer to nominal prices divided by the aggregate price level as real prices, so the real wage $\frac{w_t^{\text{nominal}}}{P_t}$, and real profits $\frac{\pi_t^{\text{nominal}}}{P_t}$. We will introduce real product prices, which is the nominal price divided by the firm-level price.

[10] Equivalently, when the growth rate of the companies $E_t/N_t - \delta = (N_{t+1} - N_t)/N_t$ is higher, the cost of setting up is higher.

The optimization conditions for the household reduce to the following equilibrium conditions:

$$w_t = C_t^{\sigma} \chi L_t^{\eta} \tag{3}$$
$$s_t = 1 + \mathcal{C}_E(E_t, N_t) \tag{4}$$
$$s_t = \beta \mathbb{E}_t \left\{ \left(\frac{C_t}{C_{t+1}} \right)^{\sigma} \left(\pi_{t+1} - \mathcal{C}_N \left(E_{t+1}, N_{t+1} \right) + (1-\delta) s_{t+1} \right) \right\} \tag{5}$$

Subscripts denote partial derivatives. Specifically, the effects of an additional firm and additional entry investment on the total adjustment cost are

$$\mathcal{C}_N \left(E_t, N_t \right) = \frac{\psi}{2} \left(\frac{E_t}{N_t} - \delta \right)^2 - \psi \left(\frac{E_t}{N_t} - \delta \right) \frac{E_t}{N_t},$$
$$\mathcal{C}_E(E_t, N_t) = \psi \left(\frac{E_t}{N_t} - \delta \right).$$

s_t is the relative price of a firm in terms of consumption. That is, the amount of consumption forgone for an extra firm.[11] The cost of entry s_t is determined by flow of firms into the market. In equilibrium the price of entry equals to the net present value of incumbency. The marginal cost of entry varies with the flow of entry due to a congestion effect in setting up new firms. Collectively, the two equilibrium conditions form a free entry condition. Equation (4) determines the endogenous entry cost and equation (5) states that this endogenous entry cost must equate to future discounted profits from owning firms.[12] If we forward iterate (5) we have

$$s_t = \beta \mathbb{E}_t \sum_{i=1}^{\infty} (\beta(1-\delta))^{i-1} \left(\frac{C_t}{C_{t+i}} \right)^{\sigma} \left[\pi_{t+i} - \mathcal{C}_N \left(E_{t+i}, N_{t+i} \right) \right].$$

In the appendix, we relate firm asset value v_t to the relative price s_t. We show that firm asset value is the sum of current and future discounted profits-less-adjustment-costs $\pi_t - \mathcal{C}_{N,t}$. We express this as a familiar Bellman equation.

[11] In Appendix B, we present the full optimization problem and explain further that s_t is the ratio of Lagrange multipliers on the two constraints.

[12] In macroeconomic models with firm entry, the free-entry condition is usually assumed, whereas here it arises endogenously. For example, it is common to assume a labour denominated fixed entry cost that must equate to the net-present value of a firm: $v_t = w_t f_E / Z_t$. This states that firm value v_t equals to a constant fixed cost f_E multiplied by the marginal cost of production w_t/Z_t, which is given by the ratio of wages w_t to technology Z_t in a model with only labour in production. This says that the value of a firm is equal to the marginal cost of producing an extra unit of existing good. Since marginal costs do not respond on impact as they depend on the stock of firms which is inert on impact, the free-entry condition causes entry to adjust such that the NPV does also not respond on impact. Entry leads to a change in the stock of firms next period.

3.1.1 Initial Conditions and Transversality

The initial value of the state variable is N_0, and the transversality condition is:

$$\lim_{t\to\infty} \mathbb{E}_0 \beta^t N_{t+1} s_t = 0.$$

3.2 Firms

We begin by solving a model of perfect competition in the final goods market, where a final goods producer aggregates the output of intermediate goods producers. Subsequently, we examine the behavior of imperfectly competitive intermediate goods firms.

3.2.1 Final goods firms

We assume the aggregate production function is a generalized CES. This has increasing external returns to scale if $\gamma > 1$ and constant external returns to scale if $\gamma = 1$.

$$Y_t = N_t^{\gamma} \left[\frac{1}{N_t} \int_0^{N_t} y_{j,t}^{\frac{\theta-1}{\theta}} \mathrm{d}j \right]^{\frac{\theta}{\theta-1}}. \tag{6}$$

The perfectly competitive final goods firms maximize their profits, taking prices as given:

$$\max_{y_j} \quad P_t Y_t - \int_0^{N_t} p_{j,t} y_{j,t} \mathrm{d}j$$

subject to the aggregate production function (6). The first-order condition gives the inverse demand function:

$$\frac{p_{j,t}}{P_t} = \left(\frac{y_{j,t}}{Y_t} \right)^{-\frac{1}{\theta}} N_t^{-1+\gamma\left(\frac{\theta-1}{\theta}\right)}. \tag{7}$$

The price elasticity of demand is constant and given by

$$\varepsilon_{p_j y_j} = -\frac{1}{\theta}, \quad \text{where } \varepsilon_{p_j y_j} \equiv \frac{\partial p_j}{\partial y_j} \frac{y_j}{p_j} = \frac{\partial \ln p_j}{\partial \ln y_j}.$$

The price index, obtained by substituting the demand function into the constraint, is:

$$P_t = N_t^{1-\gamma} \left[\frac{1}{N_t} \int_0^{N_t} p_{j,t}^{1-\theta} \mathrm{d}j \right]^{\frac{1}{1-\theta}}.$$

3.2.2 Intermediate goods firms

The production function of an intermediate goods firm includes internal returns to scale $\nu \in (0,\infty)$ and an output-denominated fixed cost $\phi \in (0,\infty)$.

$$y_{j,t} = A_t l_{j,t}^{\nu} - \phi. \tag{8}$$

The production function has decreasing returns to scale when $\nu \in (0,1)$, constant returns when $\nu = 1$ and, increasing returns when $\nu \in (1,\infty)$. For equilibrium existence the degree of increasing returns is bounded by either the price markup or the degree of external returns to scale.

The individual firm maximises real revenue less real costs, where w_t are real wages:

$$\max_{y_j,p_j,l_j} \quad \frac{p_{j,t}}{P_t} y_{j,t} - w_t l_{j,t}$$

subject to the inverse-demand curve (7) and firm-level production function (8). The optimization conditions simplify to

$$w_t \frac{P_t}{p_{j,t}} = \frac{1}{\mu} A \nu \ell_{j,t}^{\nu-1}. \tag{9}$$

We have defined the price over marginal cost markup as $\mu \equiv \frac{\theta}{\theta-1}$. The relationship states that the 'real product wage' equals the real marginal revenue product of labour.

In order for the second-order condition to hold we require

$$\frac{\nu}{\mu} < \gamma.$$

3.3 Aggregate Accounting and Exogenous Shock

There is a feasibility condition for aggregate labour:

$$L_t = \int_0^{N_t} l_{j,t} \mathrm{d}j. \tag{10}$$

Goods market clearing equates aggregate output to consumption and investment in new firms:

$$Y_t = C_t + E_t + \mathcal{C}(E_t, N_t). \tag{11}$$

The exogenous technology variable evolves according to an AR(1) stochastic process:

$$A_t = e^{\varepsilon_t} A_{t-1}^{\rho_A}, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_A^2). \tag{12}$$

3.4 Planner Problem

The social planner solves the following problem:

$$\max_{L_t, C_t, E_t, N_{t+1}} \quad \sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma} - 1}{1-\sigma} - \chi \frac{L_t^{1+\eta}}{1+\eta} \right)$$

$$\text{s.t.} \quad N_t^{\gamma} (A_t N_t^{-\nu} L_t^{\nu} - \phi) = C_t + E_t + \mathcal{C}(E_t, N_t) \tag{13}$$

$$N_{t+1} = E_t + (1-\delta) N_t. \tag{14}$$

In order for the second-order maximization conditions to hold we require $\nu < \gamma$. The optimization conditions combine to give the following:[13]

$$\chi L_t^{\eta} C_t^{\sigma} = \nu A_t L_t^{\nu-1} N_t^{\gamma-\nu} \tag{15}$$

$$s_t = 1 + \psi \left(\frac{E_t}{N_t} - \delta \right) \tag{16}$$

$$s_t = \beta \mathbb{E}_t \left\{ \frac{\lambda_{t+1}}{\lambda_t} \left[\underbrace{N_{t+1}^{\gamma-1} ((\gamma - \nu) A_{t+1} N_{t+1}^{-\nu} L_{t+1}^{\nu} - \gamma \phi)}_{\text{Marginal product of a firm} Y_{N,t+1}} - \mathcal{C}_N(E_{t+1}, N_{t+1}) + s_{t+1}(1-\delta) \right] \right\} \tag{17}$$

As before s_t is the ratio of Lagrange multipliers on the two constraints, which represents the marginal utility of consumption and marginal utility of investment in a firm. We refer to s_t as the relative price of a firm. The term we label 'marginal product of a firm', is the marginal effect of one additional firm on aggregate output, where labour is treated as constant:

$$Y_{N,t} \equiv \frac{\partial \left[A_t N_t^{\gamma-\nu} L_t^{\nu} - N_t^{\gamma} \phi \right]}{\partial N} = N_t^{\gamma-1} [(\gamma - \nu) A_t N_t^{-\nu} L_t^{\nu} - \gamma \phi].$$

If we forward iterate (17), we have

$$s_t = \beta \mathbb{E}_t \sum_{i=1}^{\infty} (\beta(1-\delta))^{i-1} \left(\frac{C_t}{C_{t+i}} \right)^{\sigma} [Y_N(N_{t+i}, L_{t+i}) - \mathcal{C}_N(E_{t+i}, N_{t+i})].$$

[13] We can also represent (15) in terms of the relative price of intermediate goods to finals goods $\rho(N_t)$ as follows: $\chi L_t^{\eta} C_t^{\sigma} = \nu A_t L_t^{\nu-1} \rho(N_t)^{\frac{\gamma-\nu}{\gamma-1}}$. This generalizes Bilbiie, Ghironi, and Melitz (2019, Eq. 9) for the case where $\nu \neq 1$.

4 Analysis

4.1 Centralised and Decentralised Comparison

The centralised equilibrium conditions differ from the decentralised equilibrium conditions in two ways. First, there is no wedge in the labour market. Second, in the Euler equation operating profits in the decentralised model are replaced with the marginal product of a firm in the centralised problem. If $\mu = \gamma = 1$ the equilibrium conditions are identical because profits in the decentralised economy equal to the marginal product of a firm in the centralised economy, and there is no markup wedge in the static labour equation.

Table 2: Comparison of Decentralised and Centralised Equilibrium Conditions

Decentralised	Centralised
$N_t^{\gamma}(A_t N_t^{-\nu} L_t^{\nu} - \phi) = C_t + E_t + \mathcal{C}(E_t, N_t)$	$N_t^{\gamma}(A_t N_t^{-\nu} L_t^{\nu} - \phi) = C_t + E_t + \mathcal{C}(E_t, N_t)$
$N_{t+1} = E_t + (1-\delta)N_t$	$N_{t+1} = E_t + (1-\delta)N_t$
$\chi L_t^{\eta} C_t^{\sigma} = \frac{\nu}{\mu} A_t L_t^{\nu-1} N_t^{\gamma-\nu}$	$\chi L_t^{\eta} C_t^{\sigma} = \nu A_t L_t^{\nu-1} N_t^{\gamma-\nu}$
$s_t = 1 + \mathcal{C}_E(E_t, N_t)$	$s_t = 1 + \mathcal{C}_E(E_t, N_t)$
$s_t = \beta \mathbb{E}_t \left\{ \left(\frac{C_t}{C_{t+1}}\right)^{\sigma} \left[N_{t+1}^{\gamma-1} \left(\left(1 - \frac{\nu}{\mu}\right) A_{t+1} L_{t+1}^{\nu} N_{t+1}^{-\nu} - \phi \right) + X_t \right] \right\}$	$s_t = \beta \mathbb{E}_t \left\{ \left(\frac{C_t}{C_{t+1}}\right)^{\sigma} \left[N_{t+1}^{\gamma-1} \left((\gamma - \nu) A_{t+1} N_{t+1}^{-\nu} L_{t+1}^{\nu} - \gamma\phi \right) + X_t \right] \right\}$
$s_t = \beta \mathbb{E}_t \sum_{i=1}^{\infty} (\beta(1-\delta))^{i-1} \left(\frac{C_t}{C_{t+i}}\right)^{\sigma} [\pi_{t+i} - \mathcal{C}_{N,t+i}]$	$s_t = \beta \mathbb{E}_t \sum_{i=1}^{\infty} (\beta(1-\delta))^{i-1} \left(\frac{C_t}{C_{t+i}}\right)^{\sigma} [Y_{N,t+i} - \mathcal{C}_{N,t+i}]$

To shorten the Euler conditions, we have defined $X_t \equiv -\mathcal{C}_N(E_{t+1}, N_{t+1}) + (1-\delta)s_{t+1}$. The final row compares the forward iteration of the Euler equations. This shows that the entry cost depends on the present discounted profits in the decentralised economy. In the centralised economy, the entry cost depends on the present discounted value of a firm's contribution to aggregate output.

4.2 Steady-state Analysis

4.2.1 Steady-state Analytical Results

Table 3: Comparison of Decentralised and Centralised Steady-state Conditions

Decentralised	Centralised
$\tilde{s} = 1$	$\tilde{s} = 1$
$\tilde{N}^{\gamma-1}\left[\left(1 - \frac{\nu}{\mu}\right)\tilde{A}\tilde{N}^{-\nu}\tilde{L}^{\nu} - \phi\right] = \tilde{\pi}$	$\tilde{N}^{\gamma-1}\left[(\gamma - \nu)\tilde{A}\tilde{N}^{-\nu}\tilde{L}^{\nu} - \gamma\phi\right] = \tilde{\pi}$
$\tilde{C} = \tilde{N}^{\gamma}(\tilde{A}\tilde{N}^{-\nu}\tilde{L}^{\nu} - \phi) - \delta\tilde{N}$	$\tilde{C} = \tilde{N}^{\gamma}(\tilde{A}\tilde{N}^{-\nu}\tilde{L}^{\nu} - \phi) - \delta\tilde{N}$
$\tilde{C}^{\sigma} = \frac{1}{\mu}\frac{\nu}{\chi}\tilde{A}\tilde{L}^{\nu-1-\eta}\tilde{N}^{\gamma-\nu}$	$\tilde{C}^{\sigma} = \frac{\nu}{\chi}\tilde{A}\tilde{L}^{\nu-1-\eta}\tilde{N}^{\gamma-\nu}$
$\tilde{E} = \delta\tilde{N}$	$\tilde{E} = \delta\tilde{N}$

Since steady-state profits and the marginal product of a firm are equal to the same constant in steady state $\tilde{\pi} = \tilde{Y}_N = \frac{1}{\beta} - (1-\delta)$, we use $\tilde{\pi}$ notation for the constant, even though the profits variable is not defined in the centralised problem.

The comparative statics are as follows

	Exog. Labour $\bar{L}$		Endog. Labour L_t	
	$\mu = 1$	$\mu > 1$	$\mu = 1$	$\mu > 1$
$\gamma = 1$	$\tilde{N}^{\mathrm{P}} = \tilde{N}^{\mathrm{M}}$	$\tilde{N}^{\mathrm{P}} < \tilde{N}^{\mathrm{M}}$	$\tilde{N}^{\mathrm{P}} = \tilde{N}^{\mathrm{M}}$	$\tilde{N}^{\mathrm{P}} \gtreqless \tilde{N}^{\mathrm{M}}$
$\gamma > 1$	$\tilde{N}^{\mathrm{P}} \gtreqless \tilde{N}^{\mathrm{M}}$	$\tilde{N}^{\mathrm{P}} \gtreqless \tilde{N}^{\mathrm{M}}$	$\tilde{N}^{\mathrm{P}} \gtreqless \tilde{N}^{\mathrm{M}}$	$\tilde{N}^{\mathrm{P}} \gtreqless \tilde{N}^{\mathrm{M}}$

Table 4: Market versus Planner Outcomes in Steady State

4.3 Steady State Endogenous Labour

The endogenous labour case is harder to study analytically. It is similar to the exogenous labour case if we assume that there is a wage tax, removing the static distortion.

With endogenous labour, it is harder to determine whether the planner number of firms is greater than or less than the market number of firms. When there are constant external returns, the derivative of $\tilde{N}^M$ with respect to μ is concave, but it is increasing in a neighbourhood of $\mu = 1$. Therefore, if we are in a Pareto optimal position, the introduction of a small markup will increase the number of firms away from optimality. However, the introduction of a sufficiently large markup will lead to fewer firms in the market outcome than the Pareto optimal outcome. Furthermore, there exists an $\mu^{\text{planner N}} \in (1,\infty)$ that is strictly greater than 1 and yields $\tilde{N}^M = \tilde{N}^P$. This is an example of the theory of the second best (Lipsey and Lancaster 1956) since although the optimal number of firms can exist in the presence of a markup, it will not yield optimality via other conditions.

4.3.1 General Case $\mu > 1$ and $\gamma > 1$

The solution to steady-state $\tilde{N}$ satisfies the following non-linear equations in each case:

$$\frac{\nu}{\mu\chi}\left(\left(\frac{1}{\beta}-(1-\delta)\right)\tilde{N}^{M,1-\gamma}+\phi\right)^{1-\frac{1+\eta}{\nu}}\tilde{N}^{M,\gamma-\sigma-(1+\eta)} = \left(1-\frac{\nu}{\mu}\right)^{-\sigma+1-\frac{1+\eta}{\nu}}\left(\frac{\nu}{\mu}\phi\tilde{N}^{M,\gamma-1}+\frac{1}{\beta}-(1-\delta)-\delta\left(1-\frac{\nu}{\mu}\right)\right)^{\sigma} \tag{Market (M)}$$

$$\frac{\nu}{\chi}\left(\left(\frac{1}{\beta}-(1-\delta)\right)\tilde{N}^{P,1-\gamma}+\gamma\phi\right)^{1-\frac{1+\eta}{\nu}}\tilde{N}^{P,\gamma-\sigma-(1+\eta)} = (\gamma-\nu)^{-\sigma+1-\frac{1+\eta}{\nu}}\left(\nu\phi\tilde{N}^{P,\gamma-1}+\frac{1}{\beta}-(1-\delta)-\delta(\gamma-\nu)\right)^{\sigma} \tag{Planner (P)}$$

The main messages from the exogenous labour case hold in the endogenous labour case. The planner and market outcomes are equivalent when there are no distortions

$\gamma = \mu = 1$. And for the case $\gamma = 1$, the steady-state number of firms has an analytical solution.

4.3.2 Steady-state Quantitative Results

Table 5 presents our benchmark calibration.

Table 5: Benchmark Calibration

Parameter	Value	Description
β	0.99	Discount factor
σ	2.0	Intertemp. Elast. Sub.
η	0.5	Frisch elasicity
γ	1.05	External returns to scale
ν	0.9	Internal returns to scale
ϕ	0.1	Fixed costs
δ	0.02	Firm death rate
χ	1	Labour disutility
ψ	2	Adjustment cost
μ	1.1	Markup

We vary parameters ψ, μ, γ in our exercises.

Our calibration follows the literature (Bilbiie, Ghironi, and Melitz 2019). The adjustment cost parameter is a distinguishing feature of the model. A $\psi = 2$ corresponds to an AR(1) coefficient in equation (18) of $\omega = 0.5$, and therefore a half-life of $t = 1$ quarter, which is consistent with the results of Table 1. The AR(1) relationship that we estimate and report results on in Table 1 can be obtained as a reduced-form from our model equations. If we log-linearize the planner system, we can reduce it to the following form:

$$\hat{N}_{t+1} = \omega \hat{N}_t. \tag{18}$$

The coefficient ω is a function of model parameters including ψ. Figure 1 shows how this coefficient varies with ψ. This allows us to study a range of ψ that yield similar AR(1) coefficients to the estimated results in Table 1.

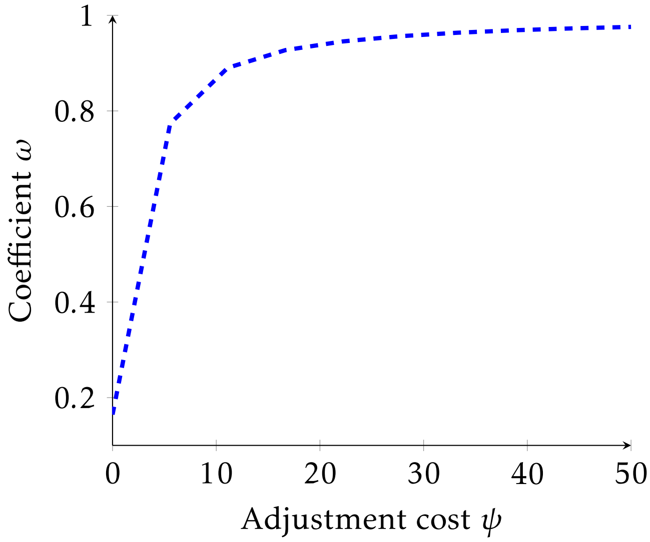


Figure 1: Policy function coefficients

The plot shows how the adjustment cost paramter ψ affects the AR(1) coefficient in equation (18).

In Table 6 we study a numerical illustration of steady state and compare welfare.[14] We define welfare as the sum of future discounted utilities

$$W_t = \sum_{T=t}^{\infty} \beta^T U_T = U_t + \beta W_{t+1}.$$

W_0 is all future discounted utilities from the initial period $t = 0$. This is *conditional* welfare:

$$W_0 = \sum_{T=0}^{\infty} \beta^T U_T = U_0 + \beta W_1.$$

Steady-steady welfare is

$$\tilde{W} = \sum_{T=t}^{\infty} \beta^T \tilde{U} = \frac{\tilde{U}}{1-\beta}.$$

Our welfare calculations compare two scenarios. Scenario 1 corresponds to W_0. The economy begins at a distorted steady state, but at $t = 0$ it jumps to the planner economy and transitions to the planner steady state. Scenario 2 corresponds to $\tilde{W}$. The economy begins at the distorted steady state and remains there forever. The percentage gains represent the welfare gain from policy that switches the economy from the distorted to the undistorted world. In other words, it is the social cost of distortion or potential gain from policy.

[14]Our analysis of welfare is similar to Turnovsky (2000) which is appropriate due to the ordinality of our utility functions. An alternative is to study consumption-equivalent welfare.

Table 6: Steady state comparison

	$\gamma = 1.05$			$\gamma = 1$		
		$\mu = 1.1$	$\mu = 1$		$\mu = 1.1$	$\mu = 1$
	1. Planner	2. Market	3. Market	4. Planner	5. Market	6. Market
$\tilde{N}$	1.20	1.52	0.78	0.78	1.53	0.78
$\tilde{E}$	0.02	0.03	0.02	0.02	0.03	0.02
$\tilde{L}$	1.07	1.04	1.05	1.05	1.05	1.05
$\tilde{C}$	0.94	0.92	0.92	0.92	0.91	0.92
$\tilde{W}$	-79.37	-79.66	-81.10	-79.76	-81.97	-79.76
W_0	-79.37	-78.99	-79.84	-79.76	-78.99	-79.76
Welf. Cost	0%	0.84%	1.55%	0%	3.64%	0%

Notes:

1. The table presents a comparison of steady state solutions in four scenarios: when external returns to scale exist and mark-up is positive or one, and when external returns to scale are absent and mark-up is positive or one. The comparison is between the Planner's steady state and the market steady state for the variables N (total firms), E (firm entry), L (aggregate labour), and C (aggregate consumption). Also note that the Planner's solution is independent of the mark-up μ but depends on external returns to scale γ.
2. W_0 is the sum of future discounted utility from moving from the distorted steady state to the planner steady state. The speed of adjustment is fixed in all these exercises to $\psi = 2$, but the steady states change according to μ and γ. Since in columns 1,4,6 the economy is already at the undistorted steady state, it cannot be moved so W_0 is equivalent to steady-state W, and the potential welfare gain from policy is 0.
3. Welfare cost row is $\left(1 - \frac{W_0}{W}\right) \times 100$.

The maximum welfare occurs when a planner allocates resources in an economy with external returns to scale. In this economy, a planner can achieve a higher level of welfare, than a planner could achieve if external returns were not present. If external returns are not present $\gamma = 1$, then removing the markup allows the market outcome to achieve the planner outcome. This follows the First Fundamental Theorem of Welfare Economics. That is, in the absence of distortions, the competitive (decentralised) equilibrium and Pareto (centralised) equilibrium coincide. However, with increasing external returns $\gamma > 1$, a higher markup moves the market outcome closer to the planner outcome because it decreases firm size and raises the number of firms.

With monopolistic competition and endogenous labour supply, there is an asymmetry in markups applied to goods/services that affect utility. The two goods are consumption and leisure. Whilst consumption costs price P leisure costs (nominal) w, and w is decreased by the markup. Hence there is a relative price difference. Consumption is more expensive than leisure. This causes excessive demand for leisure in the market outcome compared to the planner outcome and insufficient demand for consumption and labour. This is visible in columns 4 and 5. Therefore optimal policy needs to address the relative price distortion between goods and leisure.

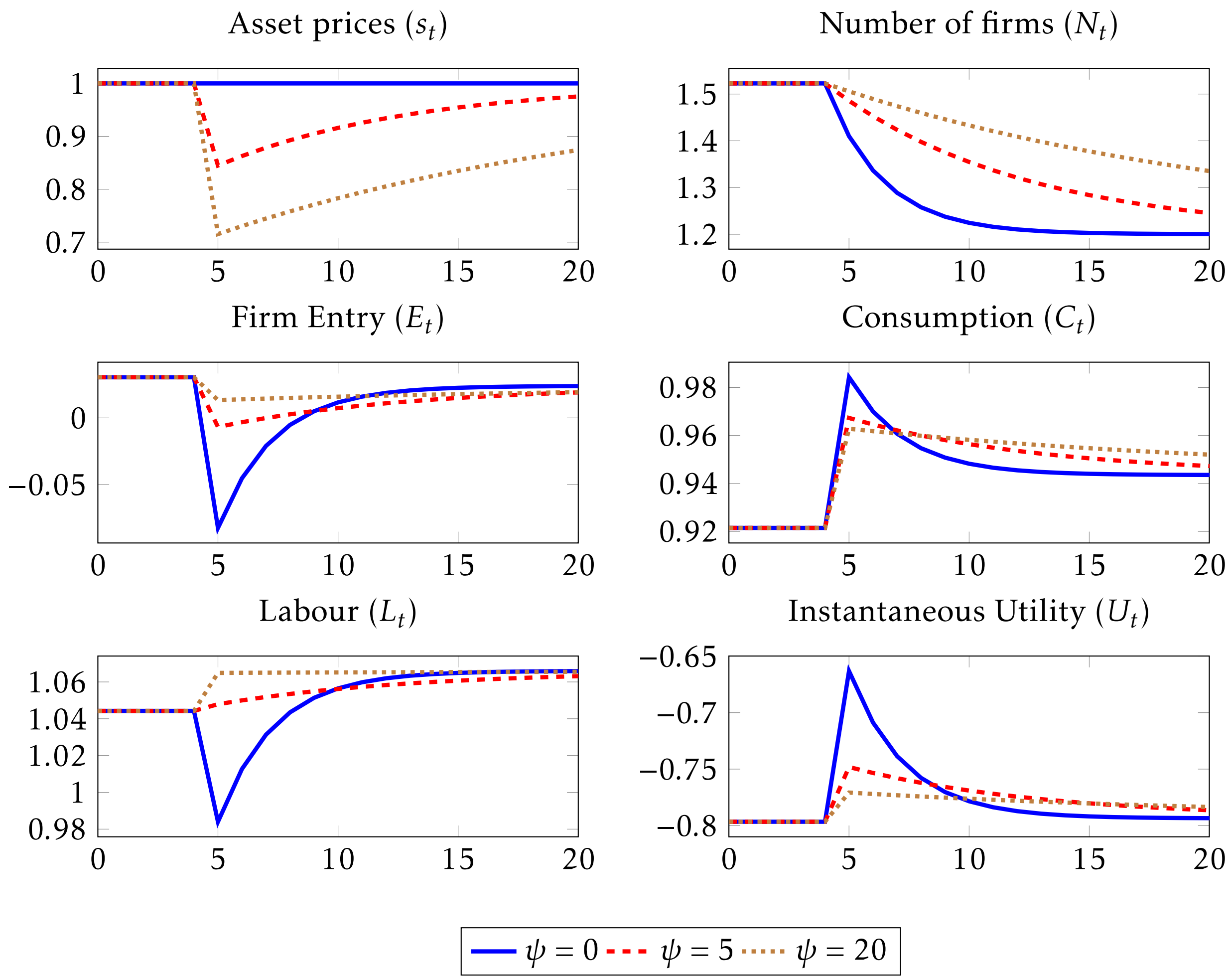
4.4 Dynamic Analysis

Figure 2 shows the transition paths of model variables moving from the distorted steady state to the undistorted steady state. The economy begins in the steady state that arises when $\mu = 1.1$, $\gamma = 1.05$ and it transitions to the planner steady state, under planner dynamics, which yields the steady-state corresponding to $\mu = 1$, $\gamma = 1.05$. In other words, the economy begins at the steady-state corresponding to column 2 in Table 6 and converges to the steady-state values in column 1 of Table 6. The three lines in each figure represent different speeds of adjustment. The three transition paths always begin and and end at the same steady-states because changing ψ does not affect steady states. However, it affects transition dynamics at a first-order approximation because changing ψ affects the eigenvalues of the model.

Number of firms N_t is the only state variable in our model. This is shown by the gradual adjustment from $t = 5$, whereas all other variables jump in $t = 5$ when the economy is shifted to the planner's tranisiton dynamics.

The dynamics are most responsive in the case of no entry adjustment costs $\psi = 0$. In this case the relative price of a firm is constant and equal to one, meaning that the marginal utility from an additional firm is always equal to the marginal utility from additional consumption.

Figure 2: Transition Paths from Distorted to Undistorted Steady State



The plot shows the convergence of variables from the decentralised steady state to the centralised steady state under the dynamics of the centralised model. Convergence is faster for lower values of ψ. The y-axis values are in levels. The variables begin at the decentralised steady-state corresponding to column 2 in Table 6 and they converge upon the steady-state values in column 1 of Table 6.

In Figure 3 we show that for slower speeds of adjustment (higher dynamic distortion), the welfare gains from moving from the distorted to the undistorted steady state are smaller when the speed of adjustment is slower. This occurs because with faster speed of adjustment, we reach the efficient steady-state faster and the utility gains during transition are discounted less because they occur sooner.

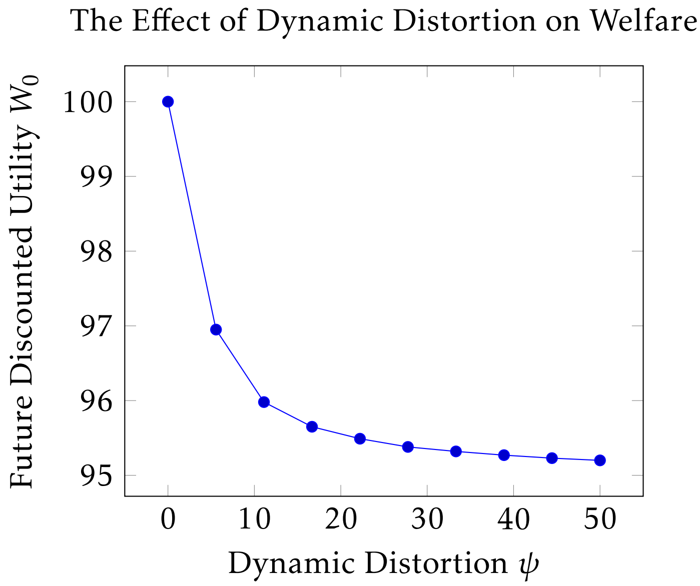


Figure 3: Welfare Effects of Slow Firm Adjustment

The points on the graph show the change in welfare moving from the distorted steady-state to the undistorted (planner) steady state, including the transition period, for different values of the speed of adjustment ψ. The y-axis values are benchmarked against the welfare cost when there are no adjustment costs $\psi = 0$. A high dynamic distortion (slow speed of adjusment) increases the welfare costs of being in a distorted steady state, relative to if there are no adjustment costs.

Figure 4 presents the trade-off between dynamic and static distortions. It emphasizes that low static distortions can be undone by higher dynamic distortions, and vice-versa. When static markups are high, a small increase in the dynamic distortion requires a large fall in markups to maintain welfare at W_0. However, in a low markup economy, a small fall in markups requires a large increase in dynamic distortions to be indifferent. More precisely, an individual in an economy with a markup of 1.23 and a $\psi = 2$, which corresponds to an AR(1) coefficient of 0.5 and a half-life of 1 quarters, is just as well off as an individual in an economy with markup 1.1 and a $\psi = 20$, which corresponds to an AR(1) coefficient of 0.72 and a half-life of 2 quarters. As an analogy, we might conclude that while the US has high markups, at least its economy adjusts quickly. In contrast, Europe has lower markups, but its slow-moving economy might cancel out that advantage.

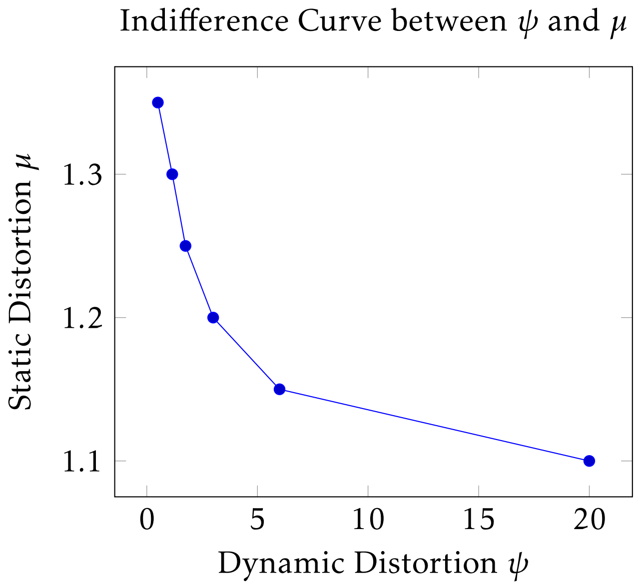


Figure 4: Iso-Utility for μ, ψ combinations

The line is a locus of constant welfare W_0, which represents future discounted utility from $t = 0$ to $t = \infty$. That is, a fixed welfare level from moving from the distorted steady-state to the undistorted (efficient) steady state, including the transition period. We fix $W_0 = -571$ along the locus. The slope of the line captures how much dynamic distortion is worth in terms of static distortion.

5 Conclusion

We analyse optimal policy in the presence of dynamic firm entry, returns to scale and markups. Our model includes constant price markups, and both internal and external returns to scale, which allows scenarios where perfect competition in conjunction with constant external returns to scale leads the decentralised, market, outcome to yield a Pareto optimal, planner, outcome.

Our static analysis stresses that markups are not always detrimental to welfare. Instead, in the presence of external returns to scale, markups can enhance welfare as they encourage greater firm entry and exploitation of scale effects.

Our dynamic analysis shows that policy aimed at speeding-up firm entry into markets can generate welfare gains equivalent to reducing static markups. We conclude that policymakers should promote dynamic competition and reduce barriers to entry, in addition to focusing on static competition measures such as markups.

References

Bahaj, Saleem, Sophie Piton, and Anthony Savagar (Mar. 2024). "Business creation during COVID-19". In: *Economic Policy* 39.119, pp. 611–648.

Baqaee, David and Emmanuel Farhi (May 2020). *Entry vs. Rents: Aggregation with Economies of Scale*. Working Paper 27140. National Bureau of Economic Research.

Barro, Robert J (July 2024). *Markups and Entry in a Circular Hotelling Model*. Tech. rep. 32660. National Bureau of Economic Research.

Bergin, James and Dan Bernhardt (2008). "Industry dynamics with stochastic demand". In: *The RAND Journal of Economics* 39.1, pp. 41–68.

Bilbiie, Florin, Fabio Ghironi, and Marc Melitz (2019). "Monopoly power and endogenous product variety: Distortions and remedies". In: *American Economic Journal: Macroeconomics* 11.4, pp. 140–74.

Decker, Ryan and John Haltiwanger (2023). "Surging Business Formation in the Pandemic: Causes and Consequences". In: *Brookings Papers on Economic Activity*, pp. 3–24.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu (2023). "How Costly Are Markups?" In: *Journal of Political Economy* 131.7, pp. 1619–1675.

Ershov, Daniel (2024). "Variety-Based Congestion in Online Markets: Evidence from Mobile Apps". In: *American Economic Journal: Microeconomics* 16.2, pp. 180–203.

Lipsey, Richard and Kelvin Lancaster (1956). "The General Theory of Second Best". In: *The Review of Economic Studies* 24.1, pp. 11–32.

Matsuyama, Kiminori (May 1991). "Increasing Returns, Industrialization, and Indeterminacy of Equilibrium". In: *The Quarterly Journal of Economics* 106.2, pp. 617–650.

Morrow, John and Swati Dhingra (2014). "Monopolistic competition and optimum product diversity under firm heterogeneity". In: *Journal of Political Economy*.

Turnovsky, Stephen J (2000). "Fiscal policy, elastic labor supply, and endogenous growth". In: *Journal of Monetary Economics* 45.1, pp. 185–210.

A Data Description: Establishment Time Series

Data is for number of establishments recorded in the BLS Quarterly Census of Employment and Wages (QCEW) program. The QCEW provides a series on the count of establishments at a quarterly frequency by industry, geography, and ownership type since 2001 Q1. We focus on supersector industries (Table 7), geography: total US, and ownership: private. We use data from 2001 Q1 up to (and including) 2019 Q1 which implies we have $t = 73$ time-series observations for each industry. There are eight industries. Table 7 outlines the Quarterly Census of Employment and Wages (QCEW) definitions of sector. We analyse 8 of the possible 12 supersectors. *'Public Administration'* (1028) is not in our dataset because we focus on *private* ownership. Also I discard *'Education and Health Services'* (1025), *'Other Services'* (1027) and *'Unclassified'* (1029). 1025 and 1027 have clear breaks (measurement induced) as underlying sectoral data is re-defined by the BLS. [15] *'Unclassified'* (1029) is a catch-all industry that we wouldn't expect to exhibit a common speed of adjustment as it does not reflect a common market structure – some sub-industries of 1029 will be fast adjusting and others will be slow adjusting.

Table 7: Quarterly Census of Employment and Wages (QCEW) Industry Definitions

Total	Domain	Super-Sector	NAICS-Sector
		1011 Natural Resources and Mining	11 Agriculture, Forestry, Fishing, and Hunting
	101 Goods-Producing		21 Mining
		1012 Construction	23 Construction
		1013 Manufacturing	31-33 Manufacturing
			42 Wholesale Trade
		1021 Trade, Transportation, and Utilities	44-45 Retail Trade
			48-49 Transportation and Warehousing
			22 Utilities
		1022 Information	51 Information
		1023 Financial Activities	52 Finance and Insurance
10 Total, All Industries			53 Real Estate and Rental and Leasing
			54 Professional, Scientific and Technical Services
	102 Service-Providing	1024 Professional and Business Services	55 Management of Companies and Enterprises
			56 Administrative and Waste Services
		1025 Education and Health Services	61 Educational Services
			62 Health Care and Social Assistance
		1026 Leisure and Hospitality	71 Arts, Entertainment, and recreation
			72 Accommodation and Food Services
		1027 Other Services	81 Other Services (Except Public Administration)
		1028 Public Administration	92 Public Administration
		1029 Unclassified	99 Unclassified

[15] For the QCEW data on number of private establishments in *'Other Services'* (1027) and *'Education and Health Services'* (1025) there are large changes in 2012 Q4 to 2013 Q1. For the case of 'Education and Health' the number increases from 939,567 to 1,414,242 and for 'Other Services' it decreases 1,279,184 to 7,878,73. The QCEW confirmed that in 2013 Q1 the two sectors were affected by an improvement in reporting. Full details see BLS News Release (26/10/2013), p.5 Notable Industry Changes. 814110 ('Private Households') which falls under Other Services and 624120 ('Services for the Elderly and Persons with Disabilities') which falls under Education and Health Services were affected. The BLS shifted non-medical, home-based services for the elderly and persons with disabilities from 814110 ('Private Households') and re-classified these establishments into services for the elderly and persons with disabilities (NAICS 624120). Hence the drop in 1027 and rise in 1025.

B Household Problem Further Details

The current value Lagrangian for the market (decentralised) problem is as follows:

$$\begin{aligned}
\mathcal{L}^M_{C_t,L_t,E_t,N_{t+1}} = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \Bigg\{ & \left(\frac{C_t^{1-\sigma}-1}{1-\sigma} - \chi \frac{L_t^{1+\eta}}{1+\eta} \right) \\
& + \lambda_t \left(w_t L_t + \pi_t N_t - \left[C_t + E_t + \mathcal{C}\left(E_t, N_t\right) \right] \right) \\
& + \varpi_t (E_t + (1-\delta) N_t - N_{t+1}) \Bigg\}
\end{aligned}$$

The first-order conditions for C_t, L_t, E_t, N_{t+1} are:

$$C_t^{-\sigma} = \lambda_t \tag{19}$$

$$\chi L_t^{\eta} = \lambda_t w_t \tag{20}$$

$$\varpi_t = \lambda_t \left[1 + \psi \left(\frac{E_t}{N_t} - \delta \right) \right] \tag{21}$$

$$\varpi_t = \beta \mathbb{E}_t \left\{ \lambda_{t+1} \left[\pi_{t+1} - \mathcal{C}_N \left(E_{t+1}, N_{t+1} \right) \right] + (1-\delta) \varpi_{t+1} \right\} \tag{22}$$

The Lagrange multiplier ϖ_t is the marginal utility of having an extra firm and λ_t is the marginal utility of consumption. So, $\frac{\varpi_t}{\lambda_t} \equiv s_t$ is the amount of consumption given up for an extra firm, or the relative price of a firm in terms of consumption.

C Relationship between firm asset price and firm value

The sale price of a firm s_t relates to firm value v_t as follows:

$$v_t = d_t + (1-\delta) s_t$$

where we define dividends as operating profits less adjustment costs $d_t \equiv \pi_t - \mathcal{C}_N(E_t, N_t)$. Therefore, the value of a firm v_t is current dividends d_t plus the sale price of a (surviving) firm $(1-\delta)s_t$, where sale price s_t equals future discounted profits. Therefore the Bellman equation is

$$v_t = d_t + m_{t,t+1} v_{t+1} = \mathbb{E}_t \sum_{i=0}^{\infty} m_{t,t+i} d_{t+i}$$

where the stochastic discount factor (SDF) is:

$$m_{t,t+i} = [\beta(1-\delta)]^i \frac{\lambda_{t+i}}{\lambda_t}, \quad i = 0, 1, 2 \cdots .$$

Hence, we can write the Bellman equation as follows

$$v_t = d_t + \beta(1-\delta)\frac{\lambda_{t+1}}{\lambda_t} v_{t+1}.$$

D Decentralised Equilibrium Conditions and Steady State

Under symmetry, all firms choose the same price p_t and hence the same amount of labour l_t. The endogenous variables are $\{\ell_t, s_t, \pi_t, w_t, y_t, N_t, E_t, C_t, L_t, Y_t, P_t, p_t\}$ and the exogenous variable is the level of technology A_t. We specify the final good at the numeraire such that $P_t = 1$. Therefore, the full symmetric equilibrium conditions are a system of 12 equations in 12 variables:

$$P_t = 1 \tag{23}$$

$$C_t + E_t + \mathcal{C}(E_t, N_t) = w_t L_t + \pi_t N_t \tag{24}$$

$$N_{t+1} = E_t + (1-\delta) N_t \tag{25}$$

$$\chi L_t^{\eta} C_t^{\sigma} = w_t \tag{26}$$

$$s_t = 1 + \mathcal{C}_E(E_t, N_t) \tag{27}$$

$$s_t = \beta \mathbb{E}_t \left\{ \left(\frac{C_t}{C_{t+1}} \right)^{\sigma} (\pi_{t+1} - \mathcal{C}_N (E_{t+1}, N_{t+1}) + (1-\delta) s_{t+1}) \right\} \tag{28}$$

$$\frac{w_t}{p_t} = \frac{1}{\mu} A_t \nu \ell_t^{\nu-1} \tag{29}$$

$$y_t = A_t \ell_t^{\nu} - \phi \tag{30}$$

$$Y_t = N_t^{\gamma} y_t \tag{31}$$

$$L_t = N_t \ell_t \tag{32}$$

$$Y_t = C_t + E_t + \mathcal{C}(E_t, N_t) \tag{33}$$

$$1 = N_t^{1-\gamma} p_t \tag{34}$$

$$\ln A_t = \rho_A \ln A_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_A^2). \tag{35}$$

D.1 Decentralised Steady-state

$$\tilde{P} = 1 \tag{36}$$

$$\tilde{A} = 1 \tag{37}$$

$$\tilde{s} = 1 \tag{38}$$

$$\tilde{\pi} = \frac{1}{\beta} - (1-\delta) \tag{39}$$

$$\tilde{p} = \tilde{N}^{\gamma-1} \tag{40}$$

$$\tilde{y} + \phi = \left(1 - \frac{\nu}{\mu}\right)^{-1} \left(\frac{\tilde{\pi}}{\tilde{p}} + \phi\right) \tag{41}$$

$$\tilde{\ell} = \left(\frac{\tilde{y} + \phi}{\tilde{A}}\right)^{\frac{1}{\nu}} \tag{42}$$

$$\frac{\tilde{w}}{\tilde{p}} = \frac{\nu}{\mu} \tilde{A} \tilde{\ell}^{\nu-1} \tag{43}$$

$$\tilde{E} = \delta \tilde{N} \tag{44}$$

$$\tilde{L} = \tilde{N} \tilde{\ell} \tag{45}$$

$$\tilde{C} = \tilde{N}\left(\tilde{p}\tilde{y} - \delta\right) \tag{46}$$

$$\tilde{w} = \chi (\tilde{N}\tilde{\ell})^{\eta} \tilde{C}^{\sigma} \tag{47}$$

Combining the steady-state decentralised conditions yields a nonlinear equation in $\tilde{N}$:

$$\frac{\nu}{\mu\chi}\left(\left(\frac{1}{\beta} - (1-\delta)\right)\tilde{N}^{1-\gamma} + \phi\right)^{1-\frac{1+\eta}{\nu}} \tilde{N}^{\gamma-\sigma-(1+\eta)} = \left(1 - \frac{\nu}{\mu}\right)^{-\sigma+1-\frac{1+\eta}{\nu}} \left(\frac{\nu}{\mu}\phi\tilde{N}^{\gamma-1} + \frac{1}{\beta} - (1-\delta) - \delta\left(1 - \frac{\nu}{\mu}\right)\right)^{\sigma}. \tag{48}$$

E Planner Problem

The current-value Lagrangian for the Planner's problem is as follows:

$$\mathcal{L}^P = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \Bigg\{ \left(\frac{C_t^{1-\sigma} - 1}{1-\sigma} - \chi \frac{L_t^{1+\eta}}{1+\eta}\right) + \lambda_t \left(N_t^{\gamma}(A_t N_t^{-\nu} L_t^{\nu} - \phi) - C_t - E_t - \mathcal{C}(E_t, N_t)\right) + \varpi_t (E_t + (1-\delta)N_t - N_{t+1}) \Bigg\}$$

The optimization conditions are:

$$\chi L_t^{\eta} = \lambda_t \nu A_t L_t^{\nu-1} N_t^{\gamma-\nu} \tag{49}$$

$$C_t^{-\sigma} = \lambda_t \tag{50}$$

$$\varpi_t = \lambda_t \left[1 + \psi\left(\frac{E_t}{N_t} - \delta\right)\right] \tag{51}$$

$$\varpi_t = \beta \mathbb{E}_t \left\{\lambda_{t+1}\left[N_{t+1}^{\gamma-1}\left(A_{t+1} N_{t+1}^{-\nu} L_{t+1}^{\nu}(\gamma-\nu) - \gamma\phi\right) - \mathcal{C}_N\left(E_{t+1}, N_{t+1}\right)\right] + \varpi_{t+1}(1-\delta)\right\} \tag{52}$$

F Planner Equilibrium Conditions and Steady State

F.1 Planner Equilibrium Conditions

The planner system consists of endogenous variables $\{N_t, C_t, E_t, L_t, s_t\}$ and exogenous $\{A_t\}$. There are five equations corresponding to the five endogenous variables.

$$N_t^{\gamma}(A_t N_t^{-\nu} L_t^{\nu} - \phi) = C_t + E_t + \mathcal{C}(E_t, N_t) \tag{53}$$

$$N_{t+1} = E_t + (1-\delta) N_t \tag{54}$$

$$\chi L_t^{\eta} C_t^{\sigma} = \nu A_t L_t^{\nu-1} N_t^{\gamma-\nu} \tag{55}$$

$$s_t = 1 + \mathcal{C}_E(E_t, N_t) \tag{56}$$

$$s_t = \beta \mathbb{E}_t \left\{\left(\frac{C_t}{C_{t+1}}\right)^{\sigma}\left[Y_N(N_{t+1}, L_{t+1}) - \mathcal{C}_N(E_{t+1}, N_{t+1}) + (1-\delta) s_{t+1}\right]\right\} \tag{57}$$

F.2 Planner steady state

The steady-state equilibrium conditions of the planner are as follows. The planner internalizes the positive effect of external returns to scale $\gamma > 1$ in equation (60).

$$\tilde{A} = 1 \tag{58}$$

$$\tilde{s} = 1 \tag{59}$$

$$\tilde{N}^{\gamma-1}\left[(\gamma-\nu)\tilde{A}\tilde{N}^{-\nu}\tilde{L}^{\nu} - \gamma\phi\right] = \frac{1}{\beta} - (1-\delta) \tag{60}$$

$$\tilde{C} = \tilde{N}^{\gamma}(\tilde{A}\tilde{N}^{-\nu}\tilde{L}^{\nu} - \phi) - \delta\tilde{N} \tag{61}$$

$$\tilde{C}^{\sigma} = \frac{\nu}{\chi}\tilde{A}\tilde{L}^{\nu-1-\eta}\tilde{N}^{\gamma-\nu} \tag{62}$$

$$\tilde{E} = \delta\tilde{N} \tag{63}$$

Combining the steady-state planner conditions yields a nonlinear equation in $\tilde{N}$:

$$\frac{\nu}{\chi}\left(\left(\frac{1}{\beta}-(1-\delta)\right)\tilde{N}^{1-\gamma}+\gamma\phi\right)^{1-\frac{1+\eta}{\nu}}\tilde{N}^{\gamma-\sigma-(1+\eta)}=(\gamma-\nu)^{-\sigma+1-\frac{1+\eta}{\nu}}\left(\nu\phi\tilde{N}^{\gamma-1}+\frac{1}{\beta}-(1-\delta)-\delta(\gamma-\nu)\right)^{\sigma} \tag{64}$$

We solve for $\tilde{N}$ numerically. In the case of $\gamma=1$, there is an analytic solution.

F.3 Planner Problem with Instantaneous Entry

If there are no adjustment costs $\psi = 0$, the household can adjust its stock of firms instantly and costlessly to its desired level. This has the following implications:

1. Entry becomes perfectly flexible. Households can instantly respond to shocks by adjusting their number of firms to the new optimal level. These adjustments are costless.

2. Firm price dynamics are removed. The price of a firm is always equal to one which removes the dynamic entry cost from the Euler condition. The Euler equation follows a standard interpretation: it determines the optimal level of firm creation by balancing the trade-off between consuming today and investing in firms to increase future consumption. The present value of future benefits from firm investment is equal to the present cost of forgoing

3. Entry volatility increases because households react instantly to shocks.[16] In the presence of adjustment costs, households smooth entry over time to avoid incurring large adjustment costs.

If there is no adjustment in firm entry the problem is as follows:

$$\begin{aligned}\max_{L_t,C_t,N_{t+1}} \quad & \sum_{t=0}^{\infty}\beta^t\left(\frac{C_t^{1-\sigma}-1}{1-\sigma}-\chi\frac{L_t^{1+\eta}}{1+\eta}\right)\\ \text{s.t.} \quad & N_t^{\gamma}(A_tN_t^{-\nu}L_t^{\nu}-\phi)=C_t+N_{t+1}-(1-\delta)N_t\end{aligned} \tag{65}$$

The current-value Lagrangian for the Planner's problem with instantaneous adjust is as follows:

$$\mathcal{L}^{P,IA}=\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t\left\{\frac{C_t^{1-\sigma}-1}{1-\sigma}-\chi\frac{L_t^{1+\eta}}{1+\eta}+\lambda_t\left(N_t^{\gamma}(A_tN_t^{-\nu}L_t^{\nu}-\phi)-C_t-N_{t+1}+(1-\delta)N_t\right)\right\}$$

[16] During the COVID-19 pandemic firm entry was volatile, particularly in industries associated with low adjustment costs like online retail (Decker and Haltiwanger 2023; Bahaj, Piton, and Savagar 2024).

The optimization conditions are:

$$\chi L_t^{\eta} = C_t^{-\sigma} \nu A_t L_t^{\nu-1} N_t^{\gamma-\nu} \tag{66}$$

$$1 = \beta \mathbb{E}_t \left\{ \left(\frac{C_t}{C_{t+1}} \right)^{\sigma} \left[N_{t+1}^{\gamma-1} \left((\gamma - \nu) A_{t+1} N_{t+1}^{-\nu} L_{t+1}^{\nu} - \gamma \phi \right) + 1 - \delta \right] \right\} \tag{67}$$

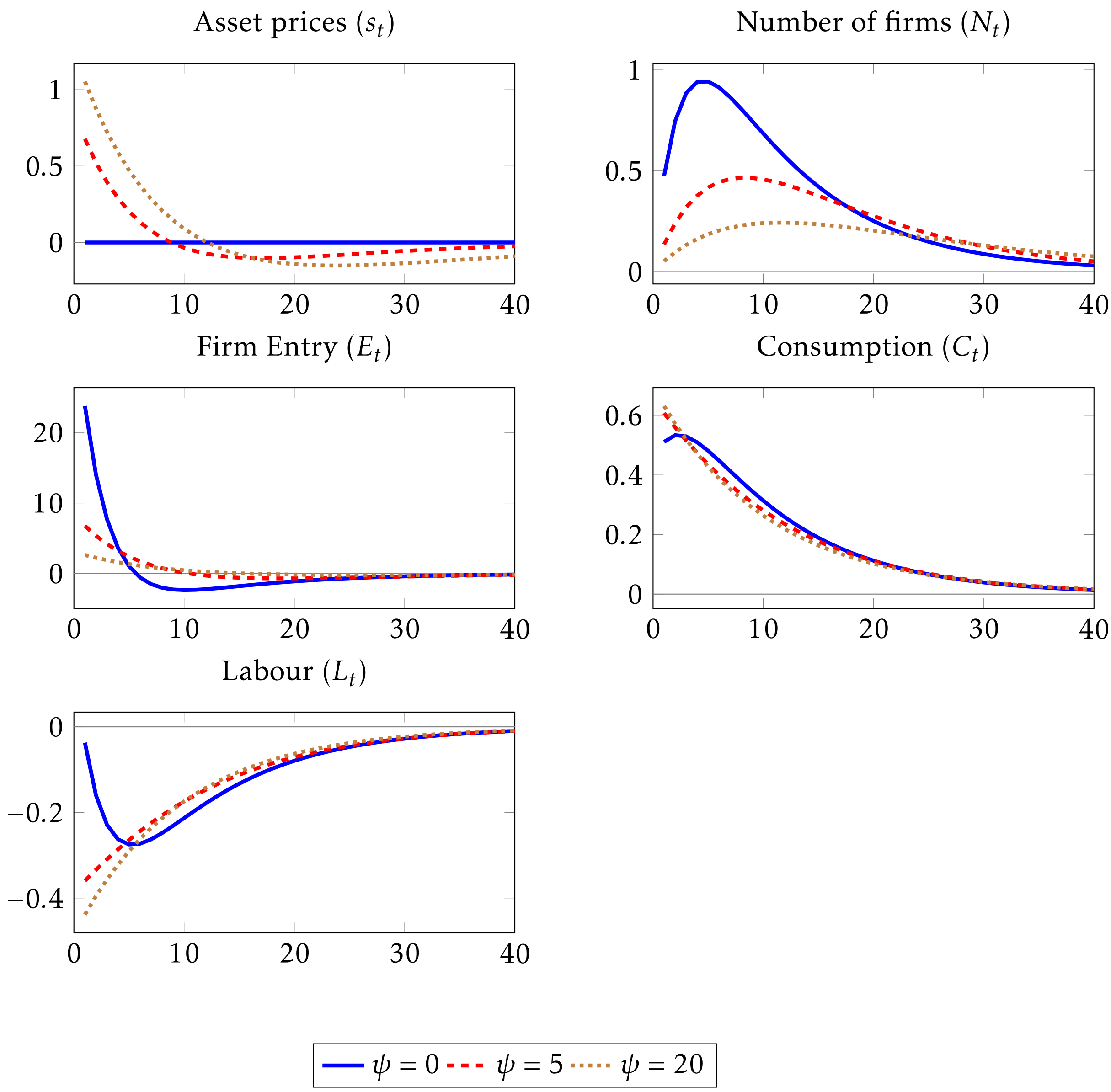
Therefore, there are three equilibrium conditions in three unknown variables C_t, L_t, N_t. By substitution, the system reduces to a single first-order difference equation in N_t.

G Stochastic Simulations

In this section, we show the impulse response functions to a 1 standard deviation technology shock under different model scenarios. The model is solved as a first-order log-linear approximation. The values on the y-axis represent the percentage deviation from steady state.

G.1 Centralised model

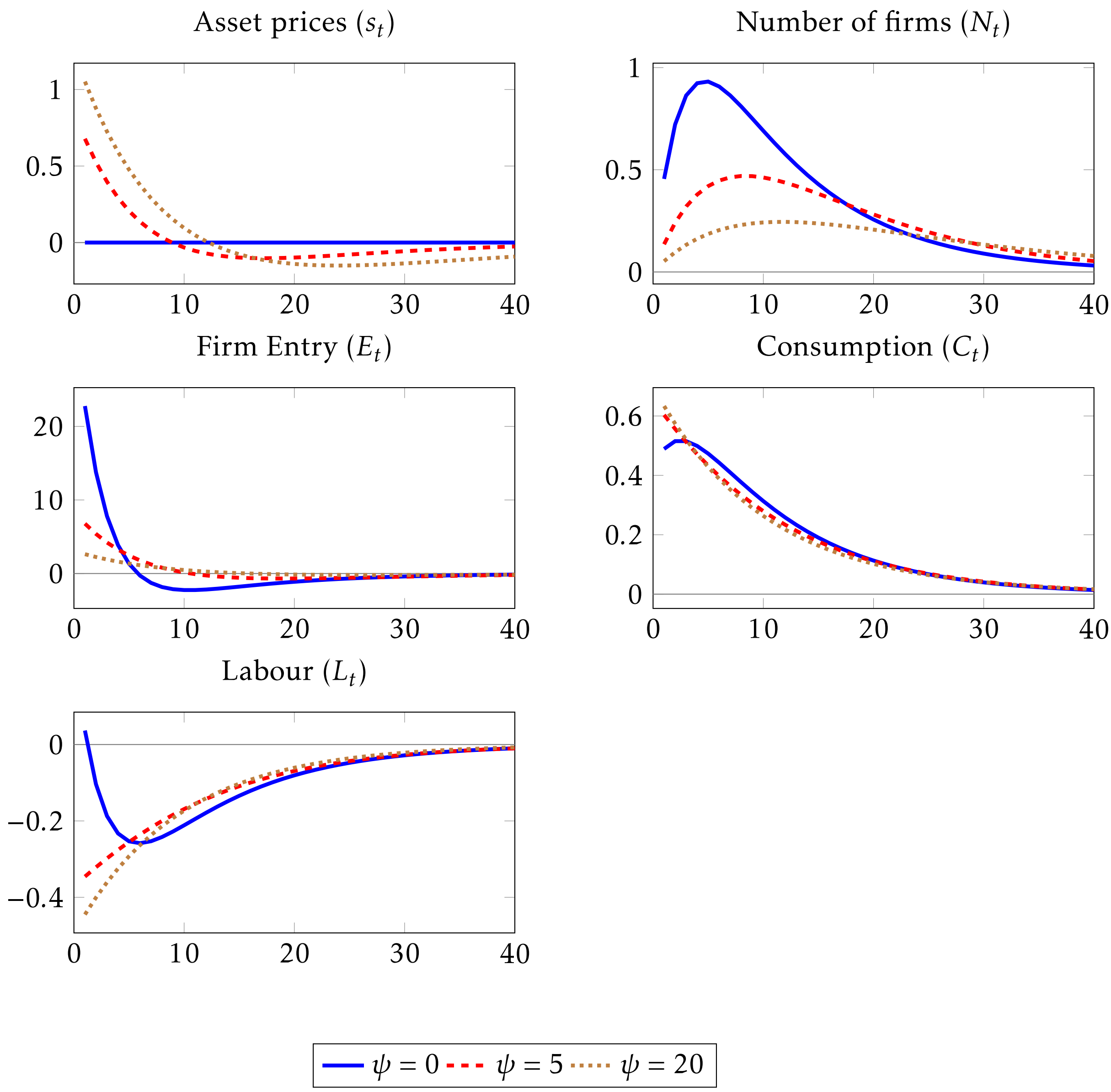
Figure 5: Technology shock IRFs for different adjustment costs (centralised)



The plot shows response of variables to a technology shock for different values of the adjustment costs ψ.

G.2 Decentralised model

Figure 6: Technology shock IRFs for different adjustment costs (decentralised)

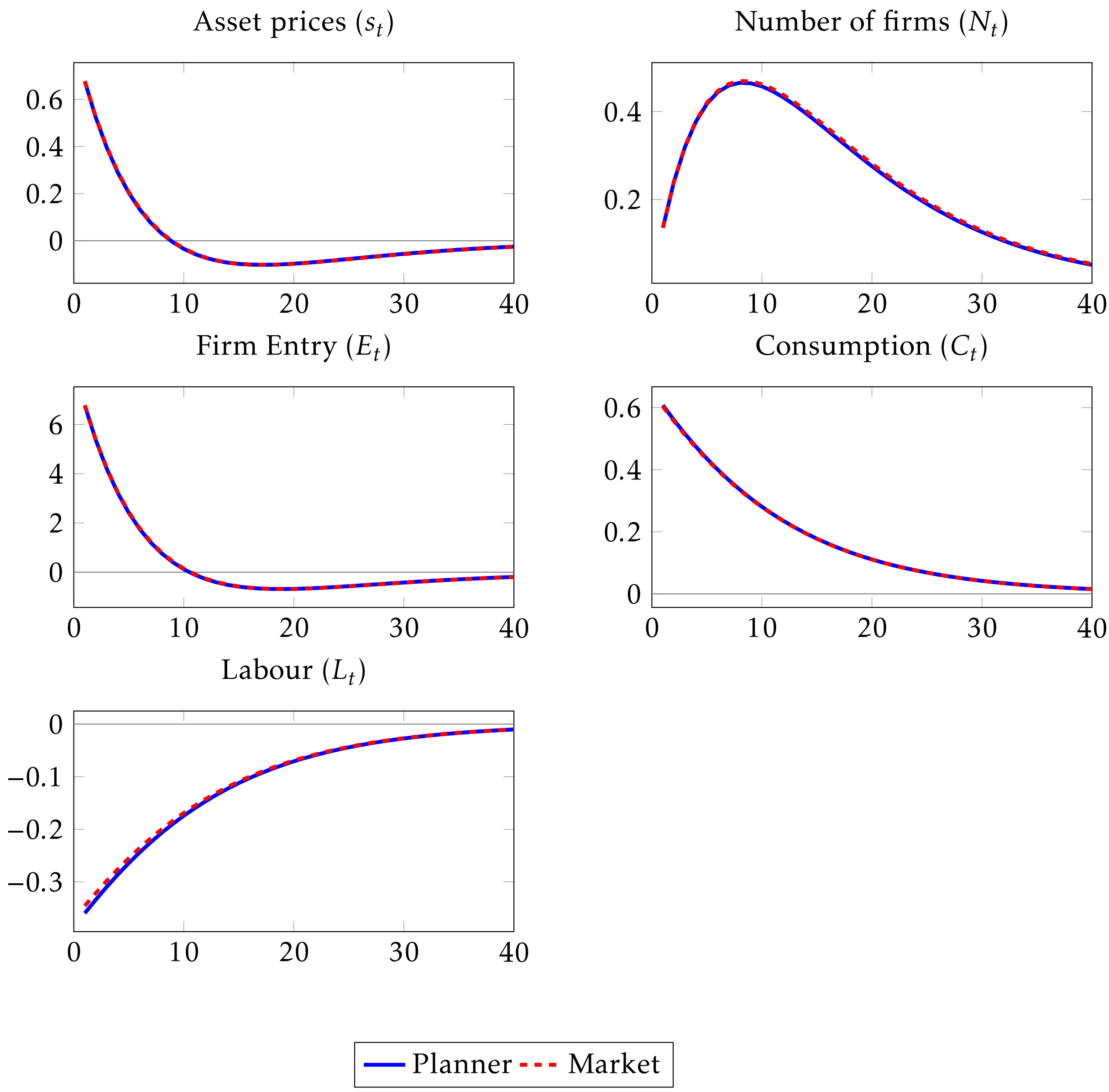


The plot shows response of variables to a technology shock for different values of the adjustment costs ψ for the decentralised model.

G.3 Planner vs. Market

In Figure 7 the impulse response functions (IRFs) for the planner and market outcomes appear equivalent. However, the similarity is by coincidence of our calibration. We show that this is not the case for an alternative calibration below.

Figure 7: Technology shock IRFs for Market and Planner



The plot shows comparison between responses of variables for centralised vs. decentralised model to a technology shock.

G.3.1 Log-Linearised Comparison

The first-order approximation of the equilibrium conditions in Table 2 is almost identical across the decentralised and centralised cases. The difference in the labour market equilibrium conditions disappears in a first-order approximation because the markup is a constant mutliplier. Thus, the only remaining difference is the first term in the square brackets of the Euler condition. For the decentralised economy this is π_{t+1} and for the centralised economy it is $Y_{N,t+1}$. The first-order approximation of these two terms is similar, particularly for the calibration that we implement. We get:

$$\hat{\pi}_t = \frac{\tilde{y}}{\tilde{\pi}}\left(1 - \frac{\nu}{\mu}\right)\hat{y}_t = \frac{1}{\frac{1}{\beta} - (1-\delta)}\left(1 - \frac{\nu}{\mu}\right)\left[y_t^{\mathrm{M}} - \tilde{y}^{\mathrm{M}}\right]$$
$$\hat{Y_N}_t = \frac{\tilde{y}}{\tilde{Y_N}}(\gamma - \nu)\hat{y}_t = \frac{1}{\frac{1}{\beta} - (1-\delta)}(\gamma - \nu)\left[y_t^{\mathrm{P}} - \tilde{y}^{\mathrm{P}}\right]$$

where $\tilde{\pi} = \tilde{Y_N} = \frac{1}{\beta} - (1-\delta)$. Therefore, for a given deviation in output from steady-state output, the difference is the multiplier which is $\left(1 - \frac{\nu}{\mu}\right)$ in the decentralised case, compared to $(\gamma - \nu)$ in the centralised case. For our baseline calibration, $\nu = 0.9$, $\mu = 1.1$, $\gamma = 1.05$ therefore:

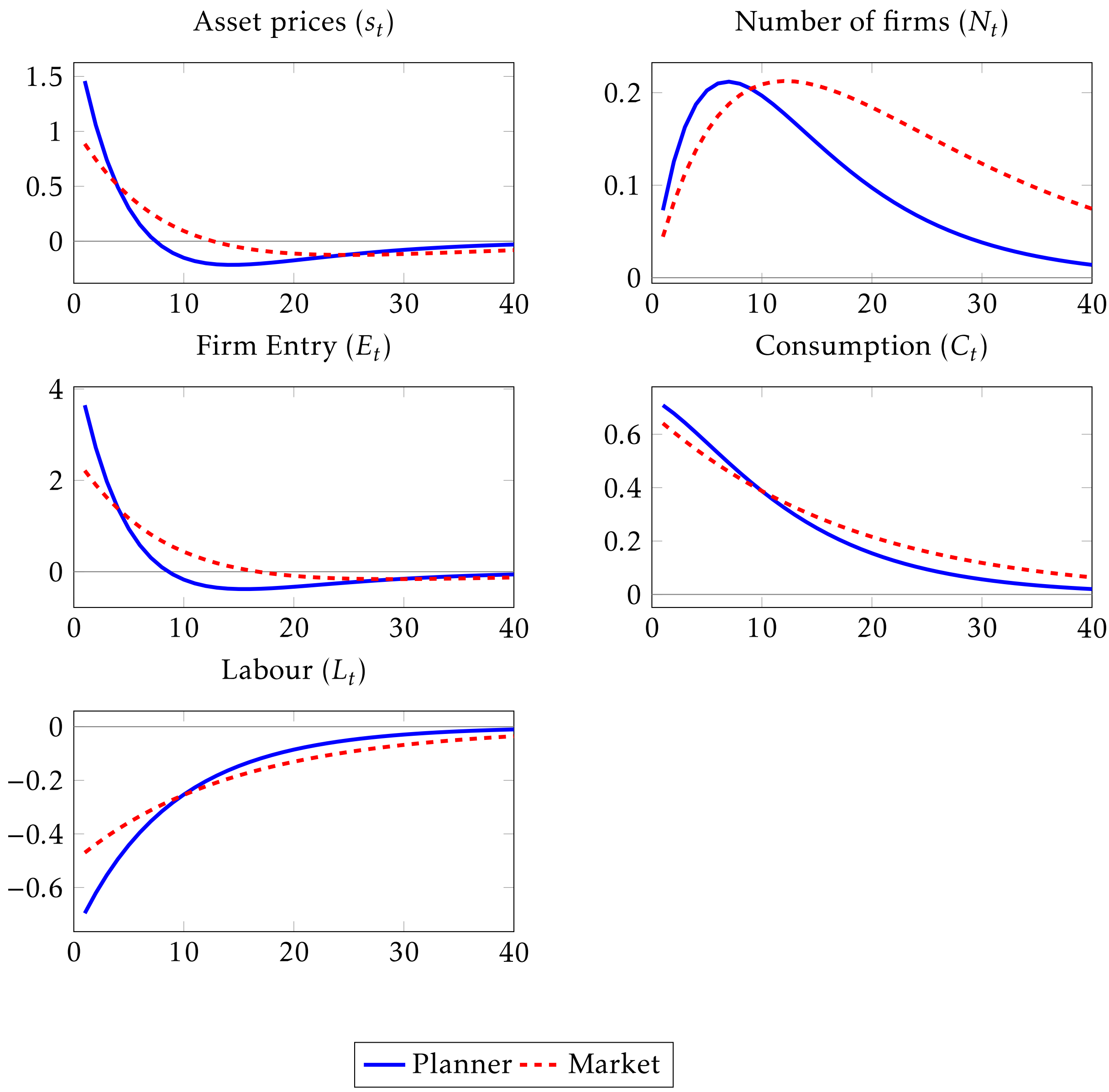
$$1 - \frac{\nu}{\mu} = 0.14$$
$$\gamma - \nu = 0.15$$

Therefore the log-linearized dynamics for $\hat{\pi}_t$ and $\hat{Y}_{N,t+1}$ are similar numerically. The implication is that the IRFs appear similar under this calibration in Figure 7. However, as an illustration of their difference, if we change γ arbitrarily such that the calibration is $\nu = 0.9$, $\mu = 1.1$, $\gamma = 2.05$, then

$$1 - \frac{\nu}{\mu} = 0.14$$
$$\gamma - \nu = 1.15$$

This calibration yields the IRFs in Figure 8.

Figure 8: Technology shock IRFs for Market and Planner (high γ)



The plot shows comparison between responses of variables for centralised vs. decentralised model to a technology shock where $\gamma = 2.05$.

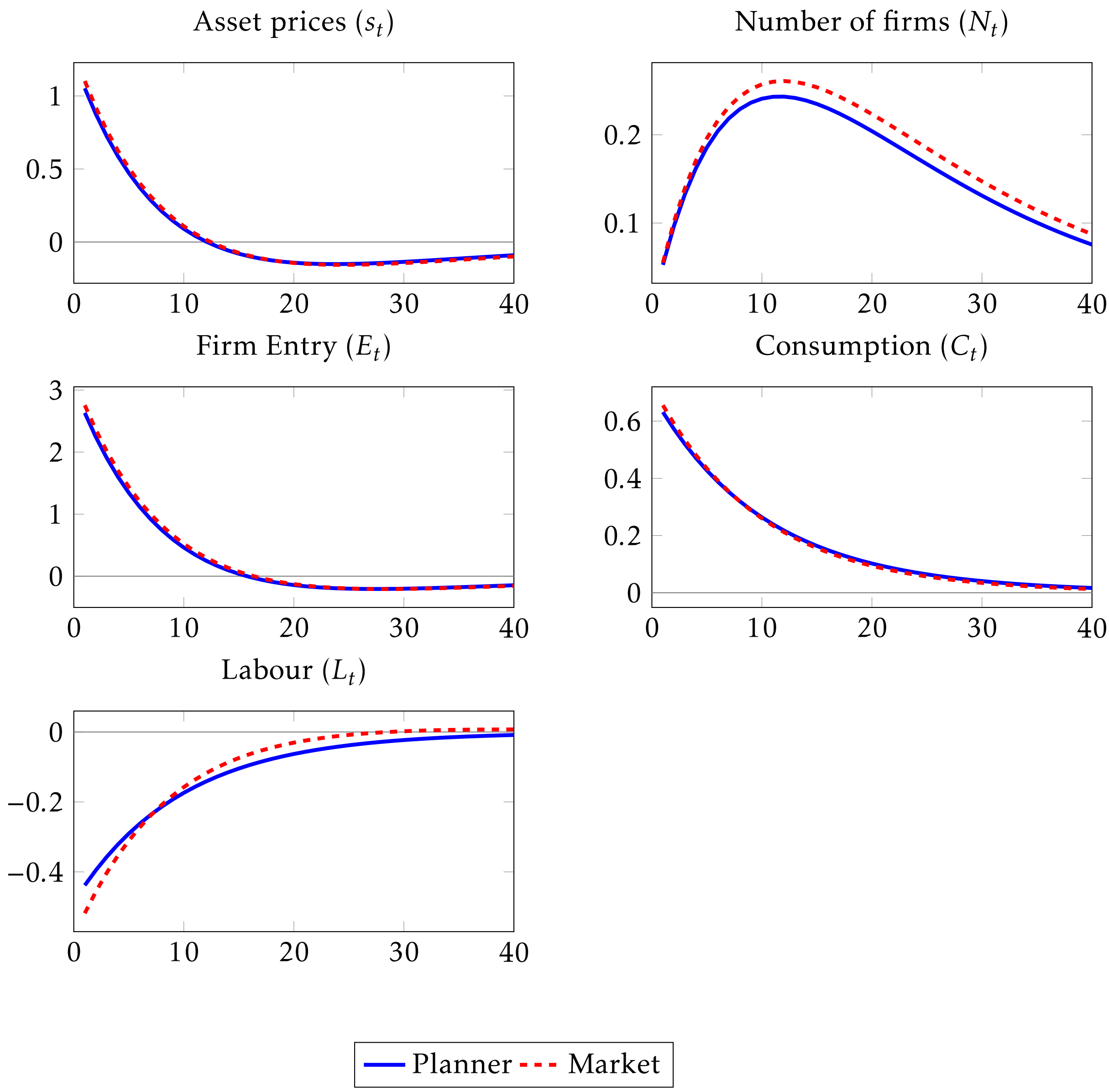
As a further illustration of their difference, if we change μ arbitrarily such that the calibration is $\nu = 0.9$, $\mu = 2.1$, $\gamma = 1.05$, then

$$1 - \frac{\nu}{\mu} = 0.57$$

$$\gamma - \nu = 0.15$$

This calibration yields the IRFs in Figure 9, which also show distinct adjustment paths.

Figure 9: Technology shock IRFs for Market and Planner (high μ)



The plot shows comparison between responses of variables for centralised vs. decentralised model to a technology shock where $\mu = 2.1$.

This shows that the introduction of dynamic firm entry causes markups to have an effect on dynamics at a first-order approximation. Whereas, in a standard RBC model with imperfect competition but no dynamic entry, the markup plays no role to a first-order approximation. Without dynamic entry, the introduction of imperfect competition creates only a static distortion through the markup reducing marginal revenue products in factor market equilibrium.[17] Therefore, for a first-order approximation, the dynamics are the same with perfect and imperfect competition, but the

[17]To maximize welfare, the planner should equate marginal revenue products of capital and labour with their factor prices. The planner can achieve this through a Pigouvian tax which subsidises labour, by the inverse of the markup, such that marginal revenue products are increased and the wage is higher.

paths converge to different steady-state values.

H Optimal Tax

In the presence of distortions, the equilibrium conditions are equivalent if there is a wage tax

$$\chi L_t^{\eta} C_t^{\sigma} = (1-\tau^w)\frac{\nu}{\mu} A_t L_t^{\nu-1} N_t^{\gamma-\nu}$$

such that

$$1-\tau^w = \mu.$$

And a profit (dividend) tax

$$s_t = \beta \mathbb{E}_t \left\{ \left(\frac{C_t}{C_{t+1}}\right)^{\sigma} \left[(1-\tau_{t+1}^D) \underbrace{N_{t+1}^{\gamma-1}\left(\left(1-\frac{\nu}{\mu}\right) A_{t+1} L_{t+1}^{\nu} N_{t+1}^{-\nu} - \phi\right)}_{\pi_{t+1}} - \mathcal{C}_N(E_{t+1}, N_{t+1}) + (1-\delta) s_{t+1} \right] \right\}$$

such that

$$1-\tau_{t+1}^D = \frac{Y_{N,t+1}}{\pi_{t+1}} = \frac{N_{t+1}^{\gamma-1}\left((\gamma-\nu) A_{t+1} N_{t+1}^{-\nu} L_{t+1}^{\nu} - \gamma\phi\right)}{\pi_{t+1}}.$$

Each period the policy maker must trade-off supply-side features with demand-side features to ensure optimality, so profits and the marginal product of a firm coincide. A negative tax (subsidy) $\tau^D < 0$ is required when $Y_N > \pi$. A positive tax $0 < \tau^D < 1$ is required when $Y_N < \pi$.

I Equilibrium Conditions Exogenous Labour Supply

The assumption of exogenous labour supply simplifies the model conditions given in Table 2. Fixing aggregate labour at a normalised level of 1 ($L^S = 1$), we remove L from the system of equations. With no choice of labour supply in the household problem, this eliminates the intratemporal labour market clearing condition, reducing the model to four equations in four unknowns $\{C_t, N_t, E_t, s_t\}$. In the decentralised model, wages are determined by labour demand: $w_t = \frac{\nu}{\mu} A_t N_t^{\gamma-\nu}$. Wages are increasing in the number of firms since $\gamma - \nu > 0$, this reflects that more firms raises labour demand, but labour supply is fixed. More firms increases labour demand through the external returns effect $\gamma > 1$ but reduces demand for higher levels of returns to scale $\nu \uparrow$ because more firms divides aggregate resources up more, meaning each individual firm benefits less from increasing returns. If there are constant external returns and a constant marginal cost curve $\gamma = \mu = 1$, then the number of firms does not determine labour demand and wages are constant. Table 8 compares the equilibrium conditions with exogenous labour.

Table 8: Comparison of Decentralised and Centralised Equilibrium Conditions with Exogenous Labour

Decentralised	Centralised
$N_t^{\gamma}(A_t N_t^{-\nu} - \phi) = C_t + E_t + \mathcal{C}(E_t, N_t)$	$N_t^{\gamma}(A_t N_t^{-\nu} - \phi) = C_t + E_t + \mathcal{C}(E_t, N_t)$
$N_{t+1} = E_t + (1-\delta)N_t$	$N_{t+1} = E_t + (1-\delta)N_t$
$s_t = 1 + \mathcal{C}_E(E_t, N_t)$	$s_t = 1 + \mathcal{C}_E(E_t, N_t)$
$s_t = \beta \mathbb{E}_t \left\{ \left(\frac{C_t}{C_{t+1}} \right)^{\sigma} \left[N_{t+1}^{\gamma-1} \left(\left(1 - \frac{\nu}{\mu}\right) A_{t+1} N_{t+1}^{-\nu} - \phi \right) + X_t \right] \right\}$	$s_t = \beta \mathbb{E}_t \left\{ \left(\frac{C_t}{C_{t+1}} \right)^{\sigma} \left[N_{t+1}^{\gamma-1} \left((\gamma - \nu) A_{t+1} N_{t+1}^{-\nu} - \gamma\phi \right) + X_t \right] \right\}$
$s_t = \beta \mathbb{E}_t \sum_{i=1}^{\infty} (\beta(1-\delta))^{i-1} \left(\frac{C_t}{C_{t+i}} \right)^{\sigma} [\pi_{t+i} - \mathcal{C}_{N,t+i}]$	$s_t = \beta \mathbb{E}_t \sum_{i=1}^{\infty} (\beta(1-\delta))^{i-1} \left(\frac{C_t}{C_{t+i}} \right)^{\sigma} [Y_{N,t+i} - \mathcal{C}_{N,t+i}]$

J Steady state Exogenous Labour

With exogenous labour, we drop the static intratemporal labour market clearing condition, and set $\tilde{L} = 1$. We have also imposed $\tilde{A} = 1$. The updated conditions are in Table 9.

Table 9: Steady-state with Exogenous Labour

Decentralised	Centralised
$\tilde{s} = 1$	$\tilde{s} = 1$
$\tilde{N}^{\gamma-1}\left[\left(1 - \frac{\nu}{\mu}\right)\tilde{N}^{-\nu} - \phi\right] = \tilde{\pi}$	$\tilde{N}^{\gamma-1}\left[(\gamma - \nu)\tilde{N}^{-\nu} - \gamma\phi\right] = \tilde{\pi}$
$\tilde{C} = \tilde{N}^{\gamma}(\tilde{N}^{-\nu} - \phi) - \delta\tilde{N}$	$\tilde{C} = \tilde{N}^{\gamma}(\tilde{N}^{-\nu} - \phi) - \delta\tilde{N}$
$\tilde{E} = \delta\tilde{N}$	$\tilde{E} = \delta\tilde{N}$

We discuss the different cases below

J.0.1 General Case

With $\mu > 1$ and $\gamma > 1$, the decentralised number of firms can be greater than, less than, or equal to the number of firms in the planner equilibrium. The market solves

$$\tilde{N}^{M,\gamma-1}\left[\left(1 - \frac{\nu}{\mu}\right)\tilde{N}^{M,-\nu} - \phi\right] = \tilde{\pi}$$

We can further understand the solution to the market outcome by taking the implicit derivative with respect to μ.

$$\frac{d\ln\tilde{N}}{d\ln\mu} = \left[-\frac{\mu}{\nu}(\gamma - 1)\tilde{N}^{1-(\gamma-\nu)}\tilde{\pi} + \mu\left(1 - \frac{\nu}{\mu}\right)\right]^{-1} \gtreqless 0$$

Depending on the size of $\gamma - 1$ the term in square brackets can be positive or negative. In the case of constant external returns $\gamma = 1$, the derivative is strictly positive.

The planner solves

$$\tilde{N}^{P,\gamma-1}\gamma\left[\left(1-\frac{\nu}{\gamma}\right)\tilde{N}^{P,-\nu}-\phi\right]=\tilde{\pi}$$

Special Case: CES

Special case of $\gamma > 1, \mu > 1$ is $\mu = \gamma$ where we have the common CES aggregator case:

$$\tilde{N}^{M,\gamma-1}\left[\left(1-\frac{\nu}{\gamma}\right)\tilde{N}^{M,-\nu}-\phi\right]=\frac{1}{\beta}-(1-\delta)$$

And the planner relationship does not change as there is no μ

$$\tilde{N}^{P,\gamma-1}\left[\left(1-\frac{\nu}{\gamma}\right)\tilde{N}^{P,-\nu}-\phi\right]=\frac{1}{\gamma}\left(\frac{1}{\beta}-(1-\delta)\right)$$

Therefore

$$\tilde{N}^{P,\gamma-1}\left[\left(1-\frac{\nu}{\gamma}\right)\tilde{N}^{P,-\nu}-\phi\right]=\frac{1}{\gamma}\tilde{N}^{M,\gamma-1}\left[\left(1-\frac{\nu}{\gamma}\right)\tilde{N}^{M,-\nu}-\phi\right]$$

The decentralised operating profits are less than the marginal product of a firm. Consequently, the decentralised equilibrium yields a lower number of firms that the planner equilibrium.

$$\tilde{N}^M < \tilde{N}^P$$

This is the implicit love-of-variety that exists in a CES aggregator.

J.0.2 No Markups Constant External Returns

With $\mu = 1$ and $\gamma = 1$ the centralised and decentralised outcomes coincide. Operating profits are equal to the marginal product of a firm.

$$\tilde{N}^{\mathrm{M}}\Big|^{\gamma=1,\mu=1}=\tilde{N}^{\mathrm{P}}\Big|^{\gamma=1,\mu=1}=\left(\frac{1-\nu}{\frac{1}{\beta}-(1-\delta)+\phi}\right)^{\frac{1}{\nu}}, \qquad \nu\in(0,1)$$

Therefore firm size is at minimum efficient scale:

$$\tilde{y}^{\mathrm{M}}\Big|^{\gamma=1,\mu=1}=\tilde{y}^{\mathrm{P}}\Big|^{\gamma=1,\mu=1}=\frac{1}{1-\nu}\left(\frac{1}{\beta}-(1-\delta)+\phi\nu\right), \qquad \nu\in(0,1)$$

Since there is a fixed cost, firms must have an upward sloping marginal cost $\nu \in (0,1)$ for existence. Firms operate at their MES.

J.0.3 Positive Markups Constant External Returns

With $\mu > 1$ and $\gamma = 1$ operating profits in the decentralised equilibrium always exceed the marginal product of a firm. Consequently, the market delivers more firms than the planner, and these firms are smaller in size than the planner solution. The planner solution is unaffected by μ and therefore remains the same as in the $\gamma = \mu = 1$ case, with the same minimum efficient scale.

$$\tilde{N}^{\mathrm{M}}\Big|^{\gamma=1} = \left(\left(1 - \frac{\nu}{\mu}\right)^{-1} (\tilde{\pi} + \phi)\right)^{-\frac{1}{\nu}}, \qquad \nu \in (0, \mu)$$

The market solution for the steady-state number of firms is monotonically increasing in the markup, with elasticity given by: $\frac{d \ln \tilde{N}^{\mathrm{M}}|^{\gamma=1}}{d \ln \mu} = \frac{1}{\mu - \nu} > 0$. Therefore, given $\mu \in (1, \infty)$ there is excess entry in the market outcomes

$$\tilde{N}^{\mathrm{M}}\Big|^{\gamma=1} > \tilde{N}^{\mathrm{P}}\Big|^{\gamma=1}$$

This is a global result. The intuition is that there are no gains from aggregation since $\gamma = 1$, so it is optimal to have all firms at their minimum efficient scale.

J.0.4 No Markups Increasing External Returns

With $\mu = 1$ and $\gamma > 1$ operating profits in the decentralised equilibrium are always less than the marginal product of a firm. Consequently, there is insufficient entry in the decentralized equilibrium relative to the planner equilibrium. The key implication is that eradicating markups in the presence of external returns to scale is not optimal. It causes too few firms to be created. This is analagous to a love of variety argument.[18]

The planner condition is unaffected

$$\tilde{N}^{P,\gamma-1}\gamma\left[\left(1 - \frac{\nu}{\gamma}\right)\tilde{N}^{P,-\nu} - \phi\right] = \frac{1}{\beta} - (1 - \delta)$$

whereas the market condition is

$$\tilde{N}^{M,\gamma-1}\left[(1 - \nu)\tilde{N}^{M,-\nu} - \phi\right] = \frac{1}{\beta} - (1 - \delta)$$

[18] This is the case of perfect competition $\mu = 1$. In this case, the labour market condition is the same in the planner and the decentralised problem. The planner already acts as if there is perfect competition since they are not constrained by a demand curve. In the decentralized model under perfect competition, there is an infinitely elastic (horizontal) demand curve and firms produce at their minimum efficient scale, where U-shaped average cost curves are minimized and intersect with price and marginal cost. There are U-shaped average cost curves in the decentralized model if $\nu < 1, \phi > 0$. The planner prefers to intervene to make firms smaller than their efficient scale, which would generate negative profits for each individual firm in their micro market, but the planner can redistribute the extra surplus from scale effects gained in the aggregate such that everyone is better off.

where $\nu \in (0, 1)$.