University of East Anglia

CENTRE FOR COMPETITION POLICY

The Natural Monopoly Paradox: Incumbent Inefficiency and Entry

Sean F. Ennis
Centre for Competition Policy and Norwich Business School,
University of East Anglia

CCP Working Paper 23-01
This version: 22 February 2023

Abstract: The underlying concept of a "natural" monopoly is that costs for a given output are minimized when one firm produces that output. This paper shows that a paradox arises for certain infrastructure and delivery products satisfying widely accepted technical criteria for such monopolies: Under well-defined conditions, total costs can be reduced after the entry of a competitor. This finding is important if special state benefits, like protection against entry or direct subsidies, are provided to firms meeting the criteria. An alternative definition of natural monopolist is proposed that removes the counter-intuitive result. Plausible boundary conditions are derived for entry to reduce total costs. Surprisingly, applying the new definition to postal delivery, incumbents with low letter volumes least merit state protection, suggesting an increasing challenge from digitalization.

Contact Details:
Sean F. Ennis S.Ennis@uea.ac.uk

ISSN 1745-9648

The Natural Monopoly Paradox: Incumbent Inefficiency and Entry

Sean F. Ennis*
University of East Anglia

Draft: 22 February 2023

Abstract

The underlying concept of a "natural" monopoly is that costs for a given output are minimized when one firm produces that output. This paper shows that a paradox arises for certain infrastructure and delivery products satisfying widely accepted technical criteria for such monopolies: Under well-defined conditions, total costs can be reduced after the entry of a competitor. This finding is important if special state benefits, like protection against entry or direct subsidies, are provided to firms meeting the criteria. An alternative definition of natural monopolist is proposed that removes the counter-intuitive result. Plausible boundary conditions are derived for entry to reduce total costs. These conditions suggest that new investments in infrastructure merit careful examination. Surprisingly, applying the new definition to postal delivery, incumbents with low letter volumes least merit state protection, suggesting an increasing challenge from digitalization.

Key words: Natural Monopoly, Competition, Infrastructure, Post, Delivery

JEL Classification: D42, L43, L87, L97, L98

*Centre for Competition Policy and Norwich Business School, University of East Anglia, NR4 7TJ, UK. email: s.ennis@uea.ac.uk. The author expresses thanks for comments to Joan-Ramon Borrell, Michael Crew, Nathan Francis, Chistos Genakos, Morten Hviid, Paul Kleindorfer, Bruce Lyons, Jens Prufer, Luis Sauri, Carlos Suarez, participants at a CRRI Conference on Postal and Delivery Economics and at seminars at the Centre for Competition Policy, DG Competition, Judge Business School, KFTC, Ofcom, and the University of Barcelona. Research assistance was provided by Haiming Wu.

1 Introduction

Natural monopolies are typically defined as activities in which entry by another firm raises total costs, an intuition formalized in the classic definition of Baumol (1977). The existence of such monopolies can justify a government reversing its presumption in favor of market competition. The main reason for such a reversal is the expectation that total economic costs from an activity will be lower as a result of production by a single firm, with society benefiting from avoidance of duplicative costs. A government's decision to reverse its presumption hinges on the criteria used in the definition of a natural monopoly. This paper suggests that the criteria are incomplete in both infrastructure industries and delivery services, which can at times result in substantial resource misallocation, for example via excessive support for incumbents extending their infrastructure or via unnecessary protections and targeted subsidies to incumbent postal services that amount to tens of billions of dollars.

The definition has broad policy significance because wide-ranging regulatory implications, with ensuing effects on citizens and firms, follow from concluding that an activity constitutes a natural monopoly. These can relate not only to whether entry restrictions are instituted to avoid duplicative costs, but also to whether a government allocates subsidies directly to an incumbent rather than openly by tender, whether price regulation is required, whether third-party access requirements are put in place, oversight of cross-subsidization between monopoly and non-monopoly activities , whether structural separation is instituted between natural monopoly and potentially competitive parts of a firm and even whether government itself chooses to own, operate, or control an activity. Industries classified as containing nat-

ural monopoly activities constitute 10 percent or more of economic activity and are politically salient. They include electricity and gas, public transport, ports, post, telecommunications, rail, water, and waste.

The main reason that entry will typically be cost increasing for a natural monopoly activity is that fixed costs of the incumbent remain and the new fixed and variable costs of an entrant will outweigh any reduction in the incumbent's costs from lower output by the incumbent. The generalized characterization of multi-product natural monopoly based on subadditivity of costs of Baumol (1977) is widely adopted.[1] Crucially, this definition assumes all firms have the same cost function. This assumption is open to question due, for example, to possible X-inefficiency, rent sharing with labor, and managerial slack. In particular, a paradoxical result can arise when the incumbent cost function is a "substantial" distance from the frontier cost function. If fixed costs are moderate or avoidable and an entrant has lower variable costs than the incumbent, a transfer of output from the incumbent to the entrant has the potential to lower total costs of production, even if the two firms continue to operate. That is, entry into an activity with subadditivity of costs can, in practice, lower total costs.

The possibility follows from potential incumbent inefficiency, ignored by the standard definition, which suggests entrants may at times perform the same activities at lower cost. A modified definition is needed that could prevent inappropriate government classifications of activities as constituting natural monopolies. This paper proposes an alternative characterization of natural monopoly that resolves the paradox and then investigates the implications of this characterization. An initial review of the parameters suggested by the alternative

[1]See Sharkey (1982)), Waterson (1988), Joskow (2007).

definition suggests that there may be many countries in which an activity commonly considered as a natural monopoly could experience a reduction of actual total costs from entry.

This paper supplements existing literature by integrating the natural monopoly definition literature of Baumol (1977), (Baumol, Bailey, and Willig (1977)) and Panzar and Willig (1977)), with the productive inefficiency literature, as exemplified by the X-inefficiency hypothesis of Leibenstein (1966) (discussed also in Borenstein and Farrell (2000), Leibenstein and Maital (1992), and Perelman (2011)), the possible labor division of monopoly rents, and managerial slack theories.[2] The core observation of this paper is that incumbent inefficiency plays a potentially important role in the definition of natural monopoly. If the possibility of incumbent inefficiency is ignored, a paradox arises in which entry into a natural monopoly industry can, at times, actually reduce total costs. Developing an approach that resolves the practical inconsistency between the two strands of literature can impact policy, firms, and customers.

The proposed resolution is demonstrated in two cases by focusing on the extent to which costs of one operator are reduced when another takes its volume, and the extent to which an operator with increasing returns to scale is at the production frontier.

The first case is for infrastructure industries where network expansion and accompanying new investment would move society from one output vector satisfying subadditivity of costs to another that also features subadditivity of costs. Although both outputs may have costs

[2]Joskow (2007) notes origins of productive cost inefficiency can include: labor seeking part of monopoly rents, as with Salinger (1984), Rose (1987), Hendricks (1977) or work rules limiting business processes, and quality distortions such as Spence (1975). Another source can be managerial slack as suggested by Hart (1983), Holmstrom (1982), Machlup (1967) and Nalebuff and Stiglitz (1983).

that exhibit subadditivity in total, if the alternative additional investment costs are lower with an entrant than the avoidable costs for the incumbent, the continued characterization of the future structure as a natural monopoly may distort social decisions. As one example, the expansion of energy transmission and distribution infrastructure capacity is considered likely to the extent that vehicle electrification and electric heating become standard, with increases in network capacity needs by up to 30 percent. If incumbent costs are not on the cost frontier, particular policy consideration must be paid to alternative governance structures, such as alternative operators of new infrastructure facilities, to avoid a default of embodying non-frontier costs in an even larger monopoly infrastructure provider. This consideration would explicitly include any cost inefficiency of the incumbents.

The second case is for delivery activities, and in particular, letter post delivery, which is typically argued to be a natural monopoly (see, for example, Panzar (1991), Bradley and Colvin (1995) and Roy (1999)). The paradox is demonstrated with a probabilistic postal delivery model.[3] The model shows that when multiple deliveries to an address on a given day are common, the natural monopoly features of postal delivery are robust. However, when multiple deliveries to an address on a given day are uncommon, the natural monopoly features of postal delivery are weak: the substitution of a letter from an incumbent to a more efficient entrant can actually reduce costs of the delivery system overall. As many as one quarter of developed countries are shown to lie in the low-volume category that is potentially susceptible to cost-lowering entry, and even a higher ratio among de-

[3]In addition to postal delivery, other delivery or home-based services – like package delivery services, food delivery, garbage pickup, grocery delivery, newspaper delivery, home care – infrastructure sectors that are commonly considered natural monopolies could also share the requisite features for the paradox under certain conditions.

veloping economies. More generally, cost-reducing substitution from the incumbent to a more efficient entrant is the key to the rationale for a modified definition of natural monopolist.

To illustrate the impacts of the modified definition, note that the modification could generally impact policies that follow designation as a natural monopoly. For concreteness, consider the case of postal services. One policy impact is that the determination that natural monopoly characteristics are present in postal delivery has nourished policies that supported a protected monopoly status for post in the EU and the US, among other countries.[4] In a recent development more consistent with the modified definition, postal services have increasingly been prosecuted for anti-competitive actions related to letter delivery competitors.[5] These conflicting government approaches towards the sector illustrate a tension in how to analyze its activities. This paper identifies conditions in which protected status would not be justified, despite an activity being classified as a natural monopoly.

Another impact affects policies to address the financial difficulties of postal services. Domestic letter volume is shrinking by an average of about five percent per year, and billions of dollars in subsidies are at stake internationally. Even if postal services have shrunk in the last decades, they still account for 5.25 million jobs, including more than 500 thousand postal staff in the US alone.[6] A natural intu-

[4]The US guarantees a letter delivery and post box access monopoly to its postal incumbent the US Postal Service under Article 37 of the Civil Code. EU regulation permitted protected status to letter delivery until 2011 or 2013 depending on the Member State, and its members still provide direct subsidies to incumbents and can create special charges just for new entrants.

[5]Examples include the UK Ofcom case CW/01122/01/14, with a Competition Appeals Decision in 2019 and Appeals Court decision in 2021 (case C3/2020/0151), and the 2021 Italian competition authority decision in A493 Poste Italiane/Prezzi Recapito.

[6]The UPU estimates 5.25 million postal employees in 2021. The U.S. Postal Service reportedly had operating losses of USD 39.2b between 2010 and 2014 (Sidak (2015)). State

ition suggests that low-volume postal services are especially in need of subsidies. This paper argues that if incumbents are inefficient, this intuition merits refinement. When a postal delivery service does not meet requirements for being a natural monopolist under the definition proposed here, due to low volume, subsidies to maintain that postal service should not go automatically to an incumbent but could, for example, be considered also for entrants.

In a qualitative sense, most prior categorizations of an activity as a natural monopoly are not questioned by the approach advocated here. This paper focuses only on exceptions to the suitability of the standard definition. These exceptions can nevertheless be considered important both in principle and in the size of potential misallocation of resources that they encompass.

The precise boundaries of the definition of natural monopoly have been relatively little examined in recent years. These boundaries merit a renewed examination due to an increased economic focus on possible increased market power, recent inconsistent government approaches to natural monopoly, rapid infrastructure restructuring for carbon reduction, and decreasing postal volume from digitalization.

The paper proceeds as follows. Section 2 outlines a definition of the natural monopolist that explicitly takes into account questions of efficiency of incumbent operator. Section 3 shows how the paradox applies in standard infrastructure industries with large fixed costs. Subsequent sections focus on delivery, with section 4 developing a cost model under a Poisson arrival process of deliveries. Section 5 derives the impact of entry and provides boundary conditions for when entry is desirable. Section 6 calculates the parameters of inefficiency

aid decisions in the European Union allowed aid of about EUR 15 billion between 2009 and 2020, with support provided in countries like Belgium, Denmark, France, Germany, Greece, Italy, Poland, Spain, and the UK.

and frequency of delivery that could alter the classification of postal delivery as a natural monopoly, showing that a growing and nontrivial number of countries may be affected by the paradox. Section 7 concludes.

2 Definition

This paper proposes a definition of a natural monopolist that addresses the question of whether entry can lower total costs of producing a service that is currently provided solely by one firm in the particular context of delivery. This modified definition may be appropriate for selected sectors without large upfront investment costs or where substantial new investment costs are expected. The definition focuses on a specific company rather than the activity (or sector) more broadly, thus enabling the possibility that, for a given activity, an efficient firm may satisfy the criteria to be a natural monopolist while an inefficient one might not satisfy them.

Previous definitions of the natural monopoly, such as Baumol (1977), provide a rigorous assessment of whether a sector, starting from no installed base or served by one firm operating at minimum costs, is a natural monopoly based on subadditivity of costs.[7] This standard definition assumes that "the menu of available techniques is fixed (no technological change), that exactly the same menu of techniques is available to the monopolist and to each of its potential competitors, *and that all input prices are fixed.*" (italics in original). One of the reasons Baumol gives to justify these assumptions is the

[7]Baumol (1977) p.810 defines the "basic criterion" of natural monopoly as strict and global subadditivity of costs. A cost function $C(y)$ is strictly and globally subadditive in the set of commodities $N = 1, ..., n$, if for any k output vectors $y^1, ..., y^k$ of the goods in N that $\sum_{i=1}^{k} y^i = y$, $y^i \neq 0$, and $k \geq 0$, we have $\sum_{i=1}^{k} C(y^i) < C(y) = C(\sum_{i=1}^{k} y^i)$.

complexity of dealing with many different potential underlying cost structures. Yet exactly by avoiding this consideration of alternative cost structures, the assumption plays a critical role in determination of whether an industry is in a state of natural monopoly. Most importantly, the assumption can lead to paradoxical outcomes due to the likelihood that incumbents may have higher cost inputs than entrants following from X-inefficiency, labor sharing of monopoly rents, managerial slack, and sub-optimal investment in innovation by the incumbent.

In two important cases, developed in this paper, the theoretical possibility of entry lowering costs can be established. As a result, modification of the common cost function assumption embodied in the standard definition needs to be considered, at least with respect to those cases where the paradoxes are most likely to apply. The definition in this paper modifies the Baumol (1977) definition to answer the applied question of whether a *given* firm, such as a monopolist incumbent, is *in practice* a "natural monopolist," as opposed to whether a given service could be classified as a natural monopoly when operated *de novo* on the cost frontier. The difference is important because an activity may be determined to have natural monopoly status, according to the standard definition, when entry would demonstrably lower costs. Therefore, policy makers may wish to promote entry into "natural monopoly" services that are currently not provided by a "natural monopolist".

The definition proposed here thus distinguishes between the actual cost function of the incumbent, that need not necessarily coincide with the frontier cost function, and that of entrants, which is assumed to lie on the cost frontier.[8] Entrants are considered less subject than incum-

[8]The new entrant is assumed to lie on the cost frontier due to an absence of legacy

bents to cost-raising effects of political and union activity, and more subject to cost-raising activities resulting from incumbent lobbying for regulatory burdens. Making this distinction between incumbent and frontier costs is important because monopolists in a variety of sectors appear, in practice, not to operate on the cost function frontier. Possible examples would include various electricity, post and telecom companies. Studies that provide evidence of inefficiency include Newbery (1997) and Weyman-Jones (1991) for electricity, Perloff and Wachter (1991) for postal services, and Brennan (1990) and Crandall and Galst (1995) for telecom. These can be supplemented by arguments for x-inefficiency that include Borenstein and Farrell (2000) and Leibenstein (1966).[9]

Definition. A producer I is a y^m-natural monopolist at a positive vector of outputs y^m if and only if

$C_I(y^m) < C_I(y^1) + C(y^2) + ... + C(y^k)$

for all $y^1, ..., y^k$ such that

(i) $\sum_{i=1}^{k} y^i = y^m$,
(ii) y^1 exhibits non-zero production of all outputs, and
(iii) at least one $y^i \neq 0, i \geq 2$

where $C_I(y)$ is the cost function of producer I for producing a vector y and $C(y)$ is the minimum cost of producing a vector y.

The following definition relies on regulators having adequate knowledge of the relevant cost and demand structures, and wishing

costs, absence of union agreements and absence of pre-existing political support for special rights and benefits that could be divided with employees such as that found in Salinger (1984), Rose (1987), and Hendricks (1977).

[9]The point of inefficiency is distinct from, but usefully supplemented, by arguments of Kwoka (2006) that consumer welfare gains could be achieved by allowing entry in natural monopolies which may be larger than profit losses from higher industry costs.

to ensure that efficient cost-minimizing production techniques are utilized (Panzar and Willig (1977)).

Definition. A producer I is a "natural monopolist" if the producer is a y^m-natural monopolist throughout the relevant range.

The relevant range may reasonably be all y^m which are consistent with demand levels over the range of prices consistent with subsidy levels and costs between incumbent actual costs and frontier costs.

Having said this, the definition does not go so far as to require that the judgement of whether an industry is a natural monopoly for all possible input prices but rather acknowledges that incumbent monopolists may have different cost structures from those that are generally available to entrants, thus limiting the necessary range of cost functions to consider to that of the incumbent and the cost frontier. In the limit, an incumbent with the frontier cost function and subadditivity of costs will be a "natural monopolist" by default.

The definition includes the possibility that even if an entrant can have lower variable costs than an incumbent, the incumbent can still be a natural monopolist due to duplicative costs from entry.

The subsequent sections explore the application of this proposed definition of "natural monopolist" and the established definition of "natural monopoly" to infrastructure and delivery activities. These show that to the extent a "natural monopoly" can be subject to cost-lowering entry, the proposed definition of "natural monopolist" resolves the paradox.

3 High investment infrastructure industries

When considering infrastructure industries (water, electricity, gas, etc.) that satisfy the classical natural monopoly condition of Baumol (1977), the result shown in this paper that entry could potentially reduce total cost of an industry can be satisfied.

The standard definition applied based on actual costs of an incumbent would generally yield the accurate conclusion that a designated incumbent natural monopoly is the minimum cost producer compared to an industry with new infrastructure-based entry. This arises from the high efficiency gains that would be needed to offset costs of duplicative investment. Such efficiency gains are possible if incumbent inefficiency is at extremely high levels.

The paradox can reasonably exist in high investment infrastructure industries when there is a prospect of new investment by the incumbent that can be avoided if alternative investments are made by an entrant. Such situations are not rare.[10]

We can conclude the above arguments with the following proposition.

Proposition 1. A necessary condition for entry to reduce total costs for an activity with subadditivity of costs and incumbent costs higher than those of an entrant is that at least one of the following applies:
(i) variable costs are avoidable,
(ii) large sunk costs are absent, or

[10]For example, substantial new electricity network infrastructure could be required to meet increased demand for electricity to use by cars and for heating. Such investment needs could potentially be reduced by using alternative companies for the new investment or by investing in alternative infrastructure, such as home-based generation or larger-scale local alternative investments.

(iii) new sunk costs for the incumbent are avoidable if performed by an entrant.

Proof. The cost function of firms with divisible costs can be generalized for the incumbent as $C_I = F_I(q_I, q_E) + q_I c_I(q_I, q_E)$ and for a potential entrant as $C_E = F_E(q_I, q_E) + q_E c_E(q_I, q_E)$. The necessary but not sufficient conditions for entry to have the possibility of reducing costs include $\partial c_I/\partial q_E < 0$ or $\partial F_I/\partial q_E < 0$. If at least one of these conditions holds, there is a possibility that an entrant with a transfer of existing or future output from the higher cost incumbent to a low cost entrant could lower total costs for the sum of the outputs. This is most obviously seen for a hypothetical entrant with fixed costs of ϵ_f and variable costs of ϵ_v that are both arbitrarily close to, but greater than, 0. For each of the alternate necessary conditions, total industry production costs can be lowered from entry. □

While the focus of the subsequent sections is on lower variable costs for the entrant, with a model that is particularly amenable to showing with closed form solutions how total costs can fall and overlapping costs be reduced, the proposition emphasizes a more general phenomenon that costs of an incumbent can be "lowered" through avoiding either production that is more cheaply done by an entrant or avoiding future high-cost investments that can be made more cheaply by an entrant.

Absent at least one of these conditions, the standard natural monopoly definition and the "natural monopolist" definition will yield the same determinations for a given activity.

4 Delivery costs

There are informative contrasts between classical infrastructure industries and delivery services. While the cost of infrastructure for other natural monopolies is high and at times unavoidable, in the sense that an entrant's arrival would not normally reduce the infrastructure in place for the incumbent, delivery has different cost features related to overlapping services. On the one hand, sunk costs vary significantly between the postal sector and other industries, due to the low physical investment for routes. Overlapping service provision does not necessarily yield high waste either. Understanding these points with respect to natural monopoly requires the development of a delivery cost model. Such a model would capture the fact that potential entrants may already be present and passing homes, due to the substantial non-postal delivery activity taking place in many areas, including rural ones. Indeed, Bailey (1981) and Bailey and Baumol (1983) showed that competition is not substantially cost raising for natural monopolies if a large proportion of sunk cost is absent.

Moving on to focus on a model of delivery costs, suppose that initially there is a postal incumbent that is the unique postal delivery provider within its territory. It has an observed fixed cost of operation of F_I and an observed marginal cost for each delivery to an address of c_I.

In the case of postal delivery services, the output vector is the number of letters delivered to a given address on a given day by a given producer i.

Delivery route paths can be readjusted without costs as a function of the number of addresses to be reached. There are n addresses in total and routes are determined based on the expected number of

delivery addresses per day.

For any given uniform price of letters, there is an expected annual volume of letters per address $D(p)$. Letters are non-divisible. Thus there are no fractional letters. Deliveries to an address occur on every day that a letter arrives for an address. When no letter is delivered, the address is bypassed costlessly.

Let the mean number of letter arrivals per day per address be given by

$$\lambda(p) = D(p)/d. \tag{1}$$

where d is the number of delivery days in a year. Letter arrivals are assumed to be independent of each other and the probabilities are identical for each address and each day.[11] Letter arrival is modeled as a Poisson process such that the probability of arrival of at least one piece of mail in one day per address is given by $1 - e^{-\lambda(p)}$.

The expected number of days on which mail is received in a year is $d(1 - e^{-\lambda(p)})$, given the volume of letters generated by the price p.

Figure 1 thus shows the expected number of days receiving mail at different volumes. It is readily seen that as the number of pieces per address increases, the expected number of days of receiving mail increases but the rate of increase slows as the volume increases. Economies of scale are largest when most addresses receive daily delivery.

An interpretation of these curves is that in markets with virtually guaranteed multiple-piece delivery on a daily basis, there are substantial *de facto* fixed costs of operation (because the costs will exist even if incumbent scale is moderately reduced by entry). Moreover, in mar-

[11] In the interest of simplicity, this abstracts away from the fact that, in practice, certain addresses receive more mail than others and that pieces are more likely to be generated at certain times of year (such as year-end holidays) and certain times of the week.

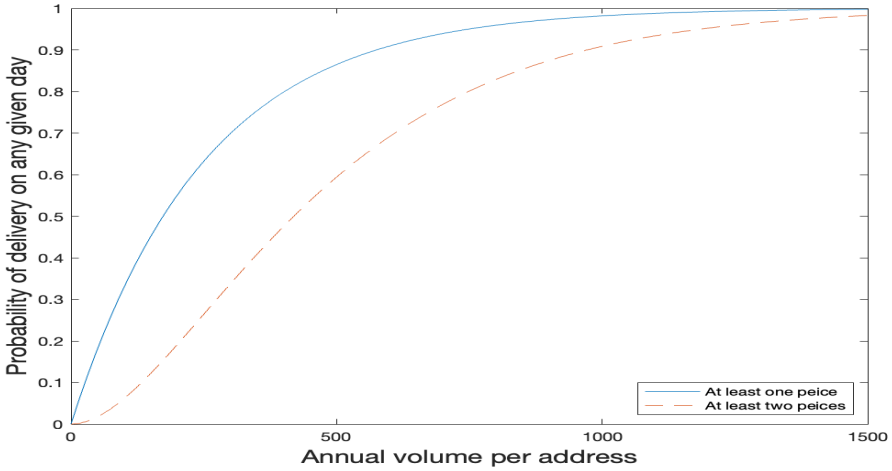


Figure 1: Probability of receiving mail on a given day

kets with one piece delivered on a day, the movement of that piece to an incumbent will not create overlapping variable costs of delivery, with such overlaps existing only when there are two or more pieces per day, as shown by the lower curve.[12]

The incumbent has only one delivery product. The incumbent's total costs for delivery in a year are given by

$$C_I = F_I + nd(1 - e^{-\lambda(p)})c_I. \tag{2}$$

The incumbent may operate at the cost frontier, but there is no requirement that this be the case.[13] Infrastructure sectors with natural

[12]Note that in markets with regular non-delivery, the incumbent delivery network does not need to operate at the same intensity when there is an entrant so fixed costs will be lower.

[13]Perloff and Wachter (1991) suggest that the wage premium for US Postal Service workers was 28 percent. Since 1990, further work has confirmed substantial wage premiums for most postal services, both internationally (Cohen and Chu (1997)) and in the U.S. where the Goldberg arbitration proceedings found that the total compensation premium was 34.2 percent as of 2001 (Goldberg(2002)). These findings do not consider work patterns that themselves may not yield the highest output per hour worked.

monopolies, such as electricity transmission networks, water networks and telecommunications wire-line networks to the home, have fundamentally different cost conditions from delivery. In particular, their fixed cost element has a high ratio to the variable cost element. As a result, overlapping physical networks have high duplicative fixed costs in these sectors.

In contrast, letter delivery features a low fixed infrastructure cost share, and a high variable cost. Much of the variable cost in delivery (arising from going to a delivery point like a home from a road) falls for the incumbent when a share of deliveries are dropped by the incumbent and completed by an entrant.[14]

The example of delivery illustrates that scale economies can come either from high "physical" investment, which leads to duplicative costs if replicated, or from scale that arises from "bundling" of multiple products (e.g., individual letters for one customer address). Entry in the bundled activity leads to lost scale when there is a division of bundles between incumbent and entrant; this may be counterbalanced by the gains from a lower cost structure or more innovative entry, if the frequency of delivering multiple units to an address is low compared to overall frequency of delivery of one unit.

5 Entry

Suppose that one entrant arrives and starts to offer postal delivery service. If it operates with the same delivery cost curve as the incumbent, then it simply increases system costs, though the entrant

[14]Cost functions may differ across route types, such as rural, suburban and apartment dwellers. The results of this paper still apply by route type. We maintain a focus on a single cost function to make the clearest distinction between the established definition of natural monopoly and the proposed definition of natural monopolist.

may potentially offer some benefits to mailers through selective entry and reduction in prices. However, this paper also considers the case in which the entrant has lower costs than the incumbent. To simplify the analysis of this paper, it is assumed that the marginal costs of serving an address are the same on all routes. Thus, this paper does not feature high-cost and low-cost routes, as in Crew and Kleindorfer (2001) but rather features high-cost and low-cost operators. For simplicity, the entrant is assumed to either enter for all routes or for none.[15]

Assume that the entrant has fixed costs $0 \leq F_E \leq F_I$, and a constant marginal cost of c_E where $0 \leq c_E \leq c_I$. The entrant thus has fixed costs and marginal costs that are weakly lower than the incumbent.[16]

Due to price regulation, the entrant is assumed to charge the same price p_0 as the incumbent. At this point, prices are assumed to be regulated. As a result, the total number of annual letters will remain constant after entry. This assumption will be relaxed later in the paper and will allow for a simpler exposition.

Suppose that s represents the share of letters that are delivered by the incumbent, where s satisfies $0 \leq s \leq 1$. The entrant then receives a share of $(1 - s)$ of letters. Then the total delivery cost after entry is

[15]In practice, differential entry could occur. For example, when a route is unprofitable for an incumbent, that does not mean that it will be unprofitable for an entrant. Thus it is possible that the entrant could move into the routes that the incumbent found unprofitable and even be paid to do so. The US Postal Service has actually contracted out the operation of certain of its (unprofitable) routes to contracted delivery suppliers, suggesting that entrants may consider such routes under certain conditions.

[16]This productivity difference could arise from greater workload per delivery person, from lower total compensation per year worked or from different combinations of capital and labor.

given by C_{IE} where

$$C_{IE} = F_I + nd(1 - e^{-s\lambda(p)})c_I + F_E + nd(1 - e^{-(1-s)\lambda(p)})c_E. \quad (3)$$

There are thus two cases to consider, one in which entrant costs are identical to those of the incumbent and the other in which entrant costs are lower than incumbent costs.

5.1 Entrant costs identical to incumbent

If the entrant's variable costs are identical to those of the incumbent, if entrant fixed costs are positive, and if the incumbent fixed costs (including future ones) are not reduced by entry, total costs will increase after entry. For positive mail volumes, variable costs will also increase, because there will be days on which both the incumbent and the entrant provide service to the same address (overlap delivery days). However, substitution from incumbent to entrant delivery may mean that, at low volumes, the variable cost increase is minimal.

Proposition 2. Entry will not reduce total cost if the entrant does not have a lower variable cost than the incumbent, i.e. unless $c_E < c_I$.

Proof. This proposition holds from recognizing that the overlapping probability is positive. □

Note that if the incumbent operates with the lowest possible marginal costs (regardless of whether its fixed costs are at the frontier), then entry will not reduce total costs because no entrant could have lower marginal costs. For such an incumbent, its own cost function and that of the fully efficient entrant are identical, so the incumbent is both a natural monopolist in the sense of this paper and operating in a natural monopoly in the sense of Baumol (1977).

This proposition shows the role for variable cost efficiency improvements for public utility provision sectors like electricity, water, and telecommunications networks, deemed as the "traditional" natural monopolies consistent with the definition by Baumol (1977). In contrast, if fixed costs are large and unavoidable, reduction of total costs will not be possible.

Proposition 3. Entry will not reduce total cost if the fixed cost for the new entrant is sufficiently large.

Proof. For the proof, see Appendix A. □

This proposition differentiates high-investment sectors from postal delivery's moderate fixed cost for new entrants, introducing competition can more easily benefit society by reducing the total cost than when fixed investment costs are a vast majority of costs. The proposition hinges on the lack of fixed cost reductions by the incumbent via substitution of new fixed costs to an entrant, i.e. that future infrastructure expenses are not avoidable for the incumbent in the face of entrant investment. The focus on the fixed cost *share* of total costs, effectively bounds the level of inefficiency in variable costs.

5.2 Entrant costs lower than incumbent

Given that prices are equal both before and after entry ($p_0 = p_1$) then a simple calculation shows the conditions that would be necessary for reducing total cost, namely that total costs after entry must be lower than total costs before entry. That is,

$$C_{IE} < C_I$$

or

$$F_I + nd(1 - e^{-s\lambda(p)})c_I + F_E + nd(1 - e^{-(1-s)\lambda(p)})c_E$$
$$< F_I + nd(1 - e^{-\lambda(p)})c_I.$$

This can be simplified to yield the proposition that follows.

Proposition 4. A sufficient condition for entry to have the potential to reduce total cost is:

$$F_E/nd + \left(1 - e^{-s\lambda}\right)\left(1 - e^{-(1-s)\lambda}\right)c_E$$
$$< \left(1 - e^{-(1-s)\lambda}\right)\left(e^{-s\lambda}\right)(c_I - c_E).$$

Proof. Full proof see appendix A. □

This states that entry reduces costs if the new fixed costs (from the entrant) and additional variable costs of an entrant's delivery to an address on the same day as an incumbent are less than the cost reduction from entrant substitution into non-overlapping deliveries. If the saving on variable costs is large and the addition of fixed costs by the new entrant is small, then the new entrant can reduce total system costs in low volume systems. The key to savings is a low probability that the one postal operator delivers to an address on the same day as the other. If this probability is high, as in high-volume markets, then the substitution effect away from incumbent towards entrant delivery is smaller.

This proposition identifies the conditions under which the incumbent is or is not a natural monopolist in the sense of this paper. From this finding, it is clear that an incumbent can fail to be a "natural monopolist", in the sense of this paper, while nonetheless operating in a "natural monopoly" in the sense of Baumol (1977). This is because the definition of natural monopoly answers the question of whether a

set of outputs would be provided at lowest cost by a firm operating at the cost frontier as opposed to an actual firm, such as an incumbent monopolist with inefficient operations.

An empirical finding of subadditivity of costs will always suggest the existence of natural monopoly. However, such a finding, when based on incumbents costs, is also consistent with unduly high production costs (if, for example, all delivery workers have a wage premium.) Thus empirical findings of subadditivity may not always be sufficient to guarantee that entry could not lower costs; instead, such findings may need to be combined with evidence that the incumbent operates on the cost frontier.[17]

5.3 Multiple entrants

The preceding model is based on entry by one firm. When multiple firms enter, the cost-lowering benefits of entry are often lower than in the case of a single entrant, provided that the total output of the incumbent and entrants, respectively, remain the same. This is because, in the case of postal delivery, the number of overlap delivery days will increase. That is, multiple entry will typically result in more total stops at an address even though the address receives the same total number of delivered items.

Important questions in addressing the impact of multiple entrants are whether the m^{th} entrant takes delivery volume exclusively from the incumbent, exclusively from other entrants or from a mix of the two. If the m^{th} entrant takes delivery volume exclusively from the incumbent, then it is possible that, because of the presumed cost advantage of the entrant, the per delivery cost advantage will cancel out the increased

[17]For one example of empirical findings of subadditivity in postal delivery, see Bradley and Colvin (1995).

cost from duplicative delivery.

Two extreme cases merit consideration: one in which a m^{th} entrant takes volume only from entrants and one in which a m^{th} entrant takes volume exclusively from the incumbent.

In the case in which the m^{th} entrant (with variable delivery cost c_E) takes volume only from prior entrants (with variable delivery cost c_E), and not from the incumbent, there is no cost reduction that can occur from entry. In fact, m^{th} entry will necessarily be cost increasing, even if fixed costs of entry are 0, because overlapping delivery will increase.

Proposition 5. Let there be $m-1$ entrants with equal market shares $Q_E/(m-1)$ and an incumbent with market share $1-Q_E$. Total volume is constant. An m^{th} entrant yielding equal market shares for Q_E/m for all entrants will necessarily increase costs.

Proof. Full proof see appendix A. □

The equal division of shares between entrants ensures that post-entry shares are not re-allocated after the m^{th} entry so that one of the m entrants has the bulk of the entrant volume while others are all fringe entrants. A contrasting result emerges if new entrant takes share exclusively from the incumbent.

Proposition 6. Let there be $m-1$ entrants with market share $Q_E/(m-1)$ and an incumbent with market share $1-Q_E$. Total volume is fixed at nd. An m^{th} entrant that achieves market share $s_m > 0$ and that takes share exclusively from the incumbent will reduce total costs if
(i) entry by the $m-1^{th}$ entrant reduced total costs and
(ii) fixed costs are 0.

Proof. This proposition follows because the costs of overlapping delivery arise from days on which an incumbent would otherwise have

delivered more than one letter. The probability of overlap for any given daily arrival rate of letters x to the incumbent is based on the probability that 2 or more letters would be in the incumbent delivery bag for a day:

$$p\{X \geq 2\} = 1 - e^{-x} - xe^{-x}$$

As the incumbent volume falls (that is, as x falls) the probability that the incumbent would have had two letters on a given day also falls. More precisely, the derivative of this expression is given by

$$\delta p/\delta x = xe^{-x}$$

.

For $x > 0$, this derivative is always positive.

If entry reduced total costs while driving incumbent mean arrival rate x to $(1 - s_{m-1})x$ for the $m-1^{th}$ entrant, then a further reduction from $(1 - s_{m-1})x$ to $(1 - s_m)(1 - s_{m-1})x$ would begin with a lower probability of delivery overlap between incumbent and entrant, so m^{th} entry would be more beneficial than $m - 1^{th}$ entry in lowering costs, for the same level of reduction in volume for the m^{th} entrant as with the $m - 1^{th}$ entrant. □

Up to this point, we considered solely cost impacts. While social welfare is often not discussed with respect to natural monopoly, there is value in doing so in this instance. The reason is that the nature of the failure to achieve natural monopoly in practice comes from a situation in which the incumbent is not on the cost frontier. In considering social welfare, but with price regulated as is generally the case in postal services, it is evident that changes in social welfare from

entry are based on the total cost of delivery.[18] Propositions 3 and 4 show the impacts of having multiple entrants under different scenarios of the alternative provider of delivery of a given letter.

Suppose the impact of the first entrant is cost reducing. The arrival of the first entrant is therefore going to improve social welfare. In contrast, the arrival of subsequent operators will only reduce social welfare depending on the source of its letters. A second entrant, operating at the cost frontier and that takes volume only from the first entrant will create higher system costs without any counterbalancing gain in consumer surplus (as price remains constant).

As a result, we have directly:

Proposition 7. If price is constant:
(i) Social welfare gains arise if total delivery costs fall as a result of entry.
(ii) If there is a y^m-natural monopolist at the actual vector of outputs m, no social welfare gain arises from entry.

Proof. At constant prices, and thus constant quantity, profits necessarily rise if total costs fall. Thus if costs fall as a result of entry, profits rise and producer surplus increases. At constant prices and quantity, consumer surplus remains constant. Thus social welfare increases from cost-reducing entry, establishing (i).

In contrast, by definition, when there is a y^m-natural monopolist, entry cannot reduce costs but would solely raise costs. Thus producer surplus would fall, while consumer surplus would remain constant, establishing (ii). □

A key intuitive point to draw from this is that each entrant creates

[18] Competition may still manifest itself via quality parameters such as speed and customization of delivery.

additional fixed costs. If the supplementary entrants after the first primarily take value from each other, supplementary entry at the cost frontier, can reduce social welfare, even if the first entrant increases it. As a result, unlimited entry may not be desirable for activities that meet the classic natural monopoly test, even in the situation in which one entrant can reduce total costs.

6 Calibration

This section derives the minimum efficiency improvement needed for entry to lower costs using the model developed previously. The realism, or lack thereof, of the relevant parameters allows us to assess the potential relevance of the model to actual situations. We begin by calculating the frequency of overlapping delivery on a given day for different volumes of mail. This frequency establishes how often there are two deliveries on the same day at the same address and assumes that the entrant has the same delivery days as the incumbent. In fact, entrants could choose to operate with fewer deliveries per week than the incumbent. This would reduce the total number of entrant deliveries in a year and would reduce the frequency of overlap.

Figure 2 illustrates the frequency of overlapping delivery as a function of the percentage of deliveries that are taken by one firm.[19] At 100 letters per address or less, there would be overlap on only 2 percent of days. These latter conditions are precisely those in which the natural monopoly definition of this paper may come up with different results than the more traditional natural monopoly industry definition. The parameters under which postal incumbents may not be y^m natural monopolists, in the sense of this paper, are explored below. While

[19]The reason to consider these percentages is that new entrants in post have not typically taken half the market but rather a much smaller share of letter delivery volume.

the model developed here would need refinement and greater focus on costs for different types of services to characterise real delivery systems, the basic parameter results suggest that one could potentially find postal natural monopolies, according to the traditional definition, in which the incumbent does not satisfy conditions of being a y^m-natural monopolist.

Figure 2: Frequency of overlap delivery based on share of deliveries handled by large firm

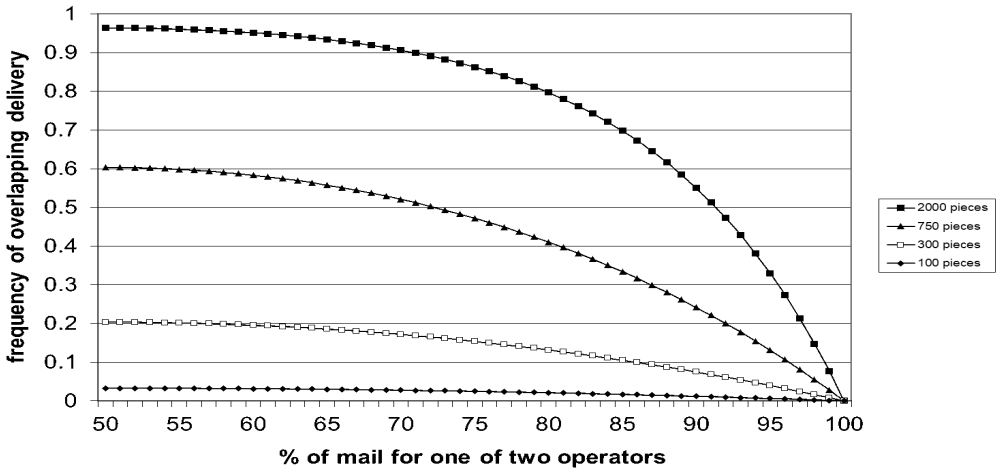


Figure 2 also shows that, as the share of the small firm approaches 0, the frequency of overlap also approaches 0, even for countries with high volumes. Given that, where liberalization has occurred, the share of letters handled by entrants has been small, the most relevant region of Figure 2, in practice, may be on the right, between 80 and 100 percent.

Figure 2 prompts the question of when countries would actually count as low-overlap countries. Table 1 shows the number of domestic post letter pieces per capita (column (1) and (2)) for reported EU and

EEA countries, household (HH) size (column (3)), letters per household (column (4)) in Europe and North America with data available in 2020. Volume per household has fallen by 33.9 percent on average since 2012 (column 5), presumably largely as a result of increasing digitalization. If current average trends continue, volumes will shrink as projected by 2030 (column 6).[20] It is immediately clear that there is a wide range in letter volumes, especially in Europe, and that many countries have quite a low volume, with the average being just below 200 and 10 countries featuring letter volumes below 100 in 2020, and 15 expected to have letter volume below 100 in 2030. In low-volume countries with fewer than 100 letters per household, overlapping delivery would prove a relatively rare phenomenon.[21] More broadly, based on Universal Postal Union data on deliveries per capita, 64 countries or more would have a volume below 100 per household, arising from the fact that postal delivery volumes are substantially lower in emerging economies, and the potential for postal delivery not to be a natural monopoly is consequently higher for emerging economies.[22]

To explore the feasibility of cost-reducing entry in a given coun-

[20]The purpose of the 2030 projections is purely illustrative. These are not intended to reflect all the parameters that could influence 2030 letter volumes but rather to show how the number of countries subject to the paradox noted in this paper could rapidly increase should current trends in letter volume reduction continue.

[21]Calculating the number of pieces per address is difficult because the figures available to the author do not provide the total number of addresses (roughly households plus business addresses) but only total number of households. For the U.S., where business address data is readily available, adding business addresses increases the number of addresses by 8.4 percent in FY2021. Therefore, the overlap simulation here will over-estimate the number of overlapping deliveries, due to excluding business addresses generally and thus having a smaller number of addresses than the actual number.

[22]The maximum value of the Universal Postal Union's "Average number of letter-post items posted per inhabitant" between 2018 and 2020 is found. Maximum values are used due to the high frequency of unreported data for individual years and to a covid effect. For simplicity, this individual figure is multiplied by four to find an approximate population of a residence. 64 economies were below 100, 31 above 100 and 79 unreported.

Table 1: Domestic letter mail service volumes

Country	(1) letters per capita 2012	(2) letters per capita 2020	(3) HH Size	(4) letters per HH 2020	(5) % Δ letters per HH 2012-2020	(6) 2030 (proj) HH 2030 (proj)
United States	474.8	354.8	3.1	1099.9	-26.0%	755.1
Switzerland	423**	335.7***	2.2	738.5	-24.1%	523.3
Finland	264.6	227.7	2.0	455.5	-18.0%	355.3
Luxembourg	286.3	186.7	2.2	410.7	-40.2%	215.9
Belgium	210.5	169***	2.3	388.7	-19.7%	295.4
United Kingdom	207.9	168.4***	2.3	387.3	-19.0%	297.6
Canada	272.7	154.3	2.5††	385.8	-43.4%	189.3
Germany	186.4	155.0	2.0	309.9	-16.9%	246.1
Austria	226.6*	140.3	2.2	308.6	-40.8%	160.2
Sweden	278.9	154.2	2.0	308.4	-47.3%	138.3
France	234.4	113.8	2.2	250.3	-51.5%	101.4
Iceland	112.8	74.6***	2.3†	171.6	-36.6%	97.0
Spain	67.3*	63.77****	2.5	159.4	-8.9%	141.9
Portugal	86.6	55.8	2.5	139.6	-38.0%	76.8
Croatia	72.8	50.5	2.7	136.4	-33.1%	82.5
Malta	79.3	49.2	2.4	118.0	-44.9%	56.0
Hungary	70.2	49.6	2.3	114.1	-32.3%	70.1
Cyprus	63.4	37.0	2.6	96.3	-45.8%	44.8
Serbia	42.3	32.2	2.9	93.5	-23.8%	66.6
Czech Republic	59.3*	39.3***	2.3	90.4	-36.5%	51.2
Poland	47.8	30.1	2.8	84.3	-37.1%	47.2
Italy	79.2	35.7	2.3	82.2	-56.8%	28.8
FYR Macedonia	22.9	16.2***	3.6	58.3	-31.2%	36.6
Romania	23.1	18.8	2.6	48.8	-21.7%	36.0
Greece	36.2	18.3	2.6	47.5	-49.6%	20.2
Latvia	32.0	19.8	2.3	45.6	-40.7%	23.7
Lithuania	24.6	17.5	2.2	38.5	-32.0%	23.8

Source: EC GROW, domestic postal data; EU Labour Force Survey; US Postal Service; Canadapost; WEO.
Individual-country geometric compounding applied for 2030 projections. EU states excluded if not present in EC GROW for relevant time period.
* 2013 ** 2014 *** 2017 ****2019
Average household (HH) size from official postal statistics, official government statistics or, where missing, † EU-28 average †† 2016 Census data

Table 2: Minimum productivity improvement required for cost reducing entry

	(1)	(2)	(3)	(4)
Country	**2020**	**2030 (projected)**	**letters $\geq$ 1 (2020)**	**letters $\geq$ 2 (2020)**
United States	97.0%	91.1%	98.8%	93.4%
Switzerland	90.6%	81.3%	94.8%	79.4%
Finland	76.7%	67.9%	83.8%	54.4%
Luxembourg	73.1%	49.9%	80.7%	48.9%
Belgium	71.2%	61.1%	78.9%	46.0%
United Kingdom	71.0%	61.4%	78.8%	45.9%
Canada	70.9%	45.4%	78.6%	45.6%
Germany	62.9%	54.5%	71.1%	35.2%
Austria	62.7%	40.1%	70.9%	35.0%
Sweden	62.7%	35.8%	70.9%	34.9%
France	55.1%	27.7%	63.3%	26.5%
Iceland	42.3%	26.7%	49.7%	15.1%
Spain	40.0%	36.5%	47.1%	13.4%
Portugal	36.0%	21.8%	42.8%	10.8%
Croatia	35.4%	23.2%	42.0%	10.4%
Malta	31.4%	16.4%	37.6%	8.2%
Hungary	30.6%	20.1%	36.6%	7.7%
Cyprus	26.5%	13.4%	32.0%	5.8%
Serbia	25.9%	19.2%	31.2%	5.5%
Czech Republic	25.1%	15.1%	30.3%	5.2%
Poland	23.6%	14.0%	28.6%	4.6%
Italy	23.1%	8.8%	28.0%	4.4%
FYR Macedonia	17.0%	11.0%	20.8%	2.3%
Romania	14.5%	10.9%	17.7%	1.7%
Greece	14.1%	6.2%	17.3%	1.6%
Latvia	13.6%	7.3%	16.7%	1.5%
Lithuania	11.6%	7.3%	14.3%	1.1%

Source: EC GROW, domestic postal traffic data; EU Labour Force Survey; USPS; Canadapost; WEO.
Individual-country geometric compounding applied for 2030 projections.

try, with our assumed cost function, we explore parameter values that yield the result that entry would lower costs. Suppose that fixed costs for the entrant are 0 due to the presence of pre-existing alternative delivery networks.[23] Then with 200 letters per address and a 50 percent penetration by the entrant, entry would reduce costs for productivity premiums of 50 percent or more.[24] With 100 letters per address and a 20 percent penetration by the entrant, entry would reduce costs for productivity improvement by the entrant of 38 percent or more. Finally, with 50 letters per address and a 10 percent penetration by the entrant, entry would reduce costs for productivity improvements of 20 percent and above.

Table 2 shows the productivity improvement for the entrant compared to the incumbent required for single firm entry (with a 20 percent market share) to reduce total costs of the letter delivery activity for both 2020 (column (1)) andfor the projected delivery volumes of 2030 from Table 1, to assess impacts of continued volume reduction on natural monopoly determination.[25] For countries with the highest volume of delivery, a very substantial productivity premium would be required, which would be unrealistic. It is interesting to note that in more than half of the countries in the sample, the productivity

[23]Package delivery and other companies regularly pass homes, even in rural areas, though precise statistics on number of homes passed daily by non-incumbent delivery services are not available.

[24]A "productivity improvement" reflects the difference in output per currency unit of compensation between the incumbent and entrant, combining the "compensation premium" with an "output premium" per hour worked.

[25]For high market shares for the entrant, the likelihood of overlapping delivery will increase, up to 50 percent, thus increasing the required productivity premium. However, this would start reducing as an entrant market share exceed 50 percent. This scenario is not considered here as entrants have generally not come close to fully displacing incumbents in those locations where entry has occurred. For some types of routes, this could nonetheless be the most desirable for both total costs and for incumbents to displace routes on which they lose money but an entrant could make profits due to higher productivity.

improvement necessary is under 40 percent for 2020. This could suggest that, if entrants would continue to have substantial productivity premiums on the order of those that could exist where incumbent productivity has been studies, more than half of countries might not be subject to natural monopoly characterisation, to the extent the simplified model is a valid perspective on delivery markets the vary substantially between rural, urban and different densities and layouts of delivery points.[26] The calibration suggests that entry would still not lower costs in the moderate-volume or high-volume countries in Europe or North America. More generally, this suggests that the question of when a postal delivery incumbent is a natural monopolist will be one whose importance increases in the years ahead.

Additionally, Columns (3) - (4) in Table 2 indicate the likelihood of more than 1 and more than 2 deliveries per day based on the country's letter volume in 2020. As shown previously, the cost reduction from entry is much less likely to occur when there would be two companies delivering to the same address on the same day. Thus Lithuania would have a 14.3 percent likelihood of at least one delivery per day and only 1.1 percent likelihood of two or more letters, meaning that a modest productivity improvement of 11.6 percent would be needed for entry to lower total costs. The minimum productivity improvement required increases with the probability of multiple letter delivery on the same day. The United States, for example has a likelihood of delivery of 2 or more letters on a given day of 93.4 percent. Only a small number of days remaining with one letter, when a non-overlap from entry would be guaranteed, and overlapping delivery would be highly likely. The required productivity improvement is consequently at a level of 97.0

[26] Any more definitive statements would require more detailed study of precise delivery conditions and their variation within a given country.

percent, one that is unlikely to be realizable, compared to 11.6 percent for Lithuania. If a 30 percent improvement in productivity is the maximum achievable level, that would be feasible only in the bottom 10 countries (in terms of volume) listed in 2020 and in the bottom 16 in 2030. These figures suggest that the number of countries with a "natural monopolist" in postal delivery can already be important for policy purposes and will increase substantially as letter volume shrinks, potentially accounting for as many as half of the countries reported here in the near future.

While postal delivery itself may meet the standards of sustainable natural monopoly, postal incumbents that operate at inefficient high costs may not have built-in cost-based protection against entry.[27]

The suggestion that low-volume countries may not have natural monopolies for postal delivery contrasts with the claim that countries with low letter volumes are those in which the postal operator has the most financial difficulty and where competition would be the most harmful.[28] Cohen et al. (2003) suggests that low-volume countries are the most susceptible to the "graveyard spiral" and thus that low-volume countries are where entry limits may be justified. In contrast, the model here suggests that these low-volume countries may be precisely those where entry limits are *least* justified because those are the ones with the lowest likelihood of duplicate delivery.[29] A cali-

[27]For a valuable discussion of sustainability, see Panzar and Willig (1977).

[28]See Okholm et al. (2018)for reference to risks, and Crew and Kleindorfer (2001) for a formalization of the graveyard spiral impacts.

[29]The reason for the difference in results is that the graveyard spiral is based on a model with a distribution of route types, ranging from dense urban to suburban to rural. In contrast, the primary model in this paper is based on a route model for a common address type, with the incumbent carrier having the highest delivery costs that are identical for each address and entrants having lower costs that are identical for each address. This simplification focuses the paper on the key issues of volume and productive efficiency that are not sufficiently considered when focusing solely on distributions of costs across routes.

bration with realistic parameters suggests that many European and low or middle income countries around the world might experience lower postal delivery costs after entry while being a natural monopoly according to the established definition.[30]

More generally, these calibrations are based on a simple model. Greater detail would be necessary before making any policy prescriptions from the model and data. However, it is clear that the extent of overlapping delivery would vary substantially between countries and many countries would likely experience little overlapping delivery from entry. In particular, many countries already have firms that operate highly inclusive daily delivery networks such as package delivery and food delivery to the home. Were a preexisting delivery network to add mail delivery, there would be little or no increase in overlapping delivery, to the extent a delivery is already made to that address. Even for mail entrants without a preexisting delivery network, if overlap is low and the incumbent is relatively inefficient, entry may, in very low volume countries, lower the total costs of delivery.

7 Conclusion

This paper illustrates that, even when standard conditions of a natural monopoly are met, entry may at times reduce total production costs, holding quantity constant. In such a circumstance, a unique incumbent firm may both operate in a natural monopoly sector but

The model can be extended to address variable route costs, noting that natural monopolist status may differ across route types with the same mail volume per household. Where unprofitable operation is found for even an entrant, the objective to meet universal service obligations may merit subsidies to the lowest cost provider.

[30]This paper does not consider sorting or the link between sorting and delivery. To the extent that, beyond a minimum efficient scale, sorting exhibits no economies of scale, this omission may not affect the results.

not, in fact, be a "natural monopolist" as defined in this paper. The conditions under which this might occur arise largely from three factors: an absence of high sunk costs in the sector, cost substitutability and productive inefficiency for an incumbent monopolist. The paper illustrates this argument in the context of two environments both of which exhibit high political saliency: high-investment infrastructure and postal delivery.

The infrastructure example mentioned illustrates the relevance of appropriate definitions of natural monopoly for the implementation of environmental policies related to reducing carbon output and directing energy provision towards electricity.

The delivery industry analysed here is an activity of particular importance in relation to the overall digitalization of the economy. Email and electronic communication result in a decline in letter mail. The cost model applied to letter services yields a surprising result: those postal incumbents that have the lowest volumes and that are increasingly facing the most serious financial challenges may be exactly the postal services where subsidies are inappropriate and competition would have the most beneficial effects. This is because the displacement of low-volume deliveries means that entrants may reduce the number of (inefficient) incumbent deliveries and thus reduce the total costs, given that the probability of overlapping deliveries for high-volume postal services is high.

The suggestion that low-volume postal services are not natural monopolies is different from saying that subsidies should be eliminated, as meeting universal service obligations may require state support. But such support may be most productive when provided to low-cost entrants. Even if the incumbent postal operator is not completely replaced by an entrant, due to political support for the incumbent,

customer loyalty, the need to support existing pension obligations, and integration of delivery with other parts of the incumbent operations, careful consideration can be given to the best way to direct any subsidies between incumbents and entrants. In practice, for an inefficient postal delivery service, substantial reductions in the total delivery cost could potentially be achieved by the incumbent adopting competitive contracting out of loss-making delivery routes to more efficient providers. Incumbents could have an incentive to do so even if the selected providers also accept mail from delivery services besides the incumbent service.

In both the energy and delivery cases discussed here, the pressure of potential and actual competition from entrants may have the beneficial effect of pushing incumbents to reduce any remediable inefficiencies in their operations. That is, the removal of statutory exclusivity protections for an incumbent can increase the pressure on that incumbent to increase its efficiency of operation and approach the cost frontier. Ironically, reducing the distance to the frontier, from competitive forces, can then enhance an incumbent's argument that it is a natural monopolist.

In general, knowing the incumbent's distance from the cost frontier matters in monopoly industries. When an industry meeting the technical requirements of a natural monopoly has an incumbent that is distant from the frontier, natural monopolist status should not be assumed. Close examination of the incumbent's cost structure, future investment plans, and avoidable costs is imperative. Such an examination can yield essential information to inform policymakers in their use of public funds and their application of regulatory powers.

Extensions to this paper would prove valuable. Further calibrations would illustrate when the definitions make a difference in ap-

plied policy decisions for a variety of sectors. With respect to delivery services, a valuable extension would consider the overall impacts of other delivery operators carrying out letter delivery, particularly since the volume of "letter" and "package" delivery may have a tendency to inverse in the future with digitalization and increasing package delivery. Such extensions would help clarify when policies based on natural monopoly characterizations merit deeper review and, potentially, reversal.

References

Bailey, E. E. (1981). Contestability and the design of regulatory and antitrust policy. *The American Economic Review*, *71*(2), 178–183.

Bailey, E. E., & Baumol, W. J. (1983). Deregulation and the theory of contestable markets. *Yale J. on Reg.*, *1*, 111.

Baumol, W. (1977). On the proper cost tests for natural monopoly in a multiproduct industry. *The American Economic Review*, *67*(5), 809-822.

Baumol, W., Bailey, E., & Willig, R. (1977). Weak invisible hand theorems on the sustainability of multiproduct natural monopoly. *American Economic Review*, *67*(2), 350-65.

Borenstein, S., & Farrell, J. (2000). Is cost-cutting evidence of X-inefficiency? *The American Economic Review*, *90*(2), 224–227.

Bradley, M. D., & Colvin, J. (1995). An econometric model of postal delivery. In M. A. Crew & P. R. Kleindorfer (Eds.), *Commercialization of postal and delivery services.* Boston, MA: Kluwer Academic Publishers.

Brennan, T. J. (1990). Cross-subsidization and cost misallocation by regulated monopolists. *Journal of Regulatory Economics*, *2*(1), 37-51.

Cohen, R. H., & Chu, E. H. (1997). A measure of scale economies for postal systems. In M. A. Crew & P. R. Kleindorfer (Eds.), *Managing change in postal delivery industries.* Boston, MA: Kluwer Academic Publishers.

Cohen, R. H., Robinson, M., Sheehy, R., Waller, J., & Xenakis, S. (2003). An empirical analysis of the graveyard spiral. In M. A. Crew & P. R. Kleindorfer (Eds.), *Competitive transformations of the postal and delivery sector.* Boston, MA: Kluwer Academic Publishers.

Crandall, R. W., & Galst, J. (1995). Productivity growth in the telephone industry since 1984. In P. Harker (Ed.), *The service productivity and quality challenge.* Boston: MA: Kluwer.

Crew, M. A., & Kleindorfer, P. R. (2001). Liberalization and the universal service obligation in postal service. In M. A. Crew

& P. R. Kleindorfer (Eds.), *Future directions in postal reform.* Boston, MA: Kluwer Academic Publishers.

Goldberg, Stephen B. (2002). *Supplemental opinion dealing with economic issues.* Opinion, Washington, D.C.

Hart, O. D. (1983). The market mechanism as an incentive scheme. *The Bell Journal of Economics*, 366–382.

Hendricks, W. (1977). Regulation and labor earnings. *The Bell Journal of Economics*, 483–496.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 324–340.

Joskow, P. L. (2007). Regulation of natural monopoly. *Handbook of Law and Economics*, *2*, 1227–1348.

Kwoka, J. E. (2006). The role of competition in natural monopoly: Costs, public ownership, and regulation. *Review of Industrial Organization*, *29*(1), 127–147.

Leibenstein, H. (1966). Allocative efficiency vs. "X-efficiency". *The American Economic Review*, *56*(3), 392–415.

Leibenstein, H., & Maital, S. (1992). Empirical estimation and partitioning of X-inefficiency: A data-envelopment approach. *The American Economic Review*, *82*(2), 428–433.

Machlup, F. (1967). Theories of the firm: Marginalist, behavioral, managerial. *The American Economic Review*, *57*(1), 1–33.

Nalebuff, B. J., & Stiglitz, J. E. (1983). Information, competition, and markets. *The American Economic Review*, *73*(2), 278–283.

Newbery, D. M. (1997). Privatisation and liberalisation of network utilities. *European Economic Review*, *41*(3-5), 357-383.

Okholm, H. B., Cerpickis, M., Möller Boivie, A., Facino, M., Gårdebrink, J., Almqvist, M., ... Apon, J. (2018). *Main developments in the postal sector (2013-2016): Study for the European Commission* (Tech. Rep.). Brussels, Belgium: Copenhagen Economics.

Panzar, J. C. (1991). Is postal service a natural monopoly? In M. A. Crew & P. R. Kleindorfer (Eds.), *Competition and innovation in postal services* (p. 219-231). Boston, MA: Kluwer Academic Publishers.

Panzar, J. C., & Willig, R. D. (1977). Free entry and sustainabil-

ity of natural monopoly. *Bell Journal of Economics*, *8*(1), 1-22.

Perelman, M. (2011). Retrospectives: X-efficiency. *Journal of Economic Perspectives*, *25*(4), 211–22.

Perloff, J. M., & Wachter, M. L. (1991). A comparative analysis of wage premiums and industrial relations in the British Post Office and the United States Postal Service. In M. A. Crew & P. R. Kleindorfer (Eds.), *Competition and innovation in postal services.* Boston, MA: Kluwer Academic Publishers.

Rose, N. L. (1987). Labor rent sharing and regulation: Evidence from the trucking industry. *Journal of Political Economy*, *95*(6), 1146–1178.

Roy, B. (1999). Technico-economic analysis of the costs of outside work in postal delivery. In M. A. Crew & P. R. Kleindorfer (Eds.), *Emerging competition in postal and delivery services.* Boston, MA: Kluwer Academic Publishers.

Salinger, M. A. (1984). Tobin's q, unionization, and the concentration-profits relationship. *The Rand Journal of Economics*, *15*(2), 159–170.

Sharkey, W. W. (1982). *The theory of natural monopoly.* Cambridge University Press.

Sidak, J. G. (2015). Maximizing the postal service's profits from competitive products. *Journal of Competition Law and Economics*, *11*(3), 617.

Spence, A. M. (1975). Monopoly, quality, and regulation. *The Bell Journal of Economics*, 417–429.

Waterson, M. (1988). *Regulation of the firm and natural monopoly.* B. Blackwell.

Weyman-Jones, T. G. (1991). Productive efficiency in a regulated industry: The area electricity boards of England and Wales. *Energy Economics*, *13*(2), 116-122.

A Proof of propositions

Proof of Proposition 3. Notice that the social cost reduction requires

$$C_{IE} < C_I,$$

or

$$F_E/nd + (1 - e^{-(1-s)\lambda})(1 - e^{-s\lambda})c_E < e^{-s\lambda}(1 - e^{-(1-s)\lambda})(c_I - c_E).$$

Denoting the average daily fixed cost of a new entrant for each household by f_E, suppose

$$f_E > -(1 - e^{-(1-s)\lambda})(1 - e^{-s\lambda})c_E + e^{-s\lambda}(1 - e^{-(1-s)\lambda})(c_I - c_E),$$

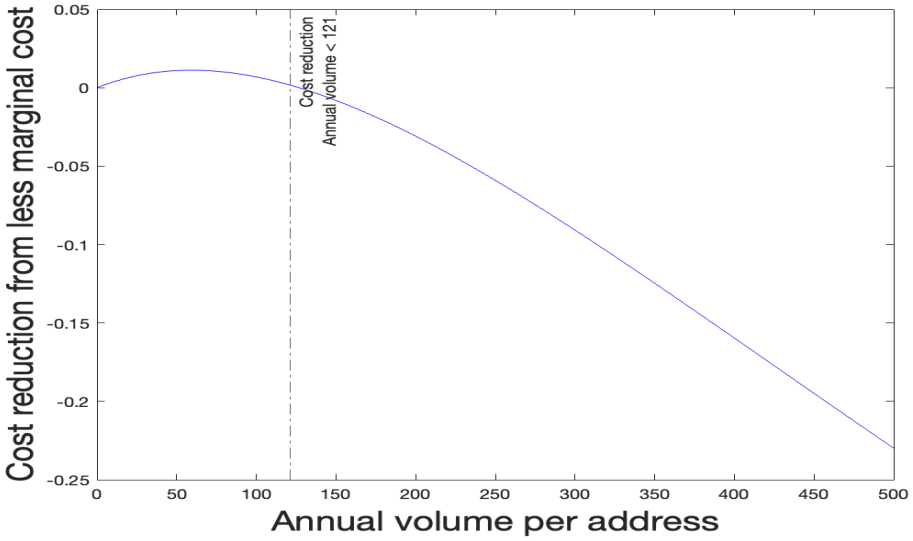
then we can never achieve total cost reduction for the high-investment infrastructure sectors from simple overlpapping competition. □

We can interpret this proposition with the help of a numerical example. Suppose the average fixed cost is $f_E = 2$, the average demand is $\lambda = 1.64$ (the average postal demand for US in 2020), the incumbent share after the entrance is $s = 0.8$, the marginal cost of incumbent and entrants are $c_I = 1.5$ and $c_E = 1$, respectively. After the calculation, we find that if we consider only the change of new entrant's marginal cost, even though c_E reduce to 0, we still cannot have a reduction in the social cost. However, if now the fixed cost of new entrant reduces to $f_E = 0.01$ and ceteris paribus, then entry will reduce the social cost.

Figure A.1 shows the case when we hold $c_I = 1.5, c_E = 1$ and $s = 0.8$, how the right-hand side expression we derived from the inequality above $(e^{-s\lambda}(1-e^{-(1-s)\lambda})(c_I-c_E)-(1-e^{-(1-s)\lambda})(1-e^{-s\lambda})c_E)$ changes with λ. We can see under our setting, new entry will reduce the total social cost only if both the average fixed cost and annual demand

per address are sufficiently low. For the parameters selected, this can occur only with an annual volume less than 121.

Figure A.1: Cost reduction from marginal change in λ (right-hand side of the inequality) for a fixed parameter combination



Proof of Proposition 4. From $C_{IE} < C_I$, we know

$$\begin{aligned}
F_I + nd(1 - e^{-s\lambda(p)})c_I + F_E + nd(1 - e^{-(1-s)\lambda(p)})c_E \\
< F_I + nd(1 - e^{-\lambda(p)})c_I) \\
\Rightarrow nd(1 - e^{-s\lambda(p)})c_I + F_E + nd(1 - e^{-(1-s)\lambda(p)})c_E \\
< nd(1 - e^{-\lambda(p)})c_I) \\
\Rightarrow F_E/nd + (1 - e^{-(1-s)\lambda})(1 - e^{-s\lambda})c_E \\
< (e^{-s\lambda} - e^{-\lambda})c_I - (1 - e^{-(1-s)\lambda})e^{-s\lambda}c_E \\
\Rightarrow F_E/nd + (1 - e^{-(1-s)\lambda})(1 - e^{-s\lambda})c_E \\
< e^{-s\lambda}(1 - e^{-(1-s)\lambda})(c_I - c_E).
\end{aligned}$$

□

Proof of Proposition 5. Notice that before m^{th} entry, the total cost in the market is

$$[F_E + nd(1 - e^{-\frac{Q_E}{m-1}\lambda})](m-1). \tag{4}$$

After m^{th} entry, the total cost becomes

$$[F_E + nd(1 - e^{-\frac{Q_E}{m}\lambda})]m. \tag{5}$$

It is easy to observe that (4) - (5) yields

$$\begin{aligned}
&m[F_E + nd(1 - e^{-\frac{Q_E}{m-1}\lambda}) - F_E \\
&\quad - nd(1 - e^{-\frac{Q_E}{m}\lambda})] - F_E - nd(1 - e^{-\frac{Q_E}{m-1}\lambda}) \\
=&mnd(e^{-\frac{Q_E}{m}\lambda} - e^{-\frac{Q_E}{m-1}\lambda}) - F_E - nd(1 - e^{-\frac{Q_E}{m-1}\lambda}) \\
=&nd(me^{-\frac{Q_E}{m}\lambda} - (m-1)e^{-\frac{Q_E}{m-1}\lambda}) - nd - F_E
\end{aligned}$$

We then need to check the value of $me^{-\frac{Q_E}{m}\lambda} - (m-1)e^{-\frac{Q_E}{m-1}\lambda}$. Notice that function $f(x) = x \cdot e^{-\frac{Q_E}{x}\lambda}$ will approximate $y = x$ but still below the line for $\lambda \geq 1$ and $0 \leq Q_E \leq 1$, since x is the dominant term here. Therefore, we have $m \cdot e^{-\frac{Q_E}{m}\lambda} - (m-1) \cdot e^{-\frac{Q_E}{m-1}\lambda} \leq 1$, this implies (4) - (5) < 0 for all m, and we show that the cost is increasing with the entry of firms. □