UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Paper Economics

Cambridge Working Paper Economics: 1680

Oligopolistic competition and welfare

Robert A.
Ritz

19 December 2016

This chapter provides a selective survey of recent developments in the study of social welfare under oligopoly. The main topics covered are (i) the rate of cost pass through as a tool to analyze market performance; (ii) the quantification of welfare losses due to market power in Cournot-style models; and (iii) new results from models with endogenous entry. The chapter highlights common themes across these topics and identi.es areas for future research.

Oligopolistic competition and welfare

Robert A. Ritz*

Judge Business School & Faculty of Economics

University of Cambridge, U.K.

E-mail: rar36@cam.ac.uk

This version: December 2016

Prepared for the *Handbook of Game Theory & Industrial Organization*, Luis Corchón & Marco Marini (*eds.*), Edward Elgar, forthcoming.

Abstract

This chapter provides a selective survey of recent developments in the study of social welfare under oligopoly. The main topics covered are (i) the rate of cost pass-through as a tool to analyze market performance; (ii) the quantification of welfare losses due to market power in Cournot-style models; and (iii) new results from models with endogenous entry. The chapter highlights common themes across these topics and identifies areas for future research.

Keywords: Oligopoly, market performance, social welfare, division of surplus, cost pass-through, endogenous entry

JEL classifications: D42 (monopoly), D43 (oligopoly), D61 (allocative efficiency), L20 (firm behavior) L40 (antitrust policy)

*I am grateful to Anette Boom, Simon Cowan, Federico Etro, Pär Holmberg, Nathan Miller, Michael Pollitt, Andrew Rhodes, John Vickers and Glen Weyl for helpful comments and suggestions. Any remaining errors are mine.

1 Introduction

1.1 Objectives

Market performance under imperfect competition has been a classic question for economists since the time of Adam Smith. It remains a central concern of the theoretical and empirical industrial-organization (IO) literature and anti-trust policymakers dealing with competition issues in practice.

The objective of this chapter is to survey recent developments in the IO theory literature that speak to oligopoly and welfare. The coverage here is explicitly selective, concentrating on areas where the literature has substantially progressed over the last 5–10 years. Related issues have been covered extensively by several authors in the past. Valuable resources remain the early survey chapter by Shapiro (1989) as well as the oligopoly-theory books by Tirole (1988) and Vives (1999) which also contain significant material on welfare.

The uniqueness of this chapter lies in the following. First, the focus is specifically on welfare; most other treatments deal with this only as a by-product. Second, it covers recent developments which have not yet found their way into textbook treatments—but hopefully will do so in the near future. Third, it discusses separate strands of the recent literature in a way that highlights their common themes.

1.2 Scope

The scope of this chapter is limited to relatively simple static oligopoly models under partial-equilibrium analysis.[1] It concentrates on theory—albeit in a way that it is informed by the empirical literature and speaks to industrial applications. Extensions to more complex settings are dealt with by other chapters contained in this volume.

Market power lies solely with firms while buyers are atomistic; there is no price discrimination. The focus is on markets with varying degrees of competitive conduct—rather than tacit collusion or price fixing. Firms are assumed to be risk-neutral profit-maximizers and are equally well informed about the market. There are no other market failures (such as environmental externalities) and no explicit role for regulation (such as price caps) or

[1] This excludes any general-equilibrium effects which, for example, could arise due to interactions between supplier market power and imperfections in input markets (such as the labour market).

other policy interventions.

The definition of "welfare" (W) is mostly taken to be social surplus, that is, the unweighted sum of aggregate consumer surplus (CS) and aggregate producer surplus (Π): $W = CS + \Pi$. A consumer-welfare standard is highlighted in some places given that recent antitrust policy in jurisdictions such as the US and EU is said to be geared more heavily towards consumers.[2]

The results discussed cover a range of models with homogeneous products as well as different forms of horizontal product differentiation. Some of the homogeneous-products results apply equally to settings with vertical differentiation in which there are (known) differences in product quality across firms. Many of the models are "aggregative games" in which a firm's competitive environment can be captured using a single summary statistic of rivals' actions.

These models have useful application across a wide array of industries. In the energy sector, similar homogeneous-product models are widely employed in the analysis of electricity, natural gas and crude oil markets—as well as energy-intensive industry such as cement and steel. The differentiated-price models covered form the basis for competition policy in sectors with branded products.

1.3 Plan for the chapter

Section 2 presents the recent literature on the rate of cost pass-through as an economic tool to understand the market performance and the division of surplus between buyers and sellers. Section 3 discusses recent papers which quantify market performance in various Cournot-style models using welfare losses, that is, the comparison between equilibrium welfare and first-best. Section 4 covers recent developments in the theory of oligopoly with endogenous entry of firms, with a focus on the quantification of welfare losses and the impact of firm heterogeneity. Section 5 gives concluding remarks and suggestions for future research.

[2] See Farrell and Katz (2006) for a discussion of welfare standards in antitrust. Armstrong and Vickers (2010) study a model in which a consumer-welfare standard can, for strategic reasons, be optimal even if the regulator cares about total welfare (because the standard affects the set of mergers that is proposed by firms).

2 Cost pass-through and the division of surplus

Consider the treatment of monopoly in a textbook on microeconomics or industrial organization. With linear demand and costs, the monopolist captures 50% of the (potential) gains from trade, with 25% as consumer surplus—and the remainder as deadweight loss. So there is a ratio of 1:2 between consumer surplus and producer surplus.

Elsewhere, the textbook may turn to the question of cost pass-through: how much of a unit tax is passed onto the market price? For a linear monopoly, the rate of pass-through ($\Delta P/\Delta MC$) equals 50%. So there is a ratio of 1:2 between the price change and the cost increase.

What textbooks do *not* say is that this is no coincidence. The ratio of consumer to producer surplus, in equilibrium, is equal to the rate of cost pass-through in that market. Weyl & Fabinger (2013) develop this insight more broadly, including for various representations of oligopoly, and argue that pass-through is a versatile tool to think about market performance.

Much earlier, Bulow & Pfleiderer (1983) noted how monopoly cost pass-through varies with the shape of the demand curve, i.e., its curvature. Kimmel (1992) exploits this link to frame the profit-impact of a unit tax in a Cournot oligopoly in terms of pass-through. Anderson & Renault (2003) study the relationship between demand curvature and the division of surplus under Cournot competition but do not explicitly cover pass-through. Weyl & Fabinger (2013) tie together these various antecedents.

2.1 Monopoly case

Consider a monopolist which produces a single good with marginal cost $c+t$, where $t \geq 0$ is a parameter. The monopolist faces inverse demand $p(Q)$; let $D(p)$ be the corresponding direct demand. At the optimum, marginal revenue equals marginal cost, $MR(Q) = c+t$.

What is the impact of a small increase in t? Let $\kappa \equiv dp/dt$ denote the rate of cost pass-through, which measures how price responds to a \$1 increase in marginal cost.[3] Denote consumer surplus $CS = \int_p^\infty D(x)dx$, and observe that $dCS/dt = -\kappa Q$, at the optimum. Similarly, by the envelope theorem, the profit impact $d\Pi/dt = -Q$, since the indirect

[3] Another formulation, more frequently used in the international trade literature, instead concerns the pass-through elasticity $(dp/p)/(dt/t) \leq \kappa$, which also incorporates the profit margin.

impact of the tax is zero since the monopolist is optimizing. Hence the burden of an infinitesimal tax, starting at zero, is split according to

$$\left.\frac{dCS/dt}{d\Pi/dt}\right|_{t=0} = \kappa(0),$$

where $\kappa(0)$ is pass-through at the price corresponding to initial zero tax rate.

Consider now a discrete increase in the tax from t_0 to $t_1 > t_0$. Write $Q(t)$ for the optimal quantity as a function of the tax. The changes in consumer surplus and monopoly profits satisfy $\Delta CS_{t_0}^{t_1} = -\int_{t_0}^{t_1} \kappa(t)Q(t)dt$ and $\Delta\Pi_{t_0}^{t_1} = -\int_{t_0}^{t_1} Q(t)dt$. Define the quantity-weighted pass-through over the interval $[t_0, t_1]$ as $\overline{\kappa}_{t_0}^{t_1} \equiv [\int_{t_0}^{t_1} \kappa(t)Q(t)dt]/\int_{t_0}^{t_1} Q(t)dt$. Define $\overline{t}$ as the hypothetical tax rate at which the market is eliminated, that is, $Q(\overline{t}) = 0$, and call the average quantity-weighted pass-through rate $\overline{\kappa} \equiv \overline{\kappa}_0^{\overline{t}}$.[4] Hence the surplus generated from the market's "birth" (at $\overline{t}$) to the equilibrium status quo (at $t = 0$) satisfies

$$\frac{\Delta CS_0^{\overline{t}}}{\Delta\Pi_0^{\overline{t}}} = \frac{\int_0^{\overline{t}} \kappa(t)Q(t)dt}{\int_0^{\overline{t}} Q(t)dt} = \overline{\kappa} = \frac{CS}{\Pi}.$$

Consumer surplus is generated by the market at a rate of monopoly profits times the pass-through rate, weighted over the inframarginal market quantities traded over the interval $[0, \overline{t}]$. This takes into account that the pass-through rate may not be a constant.

Intuitively, high pass-through means that price closely tracks marginal cost, so that (i) the monopolist's degree of market power is "low", and, conversely, (ii) realized social surplus is "high" and largely goes to consumers. With low pass-through, price follows more closely consumers' willingness-to-pay so the monopolist captures the bulk of the gains from trade.

Bulow and Pfleiderer (1983) showed that monopoly pass-through satisfies:

$$\kappa(t) = \frac{1}{[2 - \xi(t)]} = \left.\frac{\text{slope of inverse demand } p(Q)}{\text{slope of marginal revenue } MR(Q)}\right|_{Q=Q(t)},$$

where $\xi(t) \equiv -[p''(Q)/Qp'(Q)]_{Q=Q(t)}$ is the elasticity of the slope of inverse demand, which is a measure of demand curvature. The common theory assumption (Bergstrom and Bagnoli 2005) that direct demand $D(p)$ is log-concave (i.e., $\log D(p)$ is concave in

[4] Some demand curves have $\overline{t} = \infty$, though a finite choke price can be assumed.

p), corresponds to $\xi \leq 1$, and hence to pass-through (weakly) less than 100%. Loosely put, the monopolist then captures a greater share of the gains from trade than consumers. For very concave demand, $\xi \ll 0$, the "triangle" left as consumer surplus is very small; correspondingly the ratio CS/Π and pass-through κ are both small—as is the remaining deadweight loss.

For many familiar demand curves, the ratio $p'(Q)/MR'(Q)$ is constant, so pass-through is a constant with $\kappa(t) = \kappa$ for all $t \in [0, \bar{t}]$—and so the "local" properties of demand are also "global". With linear demand, marginal revenue is everywhere twice as steep as demand, so pass-through $\kappa = \frac{1}{2}$. Other examples are constant-elasticity demand, for which $\xi = 1 + 1/\eta > 1$ (violating log-concavity) where $\eta \equiv -p(Q)/Qp'(Q) > 0$ is the price elasticity, and exponential demand $D(p) = \exp((\alpha - p)/\beta)$, for which $\xi = 1$ as it is log-linear. In such cases, the marginal impact of a tax is equal to its average impact, $(dCS/dt)/(d\Pi/dt) = \Delta CS_{t_0}^{t_1}/\Delta\Pi_{t_0}^{t_1} = CS/\Pi = \kappa$.

The literature has found different ways of representing "constant" higher-order properties of demand. First, using the concept of ρ-concavity: demand $D(p)$ is ρ-concave if and only if demand curvature $\xi(Q) \leq (1 - \rho)$ (Anderson & Renault 2003). A ρ-linear demand curve thus has constant curvature $\xi = 1 - \rho$, and constant pass-through over its entire domain. Second, the demand curve $D(p)$ can be interpreted as arising from the values v of a distribution $F(v)$ of consumers with unit-demand, so $1 - F(p)$ is the quantity sold at price p. The inverse hazard rate is $h(v) \equiv [1 - F(v)]/f(v)$ where $f(v)$ is the density. The monopolist's first-order condition $(p - c) = h(p)$, so pass-through is constant whenever the inverse hazard takes the linear form $h(v) = \lambda_0 + \lambda_1 v$. Third, Rostek, Weretka & Pycia (2009) show that a distribution has a linear inverse hazard rate if and only if it belongs to the Generalized Pareto Distribution, $F(v) = 1 - \left[1 + \frac{\omega}{\sigma}(v - \mu)\right]^{-1/\omega}$, where (μ, σ, ω) respectively describe its location, scale and shape (with $\lambda_0 = (\sigma - \omega\mu)$ and $\lambda_1 = \omega$).

2.2 Oligopoly models

The preceding insights generalize to certain n-firm oligopoly models. Consider a general reduced-form model of competition in which firm i's profits $\pi_i = (p_i - c)q_i$ and the Lerner index (price-cost margin) with symmetric firms is determined as:

$$\varepsilon_D \frac{(p-c-t)}{p} = \theta,$$

where θ is a "conduct parameter" which measures the intensity of competition, and $\varepsilon_D \equiv -p(Q)/Qp'(Q)$ is the *market-level* price elasticity of demand.[5] The previous monopoly analysis corresponds to joint profit-maximization with $\theta = 1$. This setup nests various widely-used models of symmetric oligopoly, including the two following models:

Homogeneous-product oligopoly. Consider a Cournot model augmented with "conjectural variations": when firm i chooses its output it conjectures that each other firm j will adjust its quantity by $dq_j = [R/(n-1)]dq_i$. So the aggregate responses by all its rivals is given by $d(\sum_{j \neq i} q_j)/dq_i = R$. Cournot-Nash competition corresponds to $R = 0$ while Bertrand competition in effect has $R = -1$ (so the price stays fixed). Conjectural variations can be seen as a reduced-form way of incorporating (unmodelled) dynamic features of the game that firms play (Cabral 1995).

The first-order order condition for firm i has $MR_i = p(Q) + q_i p'(Q)(1+R) = c + t$, where $Q \equiv \sum_{i=1}^{n} q_i$ is industry output. This can be re-arranged to give the symmetric equilibrium (with $q_i = Q/n$):

$$\varepsilon_D \frac{(p-c-t)}{p} = \frac{(1+R)}{n} = \theta.$$

Thus a constant conjectural variation R corresponds to a constant conduct parameter θ.

Differentiated-products price competition. Consider a model of price-setting competition with symmetrically differentiated products. Firm i's demand $q_i(p_i, \mathbf{p}_{-i})$ depends on its own price and those of its $n-1$ rivals. In symmetric equilibrium (with $q_i = q = Q/n$), the corresponding price can be written as $p(q)$, which captures how each price changes in response to a simultaneous change in *all* firms' outputs.

The first-order condition, at symmetric equilibrium, for firm i is given by the inverse-elasticity rule, $(p-c-t)/p = -(q/p)/(\partial q_i/\partial p_i)$. The elasticity of market demand is

[5] A large empirical literature reviewed by Reiss and Wolak (2007) has developed structural econometric techniques for estimating the intensity of competition.

$\varepsilon_D = -(p/q) \sum_{j=1}^{n} (\partial q_i / \partial p_j)$, and so:

$$\varepsilon_D \frac{(p - c - t)}{p} = \frac{\sum_{j=1}^{n} (\partial q_i / \partial p_j)}{\partial q_i / \partial p_i} = 1 + \frac{\sum_{j \neq i} (\partial q_j / \partial p_i)}{\partial q_i / \partial p_i} = 1 - A = \theta,$$

where A is the "diversion ratio" from firm i to the rest of the industry as it raises its price (Shapiro 1996).[6] With a linear demand system, for example, A is constant—and hence the conduct parameter is also constant.

As in the monopoly case, the envelope theorem together with the symmetric demand structure imply that the marginal impact of an increase in the tax rate on consumer surplus is given by $dCS/dt = -\kappa(t)Q(t)$. Weyl and Fabinger (2013) show that the marginal impact on producers is given by $d\Pi/dt = -[1 - \kappa(t)(1 - \theta(t))]Q(t)$, where industry profits $\Pi \equiv \sum_{i=1}^{n} \pi_i$. So the burden of an infinitesimal tax, starting at zero, is split according to:

$$\left. \frac{dCS/dt}{d\Pi/dt} \right|_{t=0} = \frac{\kappa(0)}{[1 - \kappa(0)(1 - \theta(0))]}.$$

This is a clean generalization of the monopoly case, with some intuitive properties. For given pass-through $\kappa(0)$, less competitive conduct (higher θ) skews the division of surplus from consumers to producers. For given conduct θ, higher pass-through favours consumers.

The pass-through rate is here given endogenously by:

$$\kappa(t) = \left. \frac{1}{1 + \theta(t) \left[\varepsilon_{MCS} + \varepsilon_\theta \right]} \right|_{Q = Q(t)},$$

where $\varepsilon_\theta \equiv d \log \theta(q) / d \log q$ is the elasticity of the conduct parameters to changes in output, and $\varepsilon_{MCS} \equiv d \log CS'(Q) / d \log Q$ measures how responsive the *marginal* consumer surplus $CS'(Q) = -p'(Q)Q = [p(Q) - MR(Q)]$ (Bulow & Klemperer 2012) is to changes in aggregate output.[7] The pass-through rate, in general, must capture how both of these metrics may vary as the tax affects equilibrium quantities. For example, if the tax reduces per-firm output ($dq/dt < 0$) and this makes the industry more competitive ($d\theta/dq > 0$), then this will tend to enhance pass-through. Note also that pass-through depends

[6] With the symmetric demand structure, $\frac{\sum_{j \neq i} (\partial q_i / \partial p_j)}{\partial q_i / \partial p_i} = \frac{\sum_{j \neq i} (\partial q_j / \partial p_i)}{\partial q_i / \partial p_i}$.

[7] Note that $d \log CS'(Q) / d \log Q = d \log CS'(Q) / d \log q$ given the symmetric setup.

indirectly on the number of firms, since this will generally enter into $\theta(t)$.[8]

As in the monopoly case, it is possible to go from this local impact to the global division of surplus by appropriately weighting how pass-through rates changes along the demand curve.

With "constant conduct" and "constant curvature", the global division of surplus again follows immediately from its local properties. As noted above, many oligopoly models feature $\theta(t) = \theta$ so that $\varepsilon_\theta = 0$. It is also instructive to write out $\varepsilon_{MCS} = [1 - \xi(t)]$ in terms of demand curvature. (Log-concave demand $\xi < 1$ corresponds to $\varepsilon_{MCS} > 0 \Leftrightarrow CS''(Q) > 0$.) With these modifications, pass-through becomes $\kappa = 1/\left[1 + \theta(1 - \xi)\right]_{Q=Q(t)}$ which nests the well-known Cournot-Nash oligopoly result (Kimmel 1992) when $\theta = 1/n$.

2.3 Discussion

The insight that the division of surplus is pinned down by the rate of pass-through has a number of appealing features. First, it allows pass-through to be seen as a "sufficient statistic" for welfare analysis. Second, pass-through estimates already exist in the literature for many markets—based on studies of taxation, exchange rates, and other cost shifts. Third, it makes it easier to form intuitions about market performance since pass-through rates are often easier to think about than higher-order properties of demand.

Information on pass-through can also be used in the reverse direction. For price competition with differentiated products, Miller, Remer & Sheu (2013) instead emphasize how, *assuming* second-order demand properties (i.e., demand curvatures), the matrix of pass-through rates across products can be used to estimate a matrix of "first-order" cross-price elasticities. The attraction of this is that it sidesteps the problem of full-scale estimation of the demand system—which can be time-consuming or even infeasible.

While it is relatively easy to obtain empirical estimates of pass-through, it is more difficult to ascertain how pass-through itself varies along a demand curve. Yet, strictly speaking, the theory requires the *quantity-weighted* pass-through $\overline{\kappa} \equiv \overline{\kappa}_0^{\bar{t}}$. MacKay, Miller, Remer & Sheu (2014) show how reduced-form regressions of price on cost may not yield reliable estimates of the rate of cost pass-through. Loosely put, such a regression can only

[8] For Bertrand competition (with $\theta \equiv 0$), note that $CS/\Pi = \kappa/(1-\kappa)$ but also $\kappa = 1$, so that $CS/\Pi \to 0$ (since firms make zero profits).

yield consistent estimates in situations where the underlying environment is such that cost pass-through is constant over the range of prices in the data. Empirical implementation of the theory may have to resort to assuming $\kappa(t) = \kappa$ for all (or large parts of) $t \in [0, \bar{t}]$.

The above results are based on strong symmetry assumptions such as identical marginal costs and symmetrically differentiated products. These greatly simplify the welfare analysis but are likely to be violated in *any* oligopoly. Weyl & Fabinger (2013) also develop results from a general model which allows certain types of asymmetries. Other factors, such as the details of market structure, then come into play. Again, it is possible to adjust the definition of pass-through to incorporate these but this means that estimating this "adjusted" pass-through rate becomes increasingly difficult—and begins to merge into estimation of a full-scale market model. The power of pass-through is strongest for monopoly.

Another assumption is that the number of the firms' in the market is fixed, and hence invariant to changes in costs. Ritz (2014b) shows that, with log-convex demand, a higher unit-tax can induce additional entry into a market, and thus ultimately lead to a lower market price. Negative pass-through, also known as "Edgeworth's paradox of taxation", is ruled out in the models covered here. Conversely, a low pass-through rate can induce exit of weaker firms which in turn causes price to jump back up.[9]

3 Quantifying welfare losses in Cournot-style models

Consider a textbook Cournot oligopoly with symmetric firms. How large are welfare losses due to market power? With three firms, they equal $6\frac{2}{3}\%$; in other words, a highly concentrated Cournot triopoly delivers over 93% of the maximum possible welfare.[10] For a duopoly, the loss is 11%—certainly not trivial, but not large either.

A recent literature quantifies market performance directly in terms of realized welfare (Corchón 2008; Ritz 2014). It shows that welfare losses in familiar oligopoly models are often perhaps surprisingly small, and also shows what market factors can generate more

[9] Further afield, in the context of the commercial banking industry, Ritz & Walther (2015) show how risk aversion and informational frictions tend to dampen the pass-through of changes in interest rates across loan and deposit markets.

[10] For a duopoly in which one firm is a Stackelberg leader, the welfare loss also equals $6\frac{2}{3}$ percent—so the social value of leadership is equal to one additional entrant.

substantial losses.

The approach is based on calculating equilibrium welfare losses relative to the first-best benchmark. It turns out that this ratio can naturally be determined in terms of observable metrics, notably firms' market shares. In this way, this literature is potentially useful also for policy purposes as a simple initial screening tool for market performance.[11]

3.1 Cournot-Nash oligopoly

Consider a Cournot-Nash oligopoly with $n \geq 2$ active firms. Firm i has marginal cost c_i and chooses its output q_i to maximize its profits $\pi_i = (p - c_i)q_i$, where the price $p(Q)$ with industry output $Q \equiv \sum_{i=1}^{n} q_i$. Without loss of generality, firms are ordered such that $c_1 \leq c_2 \leq ... \leq c_n$. Inverse demand $p(Q) = \alpha - \beta Q^{1-\xi}$ is $(1-\xi)$-linear with constant curvature ξ, where $\xi < 2$ gives downward-sloping *industry* marginal revenue. This also ensures the uniqueness and stability of the Cournot equilibrium as well as a well-behaved consumer-surplus function.

The first-best outcome, which maximizes social welfare $W \equiv CS + \Pi$, where $\Pi \equiv \sum_{i=1}^{n} \pi_i$, has price equal to the lowest marginal cost $p^{fb} = c_1$ with output $Q^{fb} = p^{-1}(c_1) = [(\alpha - c_1)/\beta]^{1/(1-\xi)}$. Denote the corresponding welfare level as W^{fb}.

The first-order condition for firm i is $MR_i = c_i$, and the sum of first-order conditions $\sum_{i=1}^{n} MR_i \equiv [np(Q) + Qp'(Q)] = \sum_{i=1}^{n} c_i$ pins down the equilibrium industry output Q^*. Hence the equilibrium price is given by:

$$[np^* - (1-\xi)(\alpha - p^*)] = \sum\nolimits_{i=1}^{n} c_i \Longrightarrow p^* = \frac{(1-\xi)\alpha + n\bar{c}}{(n+1-\xi)}.$$

This equilibrium pricing function $p^*(\bar{c})$ is affine in the unweighted-average unit cost $\bar{c} \equiv \frac{1}{n}\sum_{i=1}^{n}$, so the pass-through of a cost change that affects all firms equally $\kappa \equiv dp^*/d\bar{c} = n/(n+1-\xi)$ is constant (i.e., $d^2p^*/d\bar{c}^2 = 0$).

Denote equilibrium welfare and consumer surplus under Cournot competition as W^*

[11] An older empirical literature going back to Harberger (1954) estimates welfare losses normalized relative to sales revenue. A disadvantage is that magnitudes are hard to intepret; for example, the ratio of equilbrium welfare to revenue can vary widely for reasons that have nothing to do with market power.

and CS^*, and define welfare losses relative to first-best as:

$$L \equiv \left(1 - \frac{W^*}{W^{fb}}\right),$$

which is a unit-free measure of welfare that lies on the unit interval, $L \in [0, 1]$.

3.1.1 Symmetric firms

To build intuition, it is useful to begin with the benchmark case in which firms have identical marginal costs, $c_i = \overline{c}$ for all i; Anderson and Renault (2003) showed that:

$$L(n, \xi) = 1 - \frac{n^{1/(1-\xi)}(n + 2 - \xi)}{(n + 1 - \xi)^{(2-\xi)/(1-\xi)}}.$$

Equilibrium welfare losses depend only on the number of (symmetric) firms and the curvature of demand. As expected, they decline with the number of firms and tend to zero in the limit as the competitors grows large. This reflects the classic result on convergence to perfect competition in large markets.

Welfare losses also tend to zero if the curvature of demand is extreme, either as $\xi \to 2$ or as $\xi \to -\infty$. The case with $\xi \to 2$ corresponds to very convex demand in which the total revenue to firms (and hence the total expenditure by consumers) become constant—and thus invariant to the number of firms competing; since production costs are symmetric, there is no other source of welfare losses. The case with $\xi \to -\infty$ corresponds to demand which becomes rectangular (infinitely concave) so all consumers have identical WTP of α for the good; then firms extract all the gains from trade with a uniform price $p^* = \alpha$ while serving all consumers efficiently. Consistent with the previous monopoly discussion, this is the limit of zero pass-through, $\kappa \to 0$.

More generally, Corchón (2008) shows that welfare losses with symmetric firms tend to be quite "small". For example, with linear demand ($\xi = 0$) the above simplifies to $L(n) = 1/(n+1)^2$. So welfare losses are of order $1/n^2$ (as the price and output inefficiencies are both of order $1/n$) and in a quantitative sense decline quickly as the number of competitors rises, e.g., $L(n) \leq 4\%$ if $n \geq 4$. For non-linear demands, Corchón (2008) derives the *maximal* welfare loss for a given number of firms, that is, $\widehat{L}(n) \equiv \max_{\xi} L(n, \xi)$. As long as there are at least four firms in the market, overall welfare losses are never greater than around 5.8%. In fact, the textbook case with linear demand generally yields fairly *high* welfare losses.

3.1.2 Asymmetric firms

The symmetric case shows that welfare losses due to market power do not tend to be "large"—say well above 5%—in Cournot-Nash models, except in some duopoly cases. However the symmetry assumption switches off any role for welfare losses due to productive inefficiency. Indeed, it is well-known that Cournot equilibria are not cost-efficient since the lowest-cost firm does produce all output; high-cost firms serve too much of the market (Lahiri & Ono 1988; Farrell & Shapiro 1990; Aiginger & Pfaffermayr 1997).[12]

More realistic results revert back to the case where firms' marginal costs may be asymmetric. The challenge that costs are typically difficult to observe (or even reliably estimate), while there is an advantage in having a welfare measure that depends on observables as far as possible. The trick to resolve this is to use the first-order conditions to "substitute out" costs for market shares which are readily available for many markets.

In particular, let firm i's equilibrium market share $s_i^* \equiv (q_i^*/Q^*)$, and recall the first-order condition $MR_i = p(Q) + q_i p'(Q) = p - (1-\xi)\beta Q^{2-\xi} s_i = c_i$. Some rearranging shows that its equilibrium market share satisfies:

$$(1-\xi)(\alpha - p^*)s_i^* = (p^* - c_i),$$

which provides a direct mapping between (observable) market share and marginal cost, for a given p^* as determined above. Note that firm 1's market share s_1^* is the highest since it has the lowest marginal cost, and $s_1^* \geq s_2^* \geq ... \geq s_n^*$.

Based on this, Corchón (2008) shows that welfare losses with asymmetric firms are given by:

$$L(s_1^*, H^*, \xi) = 1 - \frac{[1 + (2-\xi)H^*]}{[1 + (1-\xi)s_1^*]^{(2-\xi)/(1-\xi)}},$$

where $H^* \equiv \left(\sum_{i=1}^n s_i^2\right)_{\{s_i^*\}_{i=1}^n}$ is the Herfindahl index of concentration, evaluated at the

[12] More generally, marginal costs are not equalized acrosss firms (as would occure in a cost-minimizing allocation of any given industry production level).

equilibrium market shares.[13] The expression for welfare losses remains simple: they now depend on s_1^* and the Herfindahl index H^*, both of which previously boiled down to the number of firms in the symmetric case.[14] Intuitively, market performance under Cournot is described by the Herfindahl index while the largest market share captures how close this performance is to first-best—for which it should equal 100%.

Welfare losses increase with the market share of the largest firm s_1^* (holding fixed the value of the Herfindahl index). Intuitively, the largest firm must have an above-average market share; further increasing its size relative to the market pushes the equilibrium closer to monopoly.

Welfare losses decline in the Herfindahl index (holding fixed s_1^*). While perhaps initially counterintuitive, the reason for the results is that a higher industry concentration shifts market share toward the more efficient firms (which have lower costs). This mitigates the productive inefficiency of the Cournot equilibrium. The more general point is that the Herfindahl index is not a reliable guide to market performance.

Corchón (2008) shows that welfare losses in asymmetric Cournot models can be *very* large. Specifically, it is possible to find combinations of demand conditions (ξ) and market structure $(s_1^*, s_2^*, ..., s_n^*)$ which yields welfare losses which are arbitrarily close to unity, $L(s_1^*, H^*, \xi) = 1 - \epsilon$ for a small constant $\epsilon \to 0$. At the same time, the Herfindahl index may be arbitrarily low. The worst case for welfare is when the non-largest firms are symmetric, $s_2^* = s_3^* = ... = s_n^*$; then $\lim_{\xi \to -\infty} [\lim_{n \to \infty} L(s_1^*, n, \xi)] = 1 - s_1^*$, and clearly $H^* \approx 0$ while $L \approx 1$ for s_1^* small.

Welfare losses can be substantially higher than in symmetric cases, even with non-extreme assumptions about demand curvature and realistic market structures. As a numerical example, let firms' market shares $s_1^* = 40\%$, $s_2^* = 30\%$, $s_3^* = 20\%$, and $s_4^* = 20\%$ which implies a Herfindahl index $H^* = 0.3$. Assuming linear demand ($\xi = 0$), it follows that welfare losses $L(s_1^*, H^*, \xi) \approx 18\%$. This is approximately *three* times as high as the *maximal* loss with four symmetric firms.

Surprisingly, it is possible for market performance under Cournot to be *worse* than

[13] This expression simplifies to the symmetric case where $H^* = s_1^* = n^{-1}$ for all i.

[14] Again, the demand parameters (α, β) do not play any role: the influence of α is subsumed in firms' market shares and β is merely a scale factor which does not affect *relative* welfare losses. (All else equal, doubling β halves both W^* and W^{fb} so their ratio is unchanged.)

for a monopoly. Corchón (2008) shows that with log-convex demand ($\xi > 1$), monopoly indeed generates the highest welfare loss. However, with log-concavity ($\xi < 1$), the socially worst outcome involves a "high" market share (at least 50%) for one firm combined with a "tail" of very small firms. The intuition is that the small firms add little to competition but substantially reduce productive efficiency.

Finally, with asymmetric firms, market performance is no longer obviously related to cost pass-through. Pass-through $\kappa(n, \xi)$ reflects the number of competitors and demand conditions while welfare losses $L(s_1^*, H^*, \xi)$ also depend on the details of the distribution of firms' market shares. Market performance can vary widely even for a fixed underlying rate of pass-through.

3.2 Endogenous competitive conduct in two-stage games

A significant body of empirical evidence shows that many industrial markets have a competitive intensity that is tougher than Cournot-Nash but falls short of perfect competition (Bresnahan 1989). One way to model this, as in Section 2, is by adding an *exogenous* conduct parameter. Similarly, a widely-used class of two-stage strategic games comes with an conduct parameter that is endogenously determined by the interaction of the two stages. It turns out that welfare losses in such models can be *much* lower than in the standard Cournot setup.

Consider the two-stage game introduced by Vickers (1985) and Fershtman & Judd (1987). Each firm delegates decision-making in the product market to a manager. Manager i receives an incentive contract which induces maximization of an objective function $\Omega_i = (1 - \varphi_i)\pi_i + \varphi_i R_i$, where $R_i \equiv pq_i$ is the firm's sales revenue. In the first stage, each firm's shareholders choose the incentive weight φ_i to maximize their firm's profits π_i. In the second stage, each firm's manager chooses an output level q_i to maximize the objective Ω_i.

This setup reflects extensive evidence that managers across a wide range of industries appear to place significant emphasis on measures of their firm's size (Ritz 2008, 2014a). This is particularly evident in competition for rankings in "league tables" which are based on firms' sales or market share, not profits—and play a prominent role, for example, in commercial and investment banking as well as in car and aircraft manufacturing.[15]

Firms can use their Stage-1 choice of the incentive contract as a commitment device to gain strategic advantage in the product market.[16] Higher values of φ_i constitute aggressive output-increasing behaviour since they correspond to placing less weight on costs. Aggressive behaviour is optimal when firms are competing in strategic substitutes since it induces a soft response from rivals. From the firms' viewpoint, this leads to a prisoners' dilemma: each firm individually has an incentive to engage in aggressive behaviour but this ends up making them collectively worse off.

Remark. The exposition here focuses on a widely-used two-stage model of delegation. Yet the same welfare-conclusions apply to a range of other two-stage models which are strategically equivalent. This includes the seminal model of Allaz & Vila (1993) in which firms engage in forward trading of their production, hiring "overconfident" managers who overestimate the state of market demand, and models of strategic trade policy in which countries use output subsidies to commit to their firms to aggressive behaviour. (See Ritz (2008, 2014a) for further discussion.)

The game is solved backward for the subgame-perfect Nash equilibrium. Manager i's first-order condition in Stage 2 is given by:

$$\frac{\partial \Omega_i}{\partial q_i} = (1-\varphi_i)\frac{\partial \pi_i}{\partial q_i} + \varphi_i \frac{\partial R_i}{\partial q_i} = \left[p(Q) + p'(Q)q_i - (1-\varphi_i)c_i\right] = 0.$$

This implicitly defines manager i's best response in the product market. Let $q_i^*(\varphi_1, \varphi_2, ..., \varphi_n)$ denote the Nash-equilibrium output choice, as a function of all firms' incentive contracts. Given this, in Stage 1, each firm's shareholders choose their manager's incentive weight according to:

$$\frac{d\pi_i}{d\theta_i} = \left[p(Q^*) + p'(Q^*)q_i^*(1+\upsilon_{-i}) - c_i\right]\frac{dq_i^*}{d\varphi_i} = 0,$$

where $\upsilon_{-i} \equiv (dQ_{-i}/dq_i)_{\{q_i^*\}_{i=1}^n} < 0$ is the aggregate response of rivals' $Q_{-i} \equiv \sum_{j \neq i} q_j$ and $dq_i^*/d\varphi_i > 0$. Combining the two first-conditions, the contract places places positive

[15] There is also a large body of evidence which shows that executive compensation in manufacturing, service and financial industries often rewards measures of firm size in addition to profits.

[16] It is assumed that such commitment is credible; a sufficient condition for this is that managers' contracts are observable and cannot be renegotiated.

weight on sales revenue $\varphi_i^* > 0$ if and only if $\upsilon_{-i}^* < 0$. This corresponds to a conduct parameter for firm i's product-market behaviour; the only difference is that υ_{-i}^* is here determined endogenously in Stage 1.[17]

Ritz (2014a) shows that, with linear demand, $\upsilon_{-i}^* = -(n-1)/n < 0$ for all i, and equilibrium welfare losses are given by:

$$\widetilde{L}(n, s_1^*, H^*) = 1 - \frac{n(n+2H^*)}{(n+s_1^*)^2}.$$

The market share of the largest firm and the Herfindahl index play similar roles as in Cournot-Nash ($\upsilon_{-i}^* \equiv 0$); the difference is that the number of firms now also plays a crucial role—because it determines the endogenous competitive intensity as per υ_{-i}^*.

With symmetric firms, welfare losses then become $\widetilde{L}(n) = 1/(n^2+1)^2$. Losses are now of order $1/n^4$, and thus vanish *extremely* quickly as the number of firms rises. In effect, n firms now behave like n^2 Cournot competitors; even in a duopoly, losses are only 4%. The reason is that the conduct becomes endogenously more competitive with more firms; in addition to having "Cournot with more firms", "Cournot becomes more like Bertrand". Intuitively, there is more scope to manipulate rivals' behaviour if they are more numerous.

With asymmetric firms, the key point is that, given more intense competition, lower-cost firms capture larger market shares than under Cournot-Nash.[18] Turned on its head, this means that a weaker firm can sustain a given market share only if its cost disadvantage is less pronounced than under Cournot-Nash. This additional efficiency effect strongly limits welfare losses.

Ritz (2014a) shows that welfare losses now remain "small" (less than 5%) for many empirically relevant market structures. A simple sufficient condition is that the market share of the largest firm is no larger than 35%. Welfare losses are always small if firms are

[17] Instead using a differentiated-products Bertand model in which prices are strategic complements would lead to firms choosing to place *negative* weight on sales revenue ($\varphi_i^* < 0$), which seems at odds with empirical observation. In related work, Miller and Pazgal (2001) show that the equilibrium outcomes (and hence welfare) under differentiated Cournot and Bertrand can be *identical* if delegation contracts instead take the form of relative profits, e.g., $\Omega_i = \pi_i - \gamma_i \pi_j$ (for a fixed $n = 2$). While competition is as such tougher under Bertrand, this is exactly offset by the "soft" equilibrium contract featuring $\gamma_i^* < 0$—while $\gamma_i^* > 0$ under Cournot (strategic substitutes).

[18] Boone (2008) pursues this logic to develop a novel measure of competition based on how the relative profits of an efficient and a less efficient firm diverge more stronger when competition is more intense. Also related, Aghion and Schankerman (2004) study the welfare impacts of policies designed to enhance competition, and the political economy of their support, in a differentiated-products model with asymmetric costs.

not too symmetric or are sufficiently numerous (both in contrast to Cournot-Nash). In the numerical example with $s_1^* = 40\%$, $s_2^* = 30\%$, $s_3^* = 20\%$, and $s_4^* = 20\%$, welfare losses are just below 5% (instead of 18% under Cournot-Nash). These insights also extend fairly widely to non-linear demand systems.[19]

3.3 Discussion

The above welfare quantifications hold equally if firms' products are vertically differentiated in the eyes of consumers, due to (known) differences in quality. In particular, if firm i faced a demand curve $p_i = \phi_i + p(Q)$ where ϕ_i is a measure of vertical product differentiation, then the first-best has the firm with the highest "value-added", $\max_i \{\phi_i - c_i\}$, produce all output. At equilibrium, higher-quality firms tend to have too small market shares from a social viewpoint. However, like cost differentials, differences in product quality are fully captured in firms' observed market shares, allowing for welfare to be estimated.

Welfare losses, in practice, will be lower if the first-best outcome is not the relevant benchmark for comparison. For example, the most efficient firm may not be apply to supply q^{fb} because of capacity constraints or the government intervention that would be required to achieve first-best itself causes other welfare-reducing distortions. Welfare losses relative to any second-best optimum will be smaller.

These models can also speak to merger analysis. For example, as long as the *post*-merger market structure is sufficiently symmetric under Cournot-Nash or the largest firm has market share of less than 35% with delegation, then welfare "losses" remain small even after one or several mergers.[20] In this sense, the welfare impact of the mergers is limited, and there may be little rationale for policy intervention. Note that this is a different perspective from the usual approach in merger analysis: instead of testing whether or not a merger *reduces* in welfare, it focuses on whether the *level* of welfare losses remains "small" post merger (regardless of the direction of change).

[19] In related work on restructured electricity markets, Bushnell, Mansur & Saravia (2008) emphasize how retail-market commitments by vertically integrated players play a similar role to forward sales in Allaz & Vila (1993)—and how such long-term commitments can substantially improve market performance.

[20] Strictly speaking, this assumes that the underlying first-best welfare remains unaffected by the merger; this will be the case either if the most efficient firm is not involved in the merger, or if it does not experience any efficiency gains.

Conversely, welfare losses would be higher if either the mode of competition in the industry is (tacitly) collusive or if the approximate welfare standard is skewed more strongly toward consumers, e.g., $W_\lambda = \lambda\Pi + CS$ with $\lambda < 1$. Cournot-style equilibria with very concave demand (low cost pass-through) often produce high W but only low CS—and hence possibly also low W_λ.[21] For example, Cournot-Nash equilibrium with linear demand yields $CS^*/W^{fb} = 1/(1 + s_1^*)^2$, so consumer losses due to market power will be substantial—and sometimes *very* large—unless the largest firm is itself small relative to market.[22]

Other strands of the literature develop related models with endogenous conduct which may have similar welfare properties that lie between Cournot-Nash and perfect competition. One example is supply function models in which firms choose a set of price-quantity pairs to supply rather than being restricted to price *or* quantity choices (Klemperer & Meyer, 1989; Green & Newbery, 1992). In more recent work, D'Aspremont, Dos Santos Ferreira & Gerard-Varet (2007) and D'Aspremont & Dos Santos Ferreira (2009) develop a related way of endogenizing conduct parameters. Although welfare results for some limiting cases and specific examples are known, I am not aware of any general welfare analysis for such models.[23]

The finding that welfare losses due to market power are often quantitatively modest in Cournot-style models naturally leads to the question: What other market features could generate higher losses? One possibility is horizontal product differentiation which confers additional market power on firms (Corchón & Zudenkova 2009). Another possibility is welfare losses due to different forms of asymmetric information. Vives (2002) studies a symmetric (Bayesian) Cournot model in which firms have private information on their costs, and argues that informational losses can outweigh those due to classical market power. Its effect on deadweight losses is of order $1/n$ while that of market power is of order $1/n^2$. Put differently, a larger number of firms is more effective at curbing market power than reducing informational distortions. It would be interesting to know more about

[21] This may explain why policymakers often appear to have a distaste for low pass-through markets; while these often yield low deadweight losses, consumers typically capture only a small fraction of the gains from trade.

[22] This formula can be obtained heuristically by setting $\xi = 0$ and $H^* = 0$ in the expression for $L(s_1^*, H^*, \xi)$; superimposing a zero Herfindahl in effect takes away industry profits.

[23] Holmberg and Newbery (2010) study how deadweight losses vary with market structure, demand elasticity and capacity utilization in a application of the supply-function approach to electricity markets.

how such effects play out with (*ex ante*) firm heterogeneity.[24]

4 Social costs of endogenous entry

Recent work provides several refinements to the classic result that, in symmetric oligopoly, there is a tendency towards "excess entry": more firms enter than would be chosen by a social planner (Mankiw & Whinston 1986).[25]

In the long run, firms decide endogenously on whether to enter a market (at some cost, which is sunk). Amir, De Castro and Koutsougeras (2014) show for Cournot models that excess entry arises if and only if there is "business stealing": each entrant, to some degree, captures sales from incumbents rather than serving new customers; per-firm output satisfies $q'(n) < 0$.

Hence there is wedge between private and social incentives: some of an entrant's profits are a transfer from incumbent firms but yield no social gain; since entry is costly, this wedge matters for welfare. In free-entry equilibrium, each individual firm is too small from a social perspective.[26]

4.1 Quantifying welfare losses due to excess entry

Most of the existing literature examines a second-best setting in which the social planner cannot influence post-entry pricing, and focuses on qualitative results. In more recent work, Corchón (2008) quantifies the welfare losses L arising in a symmetric Cournot free-entry equilibrium, relative to the first-best social optimum—in which a single firm enters and price equals marginal cost. Welfare losses under free entry are sometimes very large,

[24] In recent work, Gabaix, Laibson, Li, Li, Resnick & de Vries (2016) highlight how price-cost margins (rather than welfare) under (symmetric) monopolistic competition can be much less sensitive to the number of firms than under Cournot. They show that, in a random utility model in which goods are homogenous but consumers are affected by random Gaussian "taste" shocks, markups are asymptotically proportional to $1/\sqrt{\ln(n)}$. One interpretation is behavioural: "consumer confusion" not captured in standard models of imperfect competition may result in significantly higher prices—even in "large" markets.

[25] Taken literally, the policy implication is that entry should be regulated or otherwise restricted. By contrast, under perfect competition the degree of entry by firms is welfare-optimal; more entry is then always a good thing for society.

[26] The same conclusions applies with a moderate degree of horizontal product differentiation, so each entrant adds only little extra variety of value to consumers. However the result can be reversed, leading to "insufficient entry", if competition in the market is very tough (e.g., undifferentiated Bertrand), even though at most by one firm "too few" (Mankiw & Whinston 1986). In recent work, Bertoletti & Etro (2016) unify many existing results from endogenous-entry models (with symmetric preferences and symmetric firms), covering Bertrand, Cournot, and monopolistic competition.

even with symmetric firms, because of the further cost misallocation.

Similar to the previous section, the approach is based on observables as far as possible. Assuming a free-entry equilibrium, the number of firms n is observed from market data.[27] The fixed cost of entry K are obtained as follows. (This can also be thought of as a fixed investment cost or R&D outlay required for market entry.) Let $\pi(n)$ denote per-firm Cournot profits (where $\pi'(n) < 0$). Since the n^{th} firm decided to enter, $K \leq \pi(n) \equiv K_{\max}$, while the $(n+1)^{\text{th}}$ firm staying out implies that $K > \pi(n+1) \equiv K_{\min}$. (This assumes a sufficiently large pool of potential entrants.) So the entry cost is bounded according to $K \in (K_{\min}, K_{\max}]$.

It is clear from Mankiw & Whinston (1986) that welfare losses are increasing in the size of the entry cost; indeed the social inefficiency disappears as the entry cost becomes small. Therefore, $L(n, K, \xi) \leq L(n, K_{\max}, \xi) \equiv L_{\max}$ and $L(n, K, \xi) > L(n, K_{\min}, \xi) \equiv L_{\min}$, where ξ is the familiar measure of (constant) demand curvature.

The limiting cases are instructive. First, with a large number of observed entrants in the industry, welfare losses tend to zero. In such cases, operating profits are driven down to almost zero, so the entry cost must have been tiny to have allowed so many firms to participate. Hence the outcome is essentially equivalent to perfect competition.

Second, with a very convex demand curve ($\xi \to 2$) industry profits are only a very small fraction of the overall surplus generated. Hence the entry cost sustaining n firms in the market cannot be very large, and so welfare losses are again tiny.

Third, and conversely, with a very concave demand curve ($\xi \to -\infty$), industry operating profits are very large relative to consumer surplus. So if some potential entrants nonetheless chose not too enter, then the fixed cost must be substantial—and so there is a lot of socially wasteful cost outlay. Indeed, if the fixed cost is large enough to wipe out all industry profits, then welfare losses tend to 100%. Specifically, Corchón (2008) shows that $L_{\min} = (n-1)/n \geq \frac{1}{2}$ while $L_{\max} = 1$.

This latter set of cases is interesting because it contrasts so strongly with a fixed number of firms. With exogenous n, welfare losses under symmetric Cournot tend to zero as $\xi \to -\infty$; the incentive for firms to withhold output disappears as they capture all

[27] This side-steps the problem of the integer constraint on n that arises when the number of firms is derived from market primitives on costs and demand.

surplus at the margin. By contrast, with endogenous n, the majority of this surplus is dissipated by fixed costs.

To get a feel for how welfare losses remain "large" in interior cases, consider the case with linear demand. Using the results in Corchón (2008), it is easy to check that $L_{\min}(4) \approx 21.8\%$ while $L_{\max}(4)$ is just over 30%. This is at least 5–7 times as high as the loss of 4% with an exogenous four firms. For a larger number of firms, the gap $[L_{\max}(n) - L_{\min}(n)]$ shrinks as per-firm profits decline. With ten firms, welfare losses are bounded by 13.5–16.0%; they remain above 5% until the number of firms exceeds 35.

4.2 Firm heterogeneity and endogenous entry

More recent work on endogenous entry has relaxed the assumption that potential entrants are symmetric, allowing for differences in firms' marginal costs and in the timing of market entry. By contrast, classic models of "excess entry" leave no room for competition to enhance productive efficiency via selection—and thus deprive it of one of its fundamental roles.

Vickers (1995) develops a simple Cournot example with unit-elastic demand (i.e., $p(Q) = S/Q$ with fixed industry revenue S) to illustrate how the adverse effects of entry may thus be overstated. Suppose that each firm discovers its unit cost (low or high) following entry; the industry already consists of three firms; the question is on the welfare impact of a fourth entrant.

If the entrant ends up being high-cost and at least two incumbents are low-cost, then it finds it optimal to not produce, so the externality from entry is zero. Even if only one incumbent is low-cost, the negative externality is less pronounced than under symmetry since business stealing mainly affects the high-cost incumbent.

The entry externality turns *positive* if two of the three incumbents are high-cost; entry by a low-cost firm then induces one of the less efficient incumbents to quit, and the efficient incumbent again expands output post-entry.[28] Surprisingly the literature does not appear to have generalized this example to richer market structures or to different forms of competition.

[28] The unit-elastic example is somewhat unusual in that an efficient incumbent regards rivals' outputs as a strategic complement.

Etro (2008) shows how a first-mover facing endogenous entry of followers typically behaves "more aggressively" than under simultaneous moves, and how this is good for social welfare. This stands in contrast with Stackelberg leadership against a fixed number of firms, which is well-known to be critically sensitive to the question of strategic substitutes (which leads to aggressive behaviour and a first-mover advantage) versus strategic complements (which yields a second-mover advantage).

Intuitively, endogenous entry means that the leaders' attention shifts from away from the reactions of followers at the margin (are strategies substitutes or complements?) to how its behaviour affects entry, that is, their participation constraints. Since products are substitutes, more aggressive behaviour (more output or lower prices) always leads to a favourable response: rivals' non-entry (or exit) becomes more likely.

To illustrate, consider quantity competition one leader and m potential followers. Demand is linear $p(Q) = 1 - Q$ and costs are zero—apart from the entry cost K.[29] The key point is that, with free entry of followers determined by a zero-profit condition, the number of *actual* entrants decreases with the leader's output. Etro (2006) shows that the equilibrium thus features strategic entry deterrence; the market leader produces $q_L = 1 - 2\sqrt{K}$, which prevents any entry, and the limit price is $p = 2\sqrt{K}$.

This simple example already has some interesting welfare implications. The price is higher than in the free-entry Cournot equilibrium (simultaneous moves), so consumers are worse off—contrary to the fixed-n Stackelberg logic. However, social surplus is nonetheless higher because of the profits made by the leader—which are associated with the saving on entry costs. The observed market structures are radically different: the market has flipped from n active firms with identical shares to a single quasi-monopolist.

Etro (2008) studies a general "aggregative game" in which each firm's profits depend on its own action and a summary statistic of those of its rivals combined, that is, firm i's payoff $\pi_i = \Pi(x_i, X_{-i}) - K$ where x_i is its own action (e.g., price or output) and $X_{-i} = \sum_{j \neq i} h(x_j)$ captures the "externalities" arising via the actions of other players, where $h(\cdot), h'(\cdot) > 0$. This setup nests as special cases quantity competition with differentiated products as well as price competition with logit and iso-elastic demand, amongst others. Typically the

[29] As is standard, the entry cost is assumed to be sufficiently low such that the market is not a natural monopoly. Both Nash and Stackelberg free-entry models converge to perfect competition as $K \to 0$.

leader produces more than under simultaneous moves or prices lower than the followers; this achieves a Pareto improvement in the allocation of resources.

The general implication is that large market shares of leading firms in an industry can be good news for social welfare; this also restores the notion of a first-mover advantage that prevails under both price and quantity competition. The details of a Stackelberg free-entry equilibrium depend on firms' strategic variables (price or quantity), the nature of product differentiation, and the shape of their cost functions.[30]

Mukherjee (2012) builds on these insights to show that the "excess entry" result can be reversed in markets with leadership and endogenous entry. The model has a single leader which enjoys a unit cost advantage relative to a tail of symmetric followers, and a linear homogeneous-products demand curve. The analysis is again second-best in that the social planner chooses the number of followers taken as given that they will engage in Stackelberg competition with the leader post-entry.

The main novelty is a "business creation" effect: the leader's optimal response to an increase in the number of followers is to *raise* production, $dq_L/dn > 0$. The reason is that its output rises with the followers' cost, and does so more stronger if there are more of them. Intuitively, the leader meets more rivals with a "fighting response" which leverages its cost advantage. (More formally, the leader's optimal output is supermodular in its cost advantage and the number of follower-entrants.)

The key insight is that the excess-entry result is reversed if the leader's cost advantage is sufficiently pronounced. Then the new business-creation effect dominates the standard business-stealing effect (which still exists amongst the followers), and more followers than delivered by the market would be socially desirable.

4.3 Discussion

The welfare metric used in the literature on endogenous entry is social surplus, so that the productive inefficiency arising from excess entry counts. Instead using consumer welfare,

[30] Anderson, Erkal & Piccinin (2015) analyze the welfare impacts of changes that affect only a subset of firms a market—such as a merger or a technology change—in a general aggregative-game setup. They show that the short-run impacts (e.g., a merger raises prices) of the change are often fully neutralized in the long-run with endogenous entry (i.e., the merger has no impact on prices). The key condition is that the marginal entrants, who make zero profits, are not directly affected, e.g., by the merger—and their actions effectively pin down the behaviour of the aggregate (and hence prices) over the long run.

an extra entrant is always socially desirable as long as it reduces prices; the market, if anything, delivers insufficient entry.[31]

The additional welfare losses that arise with endogenous entry thus have a similar effect to placing less weight on consumer surplus in the social-welfare function. Lowering λ in $W_\lambda = \lambda\Pi + CS$ pays less attention to profits either for normative reasons or because these profits are dissipated in another way. Incorporating wasteful rent-seeking costs which firms incur in securing market power (Posner 1975) has similar effects.

A central conclusion is that surplus losses remain large with endogenous entry even with a considerable number of firms in the market. Again, this conclusion can be substantially altered in two-stage models of competition. When post-entry competition is more intense, the inferred entry cost for any given number of observed entrants is well below that of Cournot competition, and so the additional source of cost misallocation is also much smaller. Welfare losses already drop below 5% whenever there are at least 4–6 observed entrants (Ritz 2014a).

Models of excess entry make the (sometimes neglected) assumption that entry occurs sequentially. While this is often reasonable, there are other examples in which a potential entrant may not know what entry decision other firms have made. Cabral (2004) provides a second-best analysis in which entry is a simultaneous process and either happens immediately or takes time as in a war of attrition. If entry costs are fairly low, then the results from sequential models are fairly robust. However, with high entry costs, the details on the timing of the entry process become important and "insufficient entry" more likely: from a societal perspective, a firm may be too fearful of an "entry mistake" (more firms enter than the market can support).

Finally, Kremer & Snyder (2015) emphasize a related source of welfare losses which arises from under-entry. Suppose that there is only a single potential entrant and that, from a social standpoint, it would be efficient for this firm to invest. However, if this firm is unable to appropriate a sufficiently large fraction of the surplus as profits, it will choose not to enter; as a result, the market does not come into existence.[32] Kremer &

[31] Some exceptions to this baseline result are known. Chen & Riordan (2008) show how more firms can sometimes lead to higher prices in a discrete-choice model with product differentiation; see also Cowan & Yin (2008) who study a related Hotelling setup.

[32] By contrast, the above models consider the welfare implications of endogenous entry where some entry has indeed occured—so there exists an observed distribution of firms' market shares.

Snyder (2015) provide worst-case bounds which take into account the possibility of such "zero entry", and argue that the resulting welfare losses can be large—for instance, in the pharmaceuticals industry in which consumers' valuations for products often vary widely.

5 Conclusion

The recent literature offers some new perspectives on a significant body of existing knowledge on oligopolistic competition and social welfare.

In a fairly broad class of oligopoly models, the division of surplus between firms and consumers is importantly determined by the rate of cost pass-through. Empirical estimates of pass-through across different markets thus offer indirect inference on welfare metrics. Yet pass-through is not a panacea in settings with firm heterogeneity, and the link between theory and the econometrics of pass-through still needs further tightening in future research.

The degree of welfare loss in widely-used Cournot-style models is often surprisingly modest, even relative to first-best and with significant industry concentration. Under Cournot-Nash competition, losses can be significantly higher due to cost asymmetries between firms yet their adverse impact is strongly limited in two-stage models with tougher competitive conduct. Losses are also typically much higher under a consumer-welfare standard. Future research could examine more closely the interaction between heterogeneity in firm's costs and asymmetric information.

Market performance is similarly reduced in dynamic models featuring "excess entry" which dissipates a significant fraction of firm profits. Recent work has extended these results to allow for Stackelberg leadership as well as differences in firms' costs. Both can be good news for social welfare, especially if the market leader also enjoys a cost advantage. Future research may focus on how these results map onto the empirical study of specific markets.

References

[1] Aghion, Philippe and Mark Schankerman (2004). On the welfare effects and political economy of competition-enhancing policies. *Economic Journal* 114, 800–824.

[2] Aiginger, Karl and Michael Pfaffenmayr (1997). Looking at the cost side of "monopoly". *Journal of Industrial Economics* 45, 245–267.

[3] Allaz, Blaise and Jean-Luc Vila (1993). Cournot competition, forward markets and efficiency. *Journal of Economic Theory* 59, 1–16.

[4] Amir, Rabah, Luciano De Castro and Leonidas Koutsougeras (2014). Free entry versus socially optimal entry. *Journal of Economic Theory* 154, 112–125.

[5] Anderson, Simon P. and Régis Renault (2003). Efficiency and surplus bounds in Cournot competition. *Journal of Economic Theory* 113, 253–264.

[6] Anderson, Simon P., Nisvan Erkal and Daniel Piccinin (2015). Aggregative oligopoly games with entry. Working Paper at University of Virginia, June 2015.

[7] Armstrong, Mark and John Vickers (2010). A model of delegated project choice. *Econometrica* 78, 213–244.

[8] Bagnoli, Mark and Ted Bergstrom (2005). Log-concave probability and its applications. *Economic Theory* 26, 445–469.

[9] Bertoletti, Paolo and Federico Etro (2016). Preferences, entry and market structure. *RAND Journal of Economics* 47, 792–821.

[10] Boone, Jan (2008). A new way to measure competition. *Economic Journal* 118, 1245–1261.

[11] Bresnahan, Timothy F. (1989). Empirical studies of industries with market power. In: *Handbook of Industrial Organization – Volume 2*, Richard Schmalensee and Robert Willig (eds.), 1011–1057. Amsterdam: North-Holland.

[12] Bulow, Jeremy and Paul Klemperer (2012). Regulated prices, rent seeking, and consumer surplus. *Journal of Political Economy* 120, 160–186.

[13] Bulow, Jeremy and Paul Pfleiderer (1983). A note on the effect of cost changes on prices. *Journal of Political Economy* 91, 182–185.

[14] Cabral, Luís M.B. (1995). Conjectural variations as a reduced form, *Economics Letters* 49, 397–402.

[15] Cabral, Luís M.B. (2004). Simultaneous entry and welfare. *European Economic Review* 48, 943–957.

[16] Chen, Yongmin and Michael H. Riordan (2008). Price-increasing competition. *RAND Journal of Economics* 39, 1042–1058.

[17] Corchón, Luis C. (2008). Welfare losses under Cournot competition. *International Journal of Industrial Organization* 26, 1120–1131.

[18] Corchón, Luis C. and Galina Zudenkova (2009). Computing welfare losses from data under imperfect competition with heterogeneous goods. *International Journal of Industrial Organization* 27, 646–654.

[19] Cowan, Simon and Xiangkang Yin (2008). Competition can harm consumers. *Australian Economic Papers* 47, 264–271.

[20] D'Aspremont, Claude and Rudolphe Dos Santos Ferreira (2009). Price-quantity competition with varying toughness. *Games and Economic Behavior* 65, 62–82.

[21] D'Aspremont, Claude, Rudolphe Dos Santos Ferreira and Louis-André Gerard-Varet (2007). Competition for market share or for market size: Oligopolistic equilibria with varying competitive toughness. *International Economic Review* 48, 761–784.

[22] Etro, Federico (2006). Aggressive leaders. *RAND Journal of Economics* 37, 146–154.

[23] Etro, Federico (2008). Stackelberg competition with endogenous entry. *Economic Journal* 118, 1670–1697.

[24] Farrell, Joseph and Michael Katz (2006). The economics of welfare standards in antitrust. *Competition Policy International* 2, 3–28.

[25] Farrell, Joseph and Carl Shapiro (1990). Horizontal mergers: An equilibrium analysis. *American Economic Review* **80**, 107–126.

[26] Fershtman, Chaim and Kenneth L. Judd (1987). Equilibrium incentives in oligopoly. *American Economic Review* **77**, 927–940.

[27] Gabaix, Xavier, David Laibson, Deyuan Li, Hongyi Li, Sidney Resnick and Casper G. de Vries (2016) The impact of competition on prices with numerous firms. *Journal of Economic Theory* 165, 1–24.

[28] Green, Richard J. and David M. Newbery (1992). Competition in the British electricity spot market. *Journal of Political Economy* 100, 929–953.

[29] Harberger, Arnold C. (1954). Monopoly and resource allocation. *American Economic Review* 44, 77–87.

[30] Holmberg, Pär and David Newbery (2010). Supply function equilibrium and its policy implications for wholesale electricity auctions. *Utilities Policy* 18, 209–226.

[31] Kimmel, Sheldon (1992). Effects of cost changes on oligopolists' profits. *Journal of Industrial Economics* 40, 441–449.

[32] Klemperer Paul D. and Margaret A. Meyer (1989). Supply function equilibria in oligopoly under uncertainty. *Econometrica* 57, 1243–1277.

[33] Kremer, Michael and Christopher M. Snyder (2015). Worst-case bounds on R&D and pricing distortions: Theory and disturbing conclusions if consumer values follow the world income distribution. Working Paper at Harvard University, January 2015.

[34] Lahiri, Sajal and Yoshiyasu Ono (1988). Helping minor firms reduces welfare. *Economic Journal* 98, 1199–1202.

[35] MacKay, Alexander, Nathan H. Miller, Marc Remer and Gloria Sheu (2014). Bias in reduced-form estimates of pass-through. *Economics Letters* 123, 200–202.

[36] Miller, Nathan H., Marc Remer and Gloria Sheu (2013). Using cost pass-through to calibrate demand. *Economics Letters* 118, 451–454.

[37] Nolan H. Miller and Amit I. Pazgal (2001). The equivalence of price and quantity competition with delegation. *RAND Journal of Economics* 32, 284–301.

[38] Mankiw, N. Gregory and Michael D. Whinston (1986). Free entry and social inefficiency. *RAND Journal of Economics* 17, 48–58.

[39] Mukherjee, Arijit (2012). Social efficiency of entry with market leaders. *Journal of Economics & Management Strategy* 21, 431–444.

[40] Posner, Richard A. (1975). The social costs of monopoly and regulation. *Journal of Political Economy* 83, 807–828.

[41] Reiss, Peter C. and Frank A. Wolak (2007). Structural econometric modeling: Rationales and examples from industrial organization. In: *Handbook of Econometrics – Volume 6A*, James J. Heckman and Edward E. Leamer (eds.), 4278–4415. Amsterdam: North-Holland.

[42] Ritz, Robert A. (2008). Strategic incentives for market share. *International Journal of Industrial Organization* 26, 586–597.

[43] Ritz, Robert A. (2014a). On welfare losses due to imperfect competition. *Journal of Industrial Economics* 62, 167–190.

[44] Ritz, Robert A. (2014b). A new version of Edgeworth's paradox of taxation. *Oxford Economic Papers* 66, 209–226.

[45] Ritz, Robert A. and Ansgar Walther (2015). How do banks respond to increased funding uncertainty? *Journal of Financial Intermediation* 24, 386–410.

[46] Rostek, Marszena, Marek Weretka and Marek Pycia (2009). Discriminatory or uniform? Design of divisible good auctions. Working Paper at University of Wisconsin-Madison, February 2009.

[47] Shapiro, Carl (1989). Theories of oligopoly behavior. In: *Handbook of Industrial Organization – Volume 2*, Richard Schmalensee and Robert Willig (eds.), 329–414. Amsterdam: North-Holland.

[48] Shapiro, Carl (1996). Mergers with differentiated products. *Antitrust*, Spring 1996. 22–30.

[49] Tirole, Jean (1988). *The Theory of Industrial Organization.* Cambridge: MIT Press.

[50] Vickers, John (1985). Delegation and the theory of the firm. *Economic Journal* 95, 138–147.

[51] Vickers, John (1995). Concepts of competition. *Oxford Economic Papers* 47, 1–23.

[52] Vives, Xavier (2000). *Oligopoly Pricing: Old Ideas and New Tools.* Cambridge: MIT Press.

[53] Vives, Xavier (2002). Private information, strategic behavior and efficiency in Cournot markets. *RAND Journal of Economics* 33, 361–376.

[54] Weyl, E. Glen and Michal Fabinger (2013). Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy* 121, 528–583.