Essays in Industrial Organization

by

Christopher Gedge

Department of Economics
Duke University

Date: ____________________

Approved:

James Roberts, Supervisor

Peter Arcidiacono

Andrew Sweeting

Yi (Daniel) Xu

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2016

Abstract

Essays in Industrial Organization

by

Christopher Gedge

Department of Economics
Duke University

Date: ____________________

Approved:

____
James Roberts, Supervisor

____
Peter Arcidiacono

____
Andrew Sweeting

____
Yi (Daniel) Xu

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2016

Abstract

This dissertation extends the empirical industrial organization literature with two essays on strategic decisions of firms in imperfectly competitive markets and one essay on how inertia in consumer choice can result in significant welfare losses. Using data from the airline industry I study a well-known puzzle in the literature whereby incumbent firms decrease fares when Southwest Airlines emerges as a potential entrant, but is not (yet) competing directly. In the first essay I describe this so-called Southwest Effect and use reduced-form analysis to offer possible explanations for why firms may choose to forgo profits today rather than wait until Southwest operates the route. The analysis suggests that incumbent firms are attempting to signal to Southwest that entry is unprofitable so as to deter its entry. The second essay develops this theme by extending a classic model from the IO literature, limit pricing, to a dynamic setting. Calibrations indicate the price cuts observed in the data can be captured by a dynamic limit pricing model. The third essay looks at another concentrated industry, mobile telecoms, and studies how inertia in choice (be it inattention or switching costs) can lead to consumers being on poorly matched cellphone plans and how a simple policy proposal can have a considerable effect on welfare.

Contents

List of Tables

List of Figures

Acknowledgements

I owe special thanks to my advisors, Jimmy Roberts and Andrew Sweeting for their unwavering support and tireless encouragement. I would also like to thank my committee members Yi (Daniel) Xu, David McAdams and Peter Arcidiacono and other members of faculty Allan Collard-Wexler and Chris Timmins for helpful suggestions and guidance. I am extremely grateful to Tom Nechyba, the Duke Economics Department and Andrew Sweeting for providing considerable data acquisition funding and to seminar participants at Duke for helpful comments. All errors are my own.

1

Introduction

The airline industry and the market for telecoms have received widespread attention in the empirical and theoretical economics literatures. This is unsurprising. Pre-deregulation both industries were dominated by state-run firms or tightly regulated private-public entities. Post-deregulation both industries have remained highly concentrated with only a few firms in each market and despite this profits have been elusive. Highly disaggregated data is fairly readily available enabling the researcher to address interesting and important questions. Combining reduced-form analysis, calibration of an extended dynamic pricing model and utilizing a novel dataset this dissertation contributes to the empirical industrial organization literature by addressing three questions.

In Chapter 2 I analyze the effect of inertia on outcomes in the telecoms industry, specifically the UK cellphone industry. It has long been a concern of the telecoms regulator, OfCom, that consumers fail to switch to plans that would be more appropriate for their usage. This likely leads to a softening of price competition as firms face less price sensitive consumers as well as there being efficiency losses from consumers being on mismatched plans. Consumers enter into contracts that stipulate allowances for minutes, data and SMS, as well as corresponding overage fees. This makes a consumer's problem complicated. Not only do they have to forecast future usage prior to selecting a plan, but they also have to keep track of usage in relation to each allowance during the month. In contrast to other papers in this area I model the usage decision as consumers solving tasks that arrive by a random process during a month. Tasks can be solved using minutes, data or SMS, or not solved should the disutility from doing so be eclipsed by the cost of solving. Using a novel dataset I use the usage decision model to construct an expected cost of being on each plan. Consumers are then seen to choose their preferred plan, taking account of all future costs. But, importantly, consumers are not assumed to evaluate all plans every month (reflecting inattention) and even when they do decide to re-optimize they are (possibly) inhibited from doing so by switching costs. By simulating a simple and practical counterfactual where carriers are required to place consumers on to the most appropriate plan, *ex ante*, I find that inertia in choice is a significant (to the order of £200m per month) detriment to consumer welfare.

In Chapter 3 I shift my focus to the airline industry. This Chapter revisits a well-known puzzle in the literature: the Southwest Effect. Economic theory typically supposes that firms' behavior is constrained when facing actual competition. Yet, as documented by Goolsbee and Syverson (2008), when Southwest Airlines operates as a potential entrant on a market it prompts incumbent airlines to reduce prices by 15-20%. Several theories have been suggested for why incumbent firms might forgo profits today when they could instead cut prices should entry occur in a subsequent period. Joint work with Jimmy Roberts and Andrew Sweeting, in this chapter I use reduced-form techniques, including an interesting approach suggested by Ellison and Ellison (2011) to assess which of the theories is most likely.

Chapter 4 builds on Chapter 3 by building a model that captures what was found to be the likely determinant of firms cutting prices. Also joint work with Jimmy

Roberts and Andrew Sweeting this chapter extends a well-known model from the theory literature, limit pricing, to a dynamic setting where firms repeatedly interact for multiple periods. In this framework incumbent airlines are cutting prices so as to *signal* to Southwest Airlines that entry would be unprofitable. The signal sent relates to the marginal cost of the incumbent firm, which is a natural instrument to signal given the opaque nature of marginal costs in the airline industry. By drawing on results from Mailath (1987) and Mailath and von Thadden (2013) and applying a simple equilibrium refinement we provide a one-to-one mapping between a firm's observed price and its signaled cost that is both incentive compatible and individually rational. With this relationship in hand we calibrate the model using demand and cost parameters estimated from data to illustrate that a model of dynamic limit pricing can match the magnitude of price cuts observed in periods where firms are exposed to the Southwest Effect.

2

Switching Costs and Inattention in Telecoms

2.1 Introduction

In the United Kingdom ("UK") households spend more than 3% of their income on telecoms the majority of which relates to cellphone services.[1] There are 129 cellphones in use for every 100 members of the population, with almost half of these on a fixed-term contract. Annual firm revenue from cellphone services has averaged £14.9bn since 2009 and since 2007 individuals have spent more time on cellphones than on fixed-line telephones. Cellphones are increasingly being used as the primary means through which individuals access the internet.

Despite the impact on budget sets consumers are often found to make poor contract decisions or to be slow to switch when cheaper and better matched plans are available. Typically around 10% of consumers switch contracts within a 12 month period, but only a further 5% consider switching, meaning 85% could potentially be on mismatched plans by rolling over into the default option. The regulator, OfCom, has expressed concern about whether low levels of switching reflect a complicated and

[1] Statistics from this section were provided by OfCom (2012b).

confusing switching process[2] and that this may soften competition between carriers.[3]

This paper will empirically investigate how two impediments to selecting the optimal contract, consumer inattention and switching costs, affect market outcomes. It is well known (see e.g. Farrell and Klemperer (2007) or DellaVigna and Malmendier (2006)) that consumers tend to exhibit inertia, choosing not to switch to a new product even though it would be optimal to do so in the absence of switching costs. It has also become increasingly common in the literature (see e.g. Ho et al. (2015), Sims (2003)) to recognize that consumers may be inattentive such that even in the absence of switching costs consumer choice remains sticky. These two effects both act to steepen the demand curve and hence dampen competition between firms.

I use a three year panel to study individual-level contract choice and usage decisions from across all of the UK's five major carriers.[4] Descriptive evidence suggests that gains to switching are available, but that switching is far more likely to occur following an overage and even then only if the potential savings are sizable. Those that switch tend to do so in the right direction with average savings of £22 per month realized by those that switched (80% of the average monthly fee). Furthermore, almost 70% are observed never to switch with many individuals consuming far below their allowances, reflecting significant downwards inertia.

While these results show that inertia is present in the data they do not allow us to distinguish between whether this is induced by switching costs, inattention or both. Furthermore policies that reduce switching costs (e.g. telephone number portability, elimination of termination charges, etc.) will typically *not* increase attention (e.g.

[2] In its consultation on cellphone switching OfCom stated that "[it] is concerned these [switching] processes may cause confusion and increase the perception that swtiching is hard, meaning consumers may miss out on the best deals" http://media.ofcom.org.uk/news/2015/consumer-switching

[3] "Consumers should feel encouraged to take advantage of the best services and be confident they can switch without unnecessary hassle or risk. If they can't of they are put off even attempting to do so their ability to benefit from the market is reduced and competition suffers" OfCom (2015)

[4] I obtained approval for the acquisition and use of the data for this paper from Duke's IRB.

actions that "nudge" consumers to scan available plans and make cross-plan comparisons easier). Policy effectiveness therefore hinges on correctly identifying the source of the distortion. This is an empirical problem that this paper will seek to resolve.

To distinguish between the sources of inertia, measure their impact and understand the effect of counterfactual policies aimed at increasing efficiency in choice I build a structural model that jointly estimates the two forms of inertia and plan and usage preferences. In the model consumers make choices that maximize expected utility over all plan options, conditional on their tastes and usage distributions. Consumers thus select into plans *prior* to realizing usage shocks in the subsequent period. To model the cost associated with usage I adapt the framework of Hendel (1999) in which the cost to being on a plan is given by solving a series of tasks that arrive during the month. I allow for heterogeneity in switching costs and the size and frequency of usage shocks to provide a rich understanding for how consumers select plans.

The choice model estimates reveal significant inertia with average switching costs of £94 and more than 80% of consumers being inattentive. This translates into consumers considering switching only once every six months and for those that do consider switching the cost to do so is more than three times the average plan cost. These two effects result in the vast majority of consumers being on poorly matched plans.

I use these estimates to study two counterfactuals that reduce the effects of inertia from the baseline estimates. The first relates to a proposal recently suggested for the energy market in the UK whereby operators are required to place individuals on to the lowest cost plan each month. Whilst this means that consumers do not consider rival carrier plans (unless they are fully attentive) their choice sets are expanded to include all of their own carrier plans. This, easy to implement, proposal increases consumer surplus by £10.5 per individual-month, being equivalent to 40% of the tariff cost. This equates to an increase in consumer surplus of approximately £200m when extrapolating to the full population on fixed-term contracts.

In the second counterfactual I examine a range of policy interventions where the degree of attention and the level of switching cost are adjusted. I consider the cases of fully inattentive to fully attentive consumers and where switching costs are at their estimated levels through to their complete elimination. Such policies would involve a combination of nudging consumers (e.g. through information provision of the plans available) and reducing the hassle costs of switching (e.g. through number portability, reduced early termination charges, etc.). Whilst this counterfactual is likely harder to implement than the first, the gains from doing so are large. In the scenario of fully attentive consumers and zero switching costs consumer surplus increases by £30 per individual-month, or £600m per month for the population, with more than 90% of consumers switching.

The rest of the paper is organized as follows. In Section 2.2 I discuss the related literature, in Section 2.3 I provide an overview of the UK's telecoms market, Section 2.4 describes the data and stylized facts relevant for modeling, Section 2.5 lays out the model, Sections 2.6 and 2.7 discuss estimation and identification, Section 2.8 reports results, Section 2.9 runs the counterfactuals and Section 2.10 concludes.

2.2 Related Literature

Switching costs arise whenever a consumer experiences a cost, in addition to price, for the action of switching suppliers (or even among products with the same supplier). This acts to "lock-in" demand creating *ex post* market power with products perceived to have an extra degree of differentiation. The theory literature has proposed several sources of switching costs each of which can differentially impact on firms' pricing strategies and with possibly ambiguous welfare effects. It is therefore important for policymakers to identify which type of switching cost(s) is present in addition to quantifying its effect.

Perhaps the most obvious form of switching cost is an explicit, monetary, cost that needs to be paid (in addition to price) in order to purchase a product different from that previously consumed. These transaction costs (see Farrell and Klemperer (2007) for theory and Dubé et al. (2010) and Marshall (2015) for empirics) directly introduce inertia into choices through raising the utility associated with a previously consumed product relative to that of all other products. For cellphone contracts these costs are likely large. Switching from one plan to another typically incurs punitive early termination charges (ETC). It also introduces time and hassle costs of calling your current provider to negotiate an intra-carrier move (or worse still the bargaining over being released from a contract to switch across carriers) and researching available plans.[5]

A second form of switching cost relates to learning about the product (see Klemperer (1995) for theory and Goettler and Clay (2011) for empirics). If a consumer currently uses a Windows PC, having expended time to understand the operating system's features, switching to an Apple product necessitates for having to go through the learning process again. This increases the likelihood that the consumer will buy another Windows PC at replacement. In cellphone contracts learning costs are likely small. Contracts are fairly straightforward and, importantly, are homogeneous across carriers: using my phone on EE's network requires no new information if I had previously been on Vodafone's network.

Switching costs also relate to impediments to search and re-optimization. Typically the literature on inertia assumes that consumers have perfect knowledge of all product characteristics. However, it is quite likely that the decision process is two-staged. In the first stage a search cost is paid to acquire information before making

[5] OfCom (2016) found that 37% of switchers paid ETC, with 50% paying more than £50. Meanwhile 38% had major difficulties with switching due to the time to cancel service, understanding the switching process and researching alternative plans.

a purchase in the second stage. This explicitly recognizes that some consumers may remain on a dominated plan, either because the search cost is high or that they have *ex ante* biased beliefs about the value of new information. Recent work has started to model this two-stage framework (e.g. Ho et al. (2015)), but the difficulty researchers face is that the decision to search is rarely observed (although Honka (2014) is an exception). In cellphone contracts search and re-optimization costs are likely present, but small. Plan characteristics are similar across carriers and many comparison websites are available to assist with re-optimization.[6]

Consumers may also favor their current product over others due to psychological costs of switching. Thaler and Sunstein (2008) provide many examples of consumers reverting to default options due, for example, to cognitive dissonance or procrastination (see Dubé et al. (2009) and Crawford et al. (2011) on this as well). Significantly, these costs need not be irrational. Remaining attentive to all available cellphone plans and characteristics is burdensome. Avoiding these costs can be captured in a rational framework (e.g. Sims (2003) or the consideration set model of Ching et al. (2009)) as has been pursued recently by Hortaçsu et al. (2015) in demand for electricity providers.

This paper also adds to the literature on estimating models of multiple-discrete choice (as pioneered by Hendel (1999)). Modeling cellphone services is complex. Consumers enter into lengthy contracts for access to all or some of voice, data and text services. These services facilitate the tackling of tasks (i.e. the quantity of minutes used is the means to the ultimate end of task resolution) that arrive throughout the contract period. Conditional on tariff choice consumers decide first whether to tackle a given task and then with which component to do so. In this light the setup is similar to Hendel who looks at firms' demands for PCs. Yet, whereas firms are free to

[6] e.g. Billmonitor which provided the data for this study. Consumers grant Billmonitor access to their billing history which is then used to search for the best match among available plans.

choose from the available set of PCs as each task arrives, cellphone users must select into a tariff *prior* to observing usage shocks. This requires an additional step in the analysis, as compared to Hendel, where consumers make contract choice decisions in expectation of future costs. This also links the paper to the literature on health expenditure and insurance (Einav et al. (2013), Einav and Finkelstein (2011), Einav et al. (2015) and Handel (2013)) whereby consumers face a similar trade off between deductibles (allowances in my case) and tariff fees before health (usage) shocks are realized.

The two-stage demand process is discrete-continuous (see Dubin and McFadden (1984) and Reiss and White (2005)) with usage decisions (continuous) conditional on tariff choice (discrete). However, whilst the contract choice is straightforward usage decisions are not. A number of papers in the tax and utilities literatures (Chetty and Saez (2013), Chetty et al. (2009), Borenstein (2009), Ito (2014)) has found that consumers tend not to make labor supply/usage decisions as would be predicted by a fully attentive agent. The observed usage tends to be smooth as opposed to the predicted bunching around kink points on the demand curve. Liebman and Zeckhauser (2004) call this phenomenon "schmeduling", whereby consumers use simplifying mechanisms (rules of thumb) when faced with complex pricing schedules. We are more likely to "schmedule" when the connection between consumption and payoffs is weak (e.g. use of a hairdryer on electricity consumption or streaming a video on your cellphone from YouTube) and when goods are bundled (as in cellphones with minutes, data and texts) as this requires consumers to keep track of a multi-dimensional state variable (usage relative to allowances) and to forecast future usage for all components.

2.3 UK Mobile Telecoms Industry

As of 2012 (being the end of the sample period) the UK mobile industry is comprised of five Mobile Network Operators ("MNO"), more than 50 Mobile Virtual Network Operators ("MVNO") and several intermediaries. Whereas MNOs own spectrum licenses, network infrastructure and retail outlets, MVNOs only operate in the downstream market. As such MVNOs require access to an MNO's network in order to allow their subscribers to connect with other mobile users. This relationship is unregulated. OfCom (the UK's communications regulator) considers that five wholesale firms is sufficient for MVNOs to be able to gain access to MNO infrastructure through a competitive bargaining process[7]. The market share for all MVNOs is 15% (an increase from 10% in 2009), although this is disproportionately with two firms (Tesco Mobile and Virgin Mobile), each with more than 5%.

Consumers can enter into contracts for cellphone services with MVNOs, MNOs or intermediaries. Unlike MNO contracts, when taking out a contract with an MVNO a consumer is not necessarily aware of which MNO will provide the network infrastructure, although it is straightforward to obtain this information (e.g. from the MVNO's website). In all other respects a contract taken out directly with an MNO is similar to that with an MVNO. Popular opinion holds that some consumers prefer to enter into a MVNO contract due to bundling with other goods or brand familiarity. For example, Tesco Mobile is a subsidiary of the UK's largest supermarket (Tesco) and mobile customers obtain loyalty points which can be redeemed against supermarket purchases.

Downstream-only intermediaries (of which the two largest are "Carphone Warehouse" and "Phones4U") tend to offer contracts on all MNO networks and carry out

[7] OfCom did raise concerns of upstream market power in the new market for 4G due to the merger of two of the five firms and wholesale infrastructure sharing by two of the remaining firms OfCom (2012a)

the post-sales billing and customer service. They are distinguished from MVNOs in that they do not negotiate for access to an MNO's network, instead they provide a means of matching consumers with networks, for which they receive a commission from the MNOs for each contract initiated (similar to other two-sided markets, e.g. Opodo). It is likely that consumers enter into contracts with these firms so as to minimize search costs.

Communicating through use of data is increasing, yet, voice remains the largest contributor to firms' revenues[8]. 102bn minutes were used in 2011, an increase of 12% from 2010. Over the same period the increase in SMS (17%) and data (43%) was more pronounced, leading to a decline in revenue from voice to 68% of total mobile revenue (from 71% in 2010). SMS provides 60% of the remaining revenue (GBP2.5bn) with data constituting GBP2bn in the same period.

2.4 Data

The dataset is a panel of monthly billing records for 37,679 individuals with pay-monthly cellphone contracts. A billing record details tariff allowances for bundle components (minutes, data and text messages), corresponding monthly usage, overage charges disaggregated by component and carrier identity. I focus on consumers on *popular* plans that constitute more than 90% of bills in the dataset. This allows for up to 10 contracts for each firm in each period. In total there are over 196,414 individual-month observations. The data include all five major carriers in the UK and are from January 2010 to December 2012.

Table 2.1 contains data on market shares and the number of plans offered. The average per-period number of plans offered is similar for the five networks. Market shares are also stable with four large networks and one small, which is representative

[8] All statistics in this section were provided by OfCom (2012b) Communications Market Report 2012.

of the market in the UK.

Table 2.1: Plans and Shares by Period

	# Plans		Mkt. Share		N
	Mean	Std. dev	Mean	Std. dev	
Network 1	7	0.707	0.204	0.003	40,131
Network 2	5.11	0.93	0.254	0.012	49,980
Network 3	5	0	0.224	0.020	43,808
Network 4	7	0	0.256	0.01	50,176
Network 5	5.11	0.782	0.061	0.023	12,319
	5.844	1.127	0.2	0.074	196,414

Table 2.2 contains details of some of the popular plans. All plans are multiple part tariffs, with fixed fees (F_j) increasing in allowances (A_{jx}). For example, Network 1 has a basic plan with allowances of 300 minutes, 0 MB of data and 300 text messages with corresponding overage charges of 30p per minute, 12p per MB and 12p per text message when the allowances are exceeded. Bundle prices are highly nonlinear with quantity discounts intra component.

All plans have fixed contract lengths, the duration of which varies by carrier and over time, although those of 12, 18 and 24 months are the most common (98% of all contracts). Consumers can switch plans intra-carrier when under contract. Typically intra-carrier switching incurs a penalty charge if allowances are decreased, but not increased. Else, once a consumer selects a plan the contract terms will remain fixed until they terminate service or switch to another carrier. Switching inter-carrier prior to contract termination date incurs a penalty of the remaining monthly fees. Figure 2.1 plots the rate of within-carrier switching and quitting for each network-period and 2.2 illustrates the frequency of switches. On average 7.5% of individuals switch within carrier each month and 7.3% leave the dataset. A majority of individuals (69%) are not observed to switch and of those that switch 75% do so only once.

Table 2.2: Popular plans

	A_{Mj}	A_{Dj}	A_{Tj}	F_j (£)	p_x (£)
Network 1					
Small	100	0	∞	13.11	(0.37,056,-)
Medium	300	0	300	17.67	(0.35,0.056,0.13)
Large	600	∞	500	26.04	(0.41,-,0.15)
X-Large	2000	∞	∞	35.8	(0.41,-,-)
Network 2					
Small	100	0	∞	13.18	(0.36,0.028,-)
Medium	300	500	∞	20.88	(0.41,0.10,-)
Large	600	500	∞	27.48	(0.38,0.10,-)
X-Large	∞	1024	∞	35.02	(-,0.10,-)
Network 3					
Small	200	0	∞	18.57	(0.36,0.078,-)
Medium	300	500	∞	24.23	(0.36,0.10,-)
Large	600	500	∞	32.35	(0.34,0.056,-)
X-Large	1200	∞	∞	43.65	(0.36,-,-)
Network 4					
Small	100	0	500	11.47	(0.40,0.02,0.14)
Medium	300	500	∞	25.29	(0.58,0.10,-)
Large	600	500	∞	31.53	(0.58,0.10,-)
X-Large	900	750	∞	40.73	(0.58,0.02,-)
Network 5					
Small	300	500	5000	21.93	(0.36,0.10,0.3)
Medium	500	1024	5000	27.83	(0.35,0.10,0.20)
Large	2000	∞	5000	34.52	(0.40,-,0.31)

Notes: additional plans are also in the data and are used in estimation. These less popular plans, representing X% of all plans, typically have negative correlation between components. $A_x : x = \{M, D, T\}$ is the allowances for each component, with p_x the overage charge (per minute, MB, text message).

Reflecting similarity in the contracts offered usage is similar across networks (see Table 2.3). As contract terms vary the marginal price of using a minute will change leading to a high degree of variability across contracts. It is also likely that this reflects individual heterogeneity in preferences for using components. It will therefore be important to allow for usage to vary by contract and for individual heterogeneity in the model in addition to allowing for individuals to make decisions over the use of

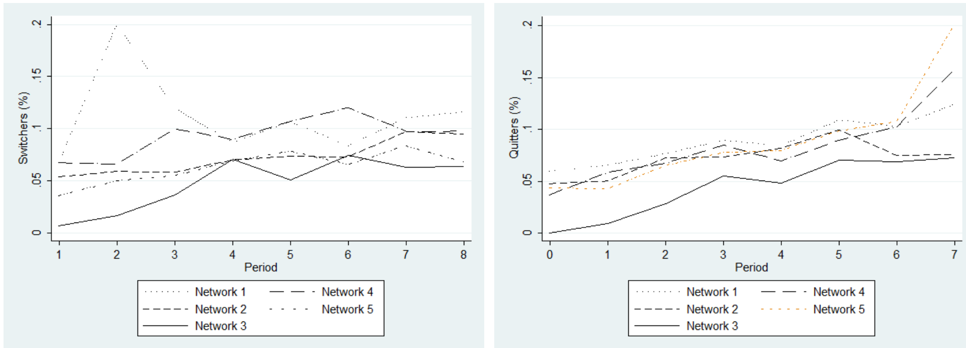

FIGURE 2.1: Percentage of individuals switching (left-hand plot) or quitting

multiple components and for varying quantities (i.e. the problem is one of multiple discrete choice as in Hendel (1999)).

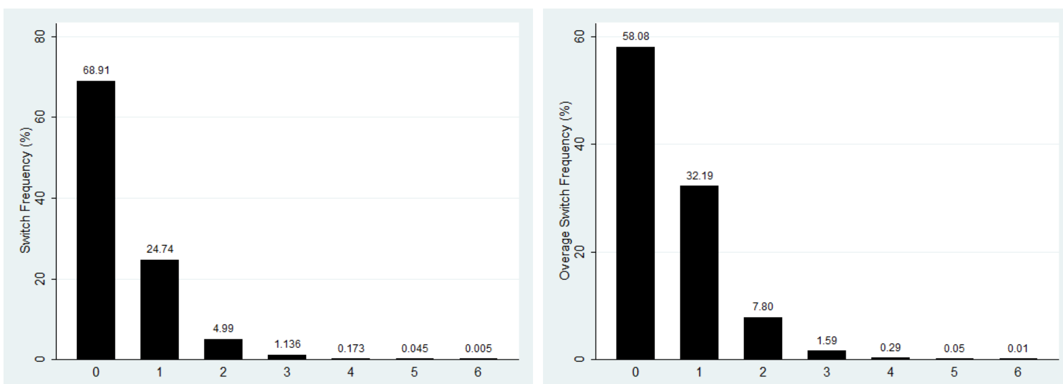

FIGURE 2.2: Frequency of switches per individual. Left plot is for all individuals, right plot conditions on incurring an overage in the previous month

Table 2.3: Usage Distribution by Period

	Minutes		Data		SMS		Overage (%)	
	Mean	Std. dev	Mean	Std. dev	Mean	Std. dev	Mean	Std. dev
Network 1	167.5	214.89	186.8	227.7	78.3	135.2	8.52	0.28
Network 2	185.73	253.16	185.02	187.44	251.05	375.03	8.89	0.284
Network 3	195.41	244.91	281.32	599.98	284.28	399.0	18.87	0.391
Network 4	161.85	176.25	149.25	257.42	237.64	355.95	8.54	0.279
Network 5	240.95	348.58	948.86	2841.33	240.10	371.97	5.10	0.22

2.4.1 Stylized Facts

There are five stylized facts observed in the data that are to be used to inform the model. Four of these relates to a consumer's monthly tariff decision, the other affects the intra-month usage decision.

Stylized Fact 1 - downwards inertia

Table 2.4 reports switching behavior. The data displays significant downwards inertia: for all bundle components it is rare that allowances are decreased following under-utilization in the previous month. However, switching behavior is predictable: allowances are most likely decreased when previous month's utilization is less than 25% and least likely when an overage was incurred in the previous month. Conversely allowances are increased after 8.24% of overages for minutes, 5.13% for data and 10.34% for text messages. Presumably reduced stickiness upwards relates to lower switching costs in this direction and that consumers are more aware of their utilization when having exceeded an allowance than when having not done so.

Stylized Fact 2 - gains to switching are available

If switching costs are zero individuals would switch whenever the cost of their current plan exceeded that of a rival plan. However, the rate of switching observed in the data is far lower than this would suggest. Table 2.5 indicates that £15.15 could be saved by switching to a rival plan. This overspending variable is constructed by comparing the cost of observed usage on the current plan to that on the cheapest available plan on the same carrier. This overspending can be decomposed into savings in the monthly fee, £10.52, and savings in usage cost, £4.63.

Table 2.4: Switching Decision

	$\mathbb{1}[A_{Xt} < A_{X(t-1)}]$	$\mathbb{1}[A_{Xt} = A_{X(t-1)}]$	$\mathbb{1}[A_{Xt} > A_{X(t-1)}]$	N
$\frac{M_{t-1}}{A_{M(t-1)}} < 0.25$	3.18	94.77	2.04	72,060
$\frac{D_{t-1}}{A_{D(t-1)}} < 0.25$	4.69	93.72	1.59	80,944
$\frac{T_{t-1}}{A_{T(t-1)}} < 0.25$	0.79	98.75	0.46	135,567
$\frac{M_{t-1}}{A_{M(t-1)}} \in [0.25, 0.5)$	2.88	94.29	2.83	43,150
$\frac{D_{t-1}}{A_{D(t-1)}} \in [0.25, 0.5)$	1.49	95.61	2.90	24,587
$\frac{T_{t-1}}{A_{T(t-1)}} \in [0.25, 0.5)$	0.48	94.71	4.8	16,513
$\frac{M_{t-1}}{A_{M(t-1)}} \in [0.5, 1)$	2.02	94.05	3.92	32,890
$\frac{D_{t-1}}{A_{D(t-1)}} \in [0.5, 1)$	1.31	94.37	4.32	23,038
$\frac{T_{t-1}}{A_{T(t-1)}} \in [0.5, 1)$	0.64	93.62	5.73	4,675
$\frac{M_{t-1}}{A_{M(t-1)}} \geqslant 1$	1.96	89.8	8.24	9,774
$\frac{D_{t-1}}{A_{D(t-1)}} \geqslant 1$	3.06	91.8	5.13	3,720
$\frac{T_{t-1}}{A_{T(t-1)}} \geqslant 1$	1.19	88.47	10.34	841

Notes: values in cells are percentages, e.g. 3.18% of individuals decreased A_M conditional on having used less than 25% of their allowance in the previous month and 10.34% increased A_T condtional on incurring an overage in $t-1$.

Table 2.5: Decompoistion of Gains from Switching

	Fee Savings (£)	Usage Savings (£)	Total (£)
All	10.52	4.63	15.16
Switchers			
All	10.86	10.98	21.85
$\mathbb{1}(Overage) = 0$	11.71	3.92	15.64
$\mathbb{1}(Overage) = 1$	6.18	50.04	56.22
Non-switchers			
All	10.51	3.54	14.04
$\mathbb{1}(Overage) = 0$	10.96	1.08	12.04
$\mathbb{1}(Overage) = 1$	5.10	30.74	35.84

Stylized Fact 3 - switching in response to shocks

From stylized fact 1 we know that switching occurs more often following an overage. Individuals also appear to respond to shocks of cheaper plans being available. Table 2.6 reports the following linear probability regression:

$$\mathbb{1}(Switched) = \beta_0 + \sum_{x=\{M,D,T\}} \beta_x \mathbb{1}(X_{t-1} > A_{x,t-1}) + \beta_1 \#Cheaper + \beta_2 Overspend \tag{2.1}$$

where the dependent variable equals 1 if the individual switched in the current period and $\#Cheaper$ is the number of cheaper within-carrier plans. The results are consistent with a story in which individuals consider switching when an overage is incurred, but then choose to switch only if the gains from switching are sufficiently high. Identifying these gains is made easier the more plans there are that are cheaper, given current usage. Similar results were found when the dependent variable equalled one if a consumer was observed to leave the dataset (i.e. moved to a rival carrier or to a pay-as-you-go contract). The results suggest that leaving the dataset is better interpreted as moving to a competitor's plan with the probability of quitting increasing in the number of plans that are cheaper across carriers.

Stylized Fact 4 - switchers make good decisions

Figure 2.3 plots the gains from switching, separately for switchers and non-switchers. For the set of switchers this illustrates the amount of overspend if an individual remained in their current plan, holding usage at observed levels. For non-switchers the graph illustrates that around £14 could be saved each period if they switched to the lowest cost plan. Whilst non-negligible (this represents a saving of 35.6%), the savings for switchers, £22 (or 67.3%), are significantly higher (see also table 2.5). These extra savings help to rationalize why some individuals switch whereas others do not. However, it doesn't necessarily imply that the switch is in the right direction.

Table 2.6: LPM: Intra-Carrier Switching

	(1)	(2)	(3)	(4)
	$\mathbb{1}(Switched)$	$\mathbb{1}(Switched)$	$\mathbb{1}(Switched)$	$\mathbb{1}(Switched)$
$\mathbb{1}[M_{t-1} > A_{M(t-1)}]$	0.0439***	0.00409	0.0147**	0.00392
	(0.00465)	(0.00541)	(0.00654)	(0.00674)
$\mathbb{1}[D_{t-1} > A_{D(t-1)}]$	0.0490***	0.0155**	0.0123	0.00646
	(0.00603)	(0.00672)	(0.00835)	(0.00836)
$\mathbb{1}[SMS_{t-1} > A_{SMS(t-1)}]$	0.119***	0.0949***	0.127***	0.119***
	(0.0151)	(0.0171)	(0.0203)	(0.0203)
# Cheaper in network plans		0.00671***		0.00443***
		(0.000524)		(0.000657)
% Overspend			0.0145***	0.00757***
			(0.00188)	(0.00203)
Constant	0.0345***	0.0384***	-0.000652	0.00353
	(0.00178)	(0.00295)	(0.00710)	(0.00704)
Household Fixed Effects	Yes	Yes	Yes	Yes
N	178463	158735	121067	121067
R^2	0.225	0.253	0.296	0.296

Robust standard errors in parentheses

** $p < 0.05$, *** $p < 0.01$

To that end, table 2.7 reports results from a regression of the amount of overspend on plan characteristics and a dummy variable for whether an individual switched. Having switched plans is negatively related to overspending.

Table 2.7 and figure 2.3 also allow us to rule out two alternative explanations for why individuals switch. In the figure the amount of overspend is fairly constant over time suggesting that individuals are not learning about usage (as in Grubb and Osborne (2015) or Goettler and Clay (2011)) . Given that long-term contracts have been in place for a number of years this is as expected. Table 2.7 allows us to rule out risk aversion. If individuals were selecting into large plans so as to avoid overage

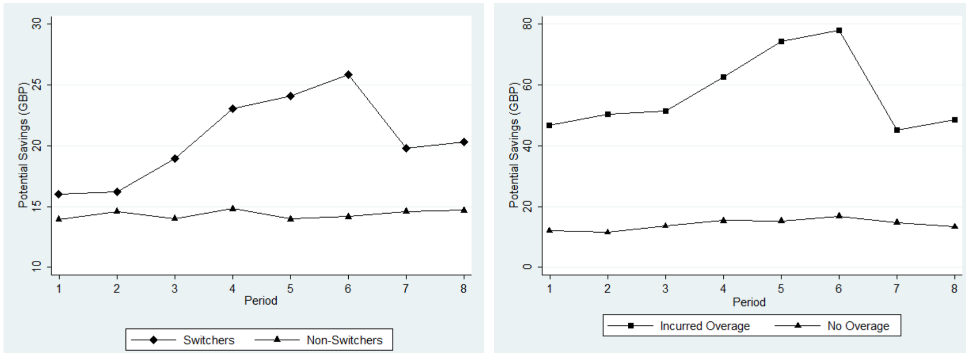

FIGURE 2.3: Potential savings from switching. Left-hand plot compares potential savings for switchers to non-switchers. Right-hand plot conditions on switching and compares potential savings for those incurring an overage at $t-1$ to those that were within allowances

fees then we would expect to see overspending increase in the size of the plan. In fact, whether we control for having switched or not, the amount of overspending is positively related to the tariff cost and negatively related to the size of plans.

Stylized Fact 5 - marginal price uncertainty

Under a three-part tariff the marginal price changes discontinuously at the allowance point. As is well known (e.g. Reiss and White (2005)), if consumers are aware of exact future and past usage there should be bunching of usage around the allowance. Figure 2.4 illustrates an absence of bunching in the data for all components. There are at least two possible explanations. First, consumers may be uncertain about future usage. Second, consumers may be responding to a price other than the marginal price, for example average or expected marginal price. Responding to average or expected marginal price seems natural when tariffs are complex and it is difficult to both keep track of usage as well as to retain information of kink points in the budget set. Borenstein (2009) suggests that California's multi-step electricity tariffs lead to consumers following rule-of-thumb techniques for electricity consumption, as confirmed by Ito (2012), who finds that these consumers are most likely responding

Table 2.7: Overspending

	(1) % Overspend	(2) % Overspend
Tariff Fee	1.120***	1.129***
	(0.00933)	(0.0106)
$\mathbb{1}(Medium_M)$	-0.665***	-0.780***
	(0.0986)	(0.105)
$\mathbb{1}(Medium_D)$	-0.711***	-0.820***
	(0.126)	(0.134)
$\mathbb{1}(Medium_{SMS})$	0.214	0.130
	(0.193)	(0.238)
$\mathbb{1}(Large_M)$	-3.695***	-3.476***
	(0.274)	(0.309)
$\mathbb{1}(Large_D)$	-2.485***	-2.024***
	(0.182)	(0.207)
$\mathbb{1}(Large_{SMS})$	-1.967***	-1.406***
	(0.271)	(0.315)
$\mathbb{1}(Switched)$		-1.731***
		(0.163)
Constant	5.539***	6.696***
	(0.325)	(0.375)
Household Fixed Effects	Yes	Yes
N	196407	158729
R^2	0.709	0.726

Robust standard errors in parentheses
*** $p < 0.01$

to the average price. Similar findings have been made in the tax literature (e.g. Chetty and Saez (2013) and Saez (2010))[9].

2.5 Model - Demand

The model is illustrated in Figure 2.5. Following an initial period (t_0) in which consumers observe their types, in all subsequent periods there are three stages: in

[9] In this literature it is often found that workers bunch around the first kink point in the income tax schedule, but not at higher kinks.

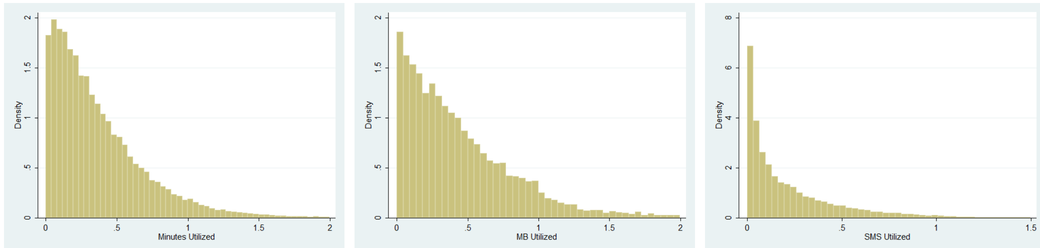

FIGURE 2.4: Densities represent the proportion of the allowance used in the month

stage one firms set the price-allowance-handset combinations; in stage two consumers make subscription decisions; and in stage three consumers make usage decisions for data, voice and text messages, conditional on plan choice. Each time period is one month. I solve the model by working backwards.

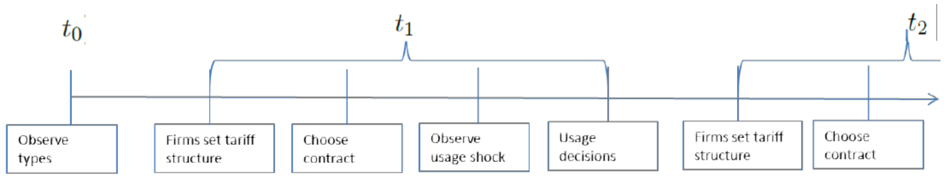

FIGURE 2.5: Timeline for Discrete/Continuous Choice

A multi-stage decision process for a consumer raises economic and econometric issues that need to be handled carefully. Economically we need to account for (at least) three effects. First, the distribution of usage shocks affects the expected cost of each tariff. Consumers therefore choose the tariff with the highest *expected* utility. Second, once usage shocks have been realized consumers decide the level of usage for each component. It seems likely that the structure of the tariff will affect usage decisions so that for a particular realization of the usage shocks different decisions are made on different tariffs, reflecting a price elasticity between mobile services and income. This is the analog of moral hazard in the insurance literature (see e.g. Einav et al. (2010), Einav et al. (2015) and Einav and Finkelstein (2011)).

The effect is that consumers can select into larger plans both because they have a high willingness to pay and because they have a lower sensitivity to the price of usage. Third, solving the optimal consumption path is a complicated problem. To the extent that consumers find it costly to track usage and/or forecast future usage shocks they may not respond to the dynamic incentives of the contract, i.e. is the relevant price to which consumers respond the static spot price, the dynamic shadow price or something else?

2.5.1 Usage Decision

During a month consumers use minutes, data and texts to solve tasks. Reflecting uncertain demand (Stlyized Fact 5) the number of potential tasks, K, is revealed to consumers subsequent to contract choice. K is assumed to be drawn from a Poisson distribution with mean $\lambda(X_{it})$, where X_{it} is a (potentially) time-varying set of consumer demographics. As tasks arrive consumers decide first, whether to solve the task and second, which component is optimal. Therefore, associated with each task is a pair (ω_i, y_i) where $\omega_i > 0$ is i's monetized disutility from not solving a task and y_i is the level of output required to solve the task (i.e. the location of the isoquant in (M,D,T)-space). These elements are assumed to be jointly distributed according to $F(y, \omega)$.

Reflecting inattentiveness to past usage consumers select a threshold at the start of the month for each component (ϕ_M, ϕ_D, ϕ_T) that represents the perceived cost of using each component, conditional on the contract choice. Following Ito (2012) suppose that consumers respond not to the true price schedule, but instead to $P(x+\epsilon)$ where $x = \{M, D, T\}$ is the volume of component x used and ϵ is the uncertainty over the amount of usage required to solve the expected number of tasks. Then the

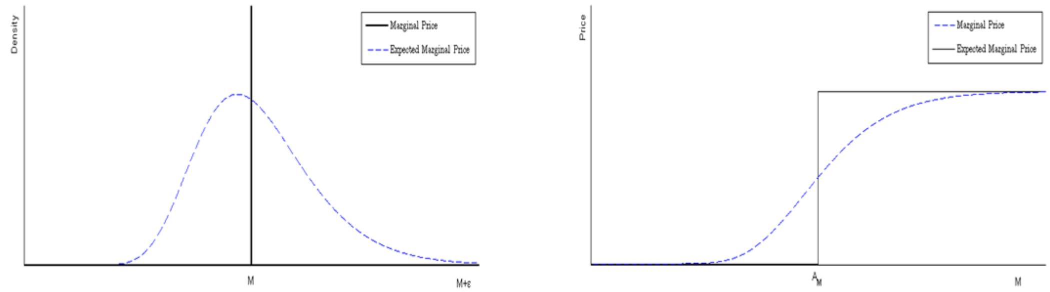

FIGURE 2.6: Uncertainty and Price Schedules

Notes: Right-hand plot illustrates two possible prices considered by the individual with uncertainty over usage distributed $LN(log(100), 0.25)$. The left-hand plot depicts the densities that convert the uncertainty into the price schedules. If consumers respond to the exact marginal price then the density is a mass point at actual usage. If instead they respond to expected marginal price then for each level of usage the price is calculated as $\phi_M = 0 * \Pr(M < A_M) + p_M * \Pr(M \geqslant A_M)$.

threshold is given by the expected marginal price,

$$\phi_x = \int P(x + \epsilon)w(\epsilon)\mathrm{d}\epsilon \tag{2.2}$$

where $w(\epsilon)$ represents the uncertainty about x. For example if consumers expect to use less than their allowance with probability one half then the expected marginal price is one half of the overage price. As all tariffs involve a maximum of one kink point with a zero marginal price up to the allowance and a constant marginal price, p_x, otherwise the threshold is given by $\phi_x = p_x * \Pr(x > A_x)$, where p_x is the overage price of component x. See Figure 2.6 for an illustration of how predicted usage will reflect the uncertainty in future usage, relative to the assumption of perfect certainty.

Individuals are assumed to respond optimally given their inattentiveness by minimizing the cost of solving the tasks that arrive during the month. Hence, conditional on tariff choice individual i solves the following problem for each task $k \in K$:

$$\begin{aligned} \tilde{c}(y_i, \phi; \Omega_U) &= \min_{M,D,T} \ \phi_M M_k + \phi_D D_k + \phi_T T_k \\ \text{s.t.} \ \ y_i &= (\alpha^i_{Mk} M_k + \alpha^i_{Dk} D_k + \alpha^i_{Tk} T_k)^\rho \end{aligned} \tag{2.3}$$

where Ω_U is a vector of parameters to be estimated. As the production function has a perfect substitutes specification a single component will be chosen for each task, the amount of which will depend on the efficiency parameters $(\alpha^i_{Mk}, \alpha^i_{Dk}, \alpha^i_{Tk})$ and the size of the task, y_i, given by:

$$M^*_k = y_i^{1/\rho}/\alpha^i_{Mk} \tag{2.4a}$$

$$D^*_k = y_i^{1/\rho}/\alpha^i_{Tk} \tag{2.4b}$$

$$T^*_k = y_i^{1/\rho}/\alpha^i_{Dk} \tag{2.4c}$$

Each of these optimal quantities has a corresponding cost of solving the task, $c(x^*_k) = \phi_x x^*_k$. In deciding whether to solve the task individual i compares the minimum cost, $\tilde{c}(y_k, \phi; \Omega_U) = \min\{c(M^*_k), c(D^*_k), c(T^*_k)\}$ to the disutility from not solving, ω_i. Therefore, conditional on task k arriving, flow cost is given by:

$$c^*(y_k, \omega_k, \phi; \Omega) = \begin{cases} \tilde{c}(y_k, \phi; \Omega) & \text{if task is solved} \\ \omega & \text{if task is not solved} \end{cases} \tag{2.5}$$

The monthly cost associated with tariff j at time t comes from summing over the cost of solving each task:

$$C^i_{jt}(\Omega_U, K^i) = \sum_{k=1}^{K} c^*(y_i, \omega_i, \phi; \Omega_U) \tag{2.6}$$

Finally, as consumers do not observe the number of tasks prior to tariff choice they form an expectation of monthly usage cost, $\mathbb{E}[C^*_{jt}(\Omega_U)]$, from using tariff j at time t by integrating equation (2.6) over F_λ.

Expected cost from usage enters the model similarly to that of medical expenditure in the model of Handel (2013). An important difference is that I need to allow for usage to depend on the tariff. In Handel's setting moral hazard is second-order with total out of pocket medical expenditure driven by exogenous health characteristics and probabilities of risk. Individuals with greater likelihood of incurring an accident will have higher expected out of pocket expenses leading them to select tariffs with lower deductibles (i.e. selection may be adverse). This is analogous to individuals with stronger preferences for using minutes to select into plans with large allowances for minutes. But whereas individuals with lower deductibles are unlikely to alter their probabilities of risk those with higher allowances are likely to choose to resolve more tasks. In other words, average usage is not invariant to the tariff choice; for the same number of tasks an individual will utilize different quantities as allowances and overage fees vary.

Parametric Assumptions

In order to estimate the model I need to make assumptions on the parametric forms of the distributions that effect consumers' usage decisions. Defining $G(\omega, y) \equiv G_2(\omega|y)G_1(y)$, I assume that:

$$\alpha^i_{xk} = \max(0, \bar{\alpha}_x + X_i'\mu + \Sigma_\alpha \nu^i_k) \quad x = \{M, D, T\} \tag{2.7}$$

$$\lambda_{it} = \gamma_1 + \gamma_2 Income_i + \gamma_3 Cars_i + \sum_{x=\{M,D,T\}} \gamma_x x_{t-1} \tag{2.8}$$

$$y_i = \bar{y} + \beta_1 Income_i + \beta_2 Age_i + \beta_3 \mathbb{1}(Homeowner_i) \tag{2.9}$$

$$\omega_i | y_i \sim \begin{cases} U[0, b] & w.p.\ p \\ b & w.p.\ (1-p) \end{cases} \tag{2.10}$$

where ν_k and are independent standard normal deviates, x_{t-1} is the quantity of component x used in the previous month and $b = c^*(y_i, \bar{\Omega})$ is the highest cost associated with y_i, i.e. when prices are at the overage levels. From equation (2.7) the efficiency parameter vector will be distributed with mean $\bar{\alpha} + X_i'\mu$ and covariance matrix $\Sigma\Sigma'$. Similarly the size of tasks (equation (2.9)) has a component common to all individuals, $\bar{y}$, as well as a portion that is influenced by an individual's income, age and homeowner status.

With these parametric forms and the assumed functional form for the production function the probability that task k is solved is given by:

$$\begin{aligned}
\Pr(solve|y_i;\Omega_U) &= \Pr(c^*(y_k;\Omega_U) < \omega_i) \\
&= \Pr(c^*(\Omega_U) < \frac{\omega_i}{y_i}) \\
&= 1 - G_2(c(\Omega_U)) \\
&= 1 - p\frac{c^*(\Omega_U)}{c^*(\bar{\Omega}_U)}
\end{aligned} \tag{2.11}$$

where the second equality follows from cost being linear in y_i. The lower is p the more likely the task will be solved, independent of cost. For a high value of p the task will only be solved if the cost to do so is sufficiently low. This intuition suggests that the distribution of ω can be identified by variation in $c(\Omega_U)$.

Individual Heterogeneity

There are several dimensions of individual heterogeneity that are introduced to allow the model to better fit the data. Some individuals will receive more tasks than others, and conditional on a task arriving there is likely heterogeneity in the size of a task and over individuals' tastes for using each of the components. These are captured, respectively, by λ, y and α.

It is useful to think about why we might observe heterogeneity in usage and how the model can rationalize this behavior. Variation in usage could be intertemporal, across individuals or both. As relative prices change a consumer would be expected to substitute between inputs, for fixed parameters on the production function and type-specific parameters. This variation can therefore help to identify the parameters of the production function, conditional on type.

More interesting heterogeneity comes from considering how individuals facing the same relative prices optimally choose different input demands. Is the observed variation due to consumers using the same production function to solve a different number and/or size of tasks (i.e. heterogeneity in(y_i, λ, p)), or do we all face similar tasks and yet have different preferences for how they should be solved (i.e. heterogeneity in α_x)? Both sound plausible. The model allows the data to reveal which, or both, of these competing theories is accurate.

Specifically, from equation (2.4) the optimal quantity of each component is influenced by y_i, but it has no affect as to *which* component is used. Instead, the choice of component is determined by the efficiency to performance ratio, ϕ/α_x, through equation (2.5).

2.5.2 Contract Discrete Choice

The previous section suggested consumers display downwards inertia in contract choice (Stylized Fact 1) increasing allowances when incurring an overage but rarely decreasing. To allow for the probability of switching to be correlated with past usage I adopt the consideration set model of Ching et al. (2009). After observing the previous month's bill consumers make an active choice with probability P_R, where:

$$P_R = \begin{cases} 1 & \text{if } M_{jt-1} > A_{jMt-1} \cup D_{jt-1} > A_{jDt-1} \cup T_{jt-1} > A_{Tjt-1} \\ a & \text{otherwise,} \end{cases} \tag{2.12}$$

Conditional on the distribution of the number of tasks the expected utility to consumer i from subscribing to contract $j \in J_t$ with carrier $c \in C$ at time t is given by:

$$\begin{aligned} V_{ijct} = \beta_{ijc} - \alpha_i(\mathbb{E}[C^*_{jct}] + F_{jct}) + \gamma \mathbb{I}\{s_{it} = jc\} + z'_{jct}\psi + \epsilon_{ijct} \\ \alpha_i = \alpha + D'_i\Gamma + \Sigma_C \nu^p_{it} \end{aligned} \tag{2.13}$$

where $\mathbb{E}[C^*_{jct}]$ is the expected cost discussed above, F_{jct} is the monthly contract fee, α_i is i's marginal utility of income that is allowed to depend on demographics and z_{jct} are other observed contract characteristics with tastes ψ. The consumer's state variable $s_{it} \in \{0, 1, ..., J_{t-1}\}$ contains the history of past contract choices, so that if the consumer was on contract k with carrier c in the previous month then $s_{it} = kc$. ϵ_{ijct} are unobserved (to the econometrician) components of utility that affect individual-tariff match, e.g. the set of available handsets. I assume that ϵ_{ijct} is distributed i.i.d. type I Extreme Value.

A finding of $\gamma > 0$ is suggestive of inertia in contract choice and is included to capture the presence of gains from switching (Stylized Fact 2). Conditional on subscribing to contract jc in the previous month the probability of repeating the choice of jc in the current month is higher than the marginal choice probability. In other words two otherwise identical individuals, one who chose jc in the previous month and one who did not would behave differently in the current period. Following Heckman (1981) past choices directly influencing current period choice probabilities is considered to be true (structural) state dependence. However, it is important to note that an observationally equivalent explanation for repeat purchases is that they are driven by preference heterogeneity, correlated over time. Failure to control for this spurious state dependence thus introduces upwards bias into γ. To disentangle these competing explanations I include contract intercepts, β_{ijc} that are allowed to vary by individual.

Modeling inertia as affecting utility additively warrants further discussion. This assumes that consumers are myopic and do not consider the "lock-in" effects of their decisions on future choice probabilities[10]. This assumption is largely made for convenience, but is considered to be reasonable: to solve the fully-dynamic problem consumers would need to forecast future prices, allowances, handset innovation and changes to network quality which would be arduous[11].

[10] e.g. as in the rational addiction model of Becker and Murphy (1988).

[11] Hartmann and Viard (2008) allow for switching costs to incorporate future lock-in, but in a simpler setting than the one considered here. In their setting consumers can perfectly predict future prices and the period of estimation is such that the choice set is plausibly exogenous.

Due to data limitations I do not observe consumers switch to rival carriers; instead they exit the data set. Mobile penetration rates averaged more than 120%[12] in the UK during the sample period. This includes individuals on long-term contracts and those using pay-as-you-go SIM cards. The interpretation of the outside good therefore requires attention. A consumer who exits from the data set is assumed to have entered into a contract with a rival firm, or to have switched to a pay-as-you-go SIM. The utility of the latter option is normalized to zero, i.e. $U_{i0t} = \epsilon_{i0t}$. The probability of leaving the current provider can therefore be thought of as the sum of the choice probabilities for all rival tariffs in the consumer's choice set at time t plus the probability of taking the outside option.

The probability that a consumer chooses contract j with carrier c is given by the product of remaining with carrier c and choosing j given carrier c is preferred. Define $\delta_{jct} = \beta_{ijc} + \gamma\mathbb{I}\{s_{it} = jc\} + z'_{jct}\psi - \alpha_i(F_{jct} + \mathbb{E}[C^*_{jct}])$, then, given the distributional assumption on ϵ_{ijct}, the probability that consumer i who makes an active choice chooses to remain with carrier c at time t is:

$$P_{it}(c|R) = \frac{\sum_{k \in Jc} exp(\delta_{jct})}{\sum_{c' \in C} \sum_{k \in Jc'} \exp(\delta_{kc't})} \tag{2.14}$$

where R denotes an active choice. Similarly, the probability of choosing contract jc, conditional on choosing carrier c, is given by:

$$P_{it}(jc|R) = \frac{\exp(\delta_{jct})}{\sum_{k \in Jc} \exp(\delta_{kct})} \tag{2.15}$$

Given the distribution for making active choices, the unconditional probability that a consumer chooses plan jc in period t is:

$$P_{it}(jc) = \begin{cases} P_R P_{it}(c|R) P_{it}(jc|R) & \text{if } s_{it} \neq jc \\ P_R P_{it}(c|R) P_{it}(jc|R) + (1 - P_R) & \text{if } s_{it} = jc \end{cases} \tag{2.16}$$

[12] Source: OfCom (2013)

where $P_{it}(jc|R)$ is the probability of choosing tariff j conditional on making an active choice, from equation 2.15. Explicitly, the probability that consumer i with $s_{it} = jc$ chooses jc is given by $P_{it}(c|R)P_{it}(jc|R)$ when an overage was previously incurred, and $aP_{it}(c|R)P_{it}(jc|R) + (1 - a)$ otherwise.

2.6 Estimation

2.6.1 Estimation of Usage Parameters

The inner loop represents the usage model (from equation (2.3)) which is estimated using simulated method of moments (see, e.g. Pakes and Pollard (1989), McFadden (1989)). The model predicts usage of minutes, data and SMS for every individual in every period as a function of observed individual characteristics, unobserved random tastes and usage shocks and the vector of parameters to be estimated.:

$$\begin{aligned} \mathbb{E}[Q_{it}] = \sum_{J_{it}=0}^{\infty} \sum_{j=0}^{J_i t} \int \int \int Q_j(X_i, \Theta_U, \alpha) f(\alpha|X_i) p(J|X_i) \partial\alpha \\ Q_{it} = [M_{it}, D_{it}, SMS_{it}] \end{aligned} \tag{2.17}$$

That is predicted usage is given by summing over task-level usage for a specific number of tasks, J_{it}, then weighting by the probability that J_{it} tasks were realized. Defining the prediction error as:

$$\epsilon_{it}(X_{it}, \Theta_U) = \mathbb{E}[Q_{it}] - Q_{it} \tag{2.18}$$

where Q_{it} is the observed usage levels, then at the true parameter vector Θ_U^0,

$$\mathbb{E}[\epsilon_{it}|X_{it}, \Theta_U^0] = 0 \;\text{ for }\; i = 1, 2, ..., I \;\text{ and }\; t = 1, 2, ..., T \tag{2.19}$$

Using equation (2.19) I can construct conditional moments, $Z_{it} = X_{it} \otimes I_3$, where X_{it} are observed demographic and usage data and I_3 is a 3-dimensional identity matrix, representing the three usage components. Following Hansen (1982) these Z_{it} are independent of the unobservables:

$$\mathbb{E}[Z_{it} * \epsilon_{it} | Z_{it}] = 0 \tag{2.20}$$

Estimation uses the sample analogs of these moments:

$$g(X_{it}, \Theta_U) = \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=1}^{T} Z_{it} * \epsilon_{it} \tag{2.21}$$

to construct the objective function over which I search for Θ_U:

$$J(\Theta_U) = [g(X_{it}, \Theta_U)' W g(X_{it}, \Theta_U)] \tag{2.22}$$

where W is the efficient weighting matrix (the inverse of the asymptotic variance of g - see Hansen). The estimates have the following asymptotic distribution:

$$\sqrt{IT}(\Theta_U - \Theta_0) \sim N(0, \Omega) \text{ with } \Omega = \left(\left(\frac{\partial g(\Theta_0)}{\partial \Theta} \right)' W \left(\frac{\partial g(\Theta_0)}{\partial \Theta} \right) \right)^{-1} \tag{2.23}$$

Estimation is performed by simulating equation (2.17) to substitute for the complex multi-dimensional integrals. For each individual-month I take R draws from the Exponential distribution (these are required to simulate the Poisson task arrival, see appendix). For each of these draws I take $(3 \times K)$ standard normals, being one for each efficiency parameter and where K is a number large enough to bound the maximum of tasks. The random draws are substituted into the equations governing optimal behavior (equations (**??**) and (2.5)) and parametric assumptions (equations (2.7), (2.8), (2.9) and (2.10)) to generate predicted usage in each month for each simulation. I proxy equation (2.17) by averaging predicted usage over the simulation draws to generate expected behavior conditional on X_{it}. In estimation I use $R = 30$ simulations.

2.6.2 Estimation of Tariff-Demand Parameters

On the outer loop individuals choose their tariff conditional on their expected cost of subsequent usage. Given the assumed distribution of ϵ_{ijct} estimation is fairly straightforward. Recall that the probability that consumer i selects plan j from carrier c at time t is given by:

$$P_{it}(jc) = \begin{cases} P_R P_{it}(c|R) P_{it}(jc|R) & \text{if } s_{it} \neq jc \\ P_R P_{it}(c|R) P_{it}(jc|R) + (1 - P_R) & \text{if } s_{it} = jc \end{cases} \tag{2.24}$$

where $P_{it}(jc|R)$ is the probability of choosing tariff j conditional on making an active choice, from equation (2.15). Explicitly, the probability that consumer i with $s_{it} = jc$ chooses jc is given by $P_{it}(c|R)P_{it}(jc|R)$ when an overage was previously incurred, and $aP_{it}(c|R)P_{it}(jc|R) + (1 - a)$ otherwise.

An individual's contribution to the likelihood is given by the sequence of observed choices:

$$L_i(\Theta_C) = \prod_{t=1}^{T} P_{it}(jc) \tag{2.25}$$

with estimation based on maximizing the log-likelihood:

$$LL(\Theta_C) = \sum_{i=1}^{I} \ln(L_i(\Theta_C)) \tag{2.26}$$

There are three complications involved with maximizing equation (2.26): the expected cost of being on a contract is not observed, I do not observe the initial choice made by an individual and heterogeneity in choice parameters (price and contract dummies coefficients) means that the likelihood does not have a closed-form solution. To resolve the first problem the simulated expected cost from the inner loop is used as a variable in the choice decision. This assumes that the simulated usage cost equals the true *ex ante* expected cost.[13]

The initial conditions problem (see, e.g. Heckman (1981)) creates two issues. First, I cannot construct the state variable for the first period that I observe each individual. Second, even if I did observe this choice it would be correlated with the brand intercept. This would, potentially, render the parameters inconsistent by confounding preference heterogeneity with structural state dependence. I address this problem in two ways. I hold out the first month's ($t = 0$) purchase history for each individual and use this to construct the state at period $t = 1$. Second, I integrate over the joint distribution of the state variable and brand intercepts using a procedure suggested by Wooldridge (2010). That is, I allow for correlation between the state variable and contract intercepts by explicitly modeling the intercepts as:

$$\beta_{ijc} = \beta_{jc} + \beta_{GL} \mathbb{1}\{s_{i0} = jc\} + \sigma_\beta \epsilon_{ij} \tag{2.27}$$

where $\mathbb{1}\{s_{i0} = jc\}$ equals 1 if contract jc was chosen in the initial period andϵ_{ij} is assumed to be distributed standard normal.

The third problem is resolved by reverting to simulated maximum likelihood see Train (2009)). For a given expected cost S draws are taken for each individual from the distributions of each random coefficient (α_i and β_{ijc}). These random terms are substituted into expected utility (see equation (2.13)) for each individual-period with choice probabilities constructed using (2.24). For each draw the optimal sequence of choices is constructed with the average of these choice probabilities being used in place of L_i in equation (2.26). This approach implicitly incorporates the effects of inertia by conditioning on past choices. If the likelihood maximizing set of parameters is $\hat{\Theta}_C$ (which includes the switching cost $\hat{\gamma}$) then the choice in period t is optimal given $\hat{\Theta}_C$ and the implied effects of inertia from period $t - 1$. In estimation I use $S = 50$ simulations.

[13] An alternative would be to integrate over the shocks explicitly (e.g. using Simpson quadrature), but given the multidimensionality of the integral I revert to simulation.

2.7 Identification

There are two sets of parameters to be estimated: (1) preferences for usage of bundle components Ω_U and (2) preferences for tariffs, to include switching costs Ω_C. The main challenge for identifying usage parameters comes from the fact that we observe the joint distribution of the number of tasks and usage per task and not task-level usage. Identification comes from considering the distinction between an individual using large amounts of one component and using small amounts of all three components. As the random tastes in the efficiency parameters are independent across tasks an individual with many tasks will tend to use some of each component, whereas an individual with few tasks will tend to use only one component.

The main challenge in estimating contract choice parameters comes from separately identifying the potential sources of persistence in choices that is observed in the data. Is it that consumers make repeat purchases of the same contract because of high switching costs, individual preference heterogeneity that is correlated over time or something else? The problem is addressed by using individual-brand intercepts that are correlated with initial choices, as discussed above. It is noted that bias will remain in γ_i if the specification fails to account for the distribution of individual heterogeneity or if there is serial correlation in ϵ_{ijct} (e.g. if the network used by family and friends affects tariff utility this would be expected to introduce upward bias into γ).

2.7.1 Identification of Usage Parameters

Predicted inputs required to solve each task are given by equation (2.4), conditional on the parameters affecting the size of tasks and efficiency of each component. Given the assumption of perfect substitutes in the production function only one component will be used for each task, with the choice governed by the minimum cost to solve (see equation (2.5)). This means that the size of task, y_i, affects the number of units chosen (through equation (2.4)), but it has no influence on *which* component is used. The choice of component is influenced (through equation (2.5)) by the price to efficiency ratios ϕ_x/X. This allows us to separately identify the size and efficiency parameters.

Identification of the joint distribution of tasks and usage comes from considering the role demographics have in the size of tasks, the mean of the Poisson process and the disutility of not solving. As income enters both λ_i and y_i it can help distinguish between having many consumption occasions and utilizing a large amount of a component for a given task. Similarly, allowing age to enter the efficiency parameters and y_i distinguishes between the amount utilized and the component used.

Finally, some tasks that arrive will not be solved (the outside good). This occurs when the disutility from not solving is lower than the minimum cost to solve (equation (2.5)). Given identification of the size of tasks, y_i, efficiency parameters, α_{xi} and arrival of tasks, λ_i variation in cost to solve each task allows us to identify the distribution of ω. To see this, the probability of solving a task is given by:

$$\begin{aligned} \Pr(solve|type, y) &= \Pr(c(y;\Omega_U) < \omega|type) \\ &= \Pr(c(\Omega_U) < \frac{\omega}{y}|type) \\ &= 1 - G_2(c(\Omega_U)) \end{aligned} \tag{2.28}$$

where the second line follows from the cost function being linear in y. Variation in $c(\Omega_U)$ trace out the distribution of ω, i.e. it's identified on the support of $c(\Omega_U)$. The idea is that with repeated observations for the same individual and the assumption of constant types variation in $c(U_\Omega)$ affects solving probabilities to identify $G_2(\omega|y)$. To be more explicit given the assumed parametric form, we have that the task is solved with probability $1 - p\frac{c(\Omega_U)}{c(\Omega_U)}$. The lower is p the more likely the task will be solved, independent of the cost. For a high value of p the task will only be solved if the cost to do so is sufficiently low. In the data I observe the number of minutes, MB of data and number of SMS that were used to solve a task, conditional on the task arriving *and* the cost being lower than the disutility of not solving. This joint distribution needs to disentangle the probability of a task arriving, λ, the location of the isoquant, $G_1(y)$, the probability of solving the task and the disutility distribution, $G_2(\omega|y)$. If tasks are frequently solved there are two possibilities:

1. p is low and hence the probability of a task being solved should be invariant to $c(\Omega_U)$: $\frac{\partial \Pr(solve)}{\partial c} \approx 0 \implies p \approx 0$

2. p is high and we only solve tasks when $c(\cdot)$ is low: $\frac{\partial \Pr(solve)}{\partial c} < 0 \implies p \to 1$

hence, variation in $c(\Omega_U)$, which is at the task level, identifies ω.

2.7.2 Identification of Tariff-Demand Parameters

Identification of Inertia

The challenge is to identify the proportion that consider switching when *not* incurring an overage, a, and the distribution of switching costs, γ. Consider first a consumer who exceeded an allowance in the previous month and hence makes an active choice in the current period. The extent to which the consumer elects not to switch away from a dominated plan identifies the distribution of switching costs, γ. Now consider a consumer who remained within each of their allowances. With probability a they make an active decision to evaluate alternative plans. Given the distribution of γ this parameter is identified by the rate at which consumers within their allowance switched plans. A key assumption therefore is that the distribution of switching costs is independent of a. Without this assumption it would not be possible to separately identify switching costs from attentiveness for those consumers within their allowance[14].

[14] To be explicit, observing a consumer remain on a dominated plan due to a high γ or due to a low a would be observationally equivalent.

Following Osborne (2011) and Dubé et al. (2010) a second source of identification comes from considering the effect of a change in period $t-1$ exogenous variables on period t's choice probabilities. For example, consider a period $t-1$ price reduction for a tariff from network 1 with allowances of 300 minutes, 500MB data and 1000 text messages. If this price cut increases the probability of purchasing a similar contract at period t then $\gamma > 0$. Sufficient variation in exogenous variables that affect tariff utility therefore provides an additional source of identification.

Identification of Other Tariff-Demand Parameters

Identification of the remaining parameters is standard. One potential problem is that the tariff fee, F_{jt}, may be correlated with the error term, ϵ_{ijt}, say due to unobserved quality or promotional activity. As unobserved quality is likely at the carrier-, as opposed to contract-level carrier fixed effects mitigate for endogeneity caused by correlation between F_{jt} and unobserved quality.

To correct for the potential bias that may remain a control function approach will be used (see Petrin and Train (2010)). Decomposing the error term into the part that is correlated with the tariff fee, $\tilde{\epsilon}_{ijt}$, and that which is uncorrelated, μ_{ijt}, as $\epsilon_{ijt} = \rho\mu_{ijt} + \tilde{\epsilon}_{ijt}$, then providing I can recover μ_{ijt} I can consistently estimate contract choice probabilities. The identifying assumption is that tariff fees and taste shocks are independent, conditional on μ. An estimate for μ is found by regressing price on product and market characteristics:

$$F_{jt} = \beta_0 + \beta_x \sum_x A_{xj} + x_{jt-1}\delta_1 + \mu_{jt} \tag{2.29}$$

Own plan allowances, A_{xj}, are included to proxy for costs of offering plans. x_{jt-1} is a vector of lagged market characteristics. These instruments are similar to those proposed by Hausman (1997), except that I use intertemporal instead of cross-market variation in market characteristics. This is necessitated by carriers using national-level pricing. Included in x_{jt-1} is the average size of rival allowances, average rival tariff fee and the number of rival plans that offer unlimited allowances. The assumption is that cost shocks (that influence tariff fees and allowances) are serially correlated, allowing lagged characteristics to affect current period fees, but that these are uncorrelated with promotional activity in the current period. Once $\hat{\mu}_{jt}$ is recovered from equation (2.29) it is then included in the expected utility function:

$$V_{ijct} = \beta_{ijc} - \alpha_i(\mathbb{E}[C^*_{jct}] + F_{jct}) + \gamma \mathbb{I}\{s_{it} = jc\} + z'_{jct}\psi + \rho\hat{\mu}_{jt} + \tilde{\epsilon}_{ijct} \qquad (2.30)$$

Standard errors are adjusted using bootstrapping to take account of the additional variance introduced by inserting an estimate of μ_{jt} into (2.30).

2.8 Results

The results for the usage decision are reported in Table 2.8 and for the contract decision in Table 2.9. The average number of tasks per consumer-month is 11.76, with an average size per task of 1.55. As we would expect the number of tasks arriving is increasing in usage in the previous month of minutes, data and SMS. There is significant heterogeneity in the efficiency parameters so that whilst the three components have similar mean values the standard deviations are as large. This suggests consumers, on average, view the three components similarly, but that the efficiency of a particular component for a particular task can vary significantly. This is to be expected. Some tasks are more amenable to voice (e.g. making a regular call to a parent), others are better suited for using data (e.g. finding out the sports scores), whereas for others a text message is likely preferred (e.g. arranging where and when to meet a friend after work).

In the contract choice results we can see the benefit of including the control function. With the control function the average price coefficient is -0.128, whereas without it the price coefficient is positive, 0.28. A positive price coefficient contradicts

Table 2.8: Usage Choice Results

Efficiency Parameters	Coeff.	Task Size/Arrival	Coeff.
$\bar{\alpha}_M$	0.1768***	$\bar{y}$	0.2311***
$\bar{\alpha}_D$	0.2002***	$\beta_1 : Income$	0.2915***
$\bar{\alpha}_{SMS}$	0.1627***	$\beta_2 : Cars$	0.2481***
$\mu : Age$	−0.0613***	$\beta_3 : \mathbb{1}(Homeowner)$	0.1548***
σ_{α_M}	0.2451***	$\lambda_1 : Age$	0.6832***
σ_{α_D}	0.1337***	$\lambda_2 : Income$	0.1162***
$\sigma_{\alpha_{SMS}}$	0.2186***	$\lambda_3 : Cars$	0.5916***
$\rho : curvature$	0.7891***	$\lambda_M : M_{t-1}$	0.3586***
$p : Pr(Solve)$	0.1062***	$\lambda_D : D_{t-1}$	0.2714***
		$\lambda_{SMS} : SMS_{t-1}$	0.3637***

the expectation of downward-sloping demand and renders the model implausible for welfare analysis and predictive purposes. The residuals enter significantly and with the expected sign. Positive residuals suggest that the price of the product is higher than the observed attributes would suggest, hence a contract has desirable attributes that the model fails to capture. These (unobserved) attributes should therefore have a positive effect on utility as represented by the coefficient on the residuals (0.4328).

Income interacts positively with price suggesting individuals with more income are less price sensitive, which agrees with intuition. The average switching cost is 12.13 (utils). Although income enters positively into the switching cost, which could represent increased time value of money for the hassle cost of contacting the carrier, it, like the amount of overspend, are not statistically different from zero. The level of switching cost corresponds to £94.8 (derived by dividing the average switching cost by the price coefficient). This is similar to industry reported figures for how much consumers are leaving on the table by sticking with their current plans. Also in line with industry research the probability of making an active decision is 0.197,

Table 2.9: Contract Choice Results

	CF	No CF		CF	No CF
$\bar{\alpha}$	-0.1544^{***}	0.1521	$\bar{\beta}_1$	15.887^{***}	6.803^{***}
	(0.0266)	(0.5532)			
$\alpha_1 : Income$	0.0060^{**}	0.0293	$\bar{\beta}_2$	16.2178^{***}	7.860^{***}
	(0.0030)	(0.1324)			
σ_α	0.0072	0.0852^{***}	$\bar{\beta}_3$	17.178^{***}	7.463^{***}
	(0.0062)	(0.0203)			
β_{GL}	-3.2542^{***}	-7.897^{***}	$\bar{\beta}_4$	10.3282^{***}	10.850^{***}
	(0.2377)	(1.1194)			
γ_1	11.288^{***}	9.690^{***}	$\bar{\beta}_5$	17.0196^{***}	17.151^{***}
	(0.7205)	(1.228)			
$\gamma_2 : Income$	0.1913	3.074^{***}	$\Pr(active)$	0.197^{***}	0.253^{***}
	(0.1548)	(0.304)		(0.057)	(0.0437)
$\gamma_3 : Overspend$	0.0013	0.368^{***}			
	(0.0032)	(0.0123)			
$\mu : CF$	0.4328^{***}	-			
	(0.0255)	-			

suggesting consumers re-evaluate their plans roughly every six months. Finally, carriers are estimated to be of fairly homogeneous quality.

2.9 Counterfactuals

The counterfactuals attempt to quantify the gains that could be made from consumers being better matched with plans. One alternative that has been pursued elsewhere (e.g. Handel (2013)) is to evaluate the effects of removing switching costs on contract choices. Consumers are assumed to always evaluate all plans, but the presence of a switching cost makes the probability of remaining with a current plan higher than it would be for an identical individual who was not locked-in by his previous choice. Removing switching costs from the choice model eliminates this distortion, flattening out the demand curve. Assuming the quality and number of plans offered and the marginal cost of each plan remains fixed consumers will be unambiguously better off.

However, in my environment such a policy would be difficult to implement. Consumers make sticky choices due both to being inattentive and facing switching/hassle costs of moving plans. With estimated average switching costs of £95 and low attentiveness (19.7% consider switching in any period) the standard mechanisms for correcting these distortions (e.g. a monthly notification campaign to improve attentiveness, improved customer service to reduce the hassle cost of switching and the removal of early termination charges to reduce the monetary cost of switching) are likely prohibitive. Instead, I consider a counterfactual where firms are required to re-optimize for consumers each month, placing them on the lowest cost plan, *ex ante*.[15] With only five carriers this policy would be straightforward to implement and is similar to a proposal made for the energy market in the UK.[16]

In the second counterfactual instead of eliminating switching costs and making consumers fully attentive I consider a range of values for $\tilde{a}$ and γ to disentangle how the two sources of inertia impact on welfare. Given that the appropriate policy response depends on which type of inertia is present it is interesting to see how a and γ interact to affect welfare. This counterfactual can be seen to inform the regulator of the gains that can be made from different policies that reduce inertia. The optimal degree of intervention can then be determined by comparing these gains to the cost of implementation. These costs could easily be obtained by the regulator.

It should be noted that in both counterfactuals prices and plans are held constant. Calculating optimal counterfactual prices with full attention and no switching costs is straightforward as the solution to a set of static first order conditions. However,

[15] This is similar to a counterfactual analyzed by Ho et al. (2015) whereby an intermediary (pharmacist) reallocates individuals on to lower cost Medicare Part D plans.

[16] See http://www.theguardian.com/business/2012/nov/20/energy-companies-customers-cheapest-tariff.

deriving optimal prices in the baseline and the counterfactual with some inertia necessitates solving a dynamic oligopoly pricing game that is beyond the scope of the paper. Whether prices fall or rise once switching costs are removed is ambiguous. Even if firms' costs remained constant (i.e. the cost of a plan doesn't depend on the set of individuals on the plan. In other words there is no adverse selection) the direction of any price change would depend on which of the "harvest" or "invest" strategies dominated in the presence of switching costs.

2.9.1 Firm-led re-optimization

In the first counterfactual at the beginning of each month (before usage shocks have been realized) firms reallocate consumers on to the lowest (expected) cost plan. Choices are predicted by using the demand estimates of section 2.8 and simulating random terms for usage and contract choice. In the baseline consumers search with probability a, whereas in the counterfactual the firm re-optimizes for all individuals (equivalent to $a = 1$).

It is important to consider whether switching costs should be incorporated into welfare calculations. γ, the switching cost parameter enters positively into the utility associated with an individual's state plan, increasing the likelihood that the plan is chosen (relative to an identical first-time buyer). Whereas some sources of inertia reflect a tangible social cost that is avoided when not switching and hence naturally should be included in a welfare calculation, others, such as biased beliefs should be excluded (see Farrell and Klemperer (2007) and Klemperer (1995) for excellent summaries on the sources of switching costs). For the baseline welfare calculations I therefore allow for switching costs to be $\kappa\gamma$ where κ is between 0 and 1. This term, $\kappa\gamma$, relates to that part of switching cost that is a tangible social cost. Although the full switching cost is allowed to affect choice probabilities, the welfare calculation will be based on $\kappa\gamma$ in the baseline and zero in the counterfactual.

Given that the error term is T1EV the change in consumer surplus is given by the difference in the log-sum of the indirect utilities under the baseline and counterfactual environments (see Small and Rosen (1981)):

$$\Delta W_{it}(\hat{a}, \hat{\gamma}\kappa, 0, 0) =$$

$$\int \frac{1}{\alpha_i}[\ln\Big(\sum_{j \in J_t^a} \exp\{(\beta_{ijc} - \alpha_i(\mathbb{E}[C^*_{jct}] + F_{jct}) + \gamma\kappa\mathbb{1}\{s_{it} = jc\} + z'_{jct}\Psi\Big)\} \tag{2.31}$$
$$-\ln\Big(\sum_{j \in J_t} \exp\{(\beta_{ijc} - \alpha_i(\mathbb{E}[C^*_{jct}] + F_{jct}) + z'_{jct}\Psi\Big)\}]dG(a)dF(\nu)$$

where the choice set in the baseline, J_t^a, is equal to all products of all firms when the consumer is attentive (which occurs with probability a), otherwise it is the current plan. ν is all individual random terms that affect choices, i.e. the random coefficients on price, switching costs and carrier intercepts.

In order to calculate (2.31) I simulate whether a consumer is attentive from a Bernoulli distribution with parameter $\hat{a}$ (in addition to simulating the random terms, ν). For those with $a_c = 0$ the choice set is the current plan, whereas those with $a_c = 1$ choose from among all plans on all carriers. In either case choice probabilities include switching costs of γ (it is only for welfare that switching costs are $\kappa\gamma$). In the counterfactual those with $a = 1$ continue to choose from among all plans on all carriers whereas those with $a = 0$ are reallocated within their current carrier meaning consumer surplus is calculated by summing over all of carrier c's plans.

In the baseline and counterfactual scenarios the strong degree of path dependence requires me to simulate a sequence of choices for each individual (to analyze all possible paths would be computationally prohibitive). I create 10 such sequences, calculate consumer surplus for each sequence and average over the set to calculate the effect on welfare.

2.9.2 *Consumer nudging*

In this counterfactual welfare is calculated by simulating choices under the baseline and counterfactual scenarios. As in the first counterfactual the strong degree of path dependence requires me to simulate a sequence of choices for each individual in both the baseline and counterfactual scenarios. (It is only in the counterfactual when switching costs are zero and consumers are fully attentive that the decision is static and I do not need to simulate a sequence of choices.) Explicitly, in the baseline for each consumer in each month I draw a consumer's attentiveness from a Bernoulli with mean $\hat{a}$. Those with $a = 1$ choose from among all plans on all carriers with switching cost of γ. Those with $a = 0$ remain on the same plan. In the counterfactual for each consumer in each month I draw a consumer's attentiveness from a Bernoulli with parameter $\tilde{a} \in [0, 1]$. Those with $\tilde{a} = 1$ search among all plans on all carriers incurring γZ to switch plans. Those with $\tilde{a} = 0$ remain on the same plan. I create 10 such sequences for each individual and average over the consumer surplus to calculate the effect on welfare for each $(\tilde{a}, Z)$ pair.

2.9.3 *Counterfactual Results*

Firm-led re-optimization

In the first counterfactual those consumers that make an active decision choose from among all of the available plans, across all carriers. These, attentive, consumers achieve the same consumer surplus in the benchmark and counterfactual scenarios as they do not benefit from the firm re-optimizing on their behalf. The larger set of inattentive consumers are reallocated on to the lowest cost plan by the carrier at the beginning of the month. For these consumers the welfare change is given by comparing consumer surplus from among all plans within carrier to that achieved from when the choice set is only the current plan.

The results are reported in Table 2.10 with $\kappa = 0.5$. The average change in consumer surplus is £10.6 per individual-month with 39% of consumers being reallocated on to a cheaper contract each month. This is equivalent to approximately £200m per month when extrapolating to the population of 20m consumers that are on fixed-term contracts. In the baseline, of the 20% of attentive consumers 70% remain on the same plan, less than 1% switch within carrier and 29% switch to a rival carrier. In the counterfactual the amount of switching is similar to the baseline with 25.7% switching across carrier and only 0.1% switching within carrier. This is to be expected. For those that are attentive and were reallocated in a previous period the switching cost has the same effect (albeit on a different state contract) as when they had not been reallocated. The slight decrease in the proportion switching can be rationalized by the removal of the brand loyalty effect. As this enters negatively into utility it induces consumers to switch away from this contract more than from any other contract. With consumers being reallocated away from their initial state contract the reduction in utility from the brand loyalty effect is removed, making other state contracts more attractive, hence reducing the degree of subsequent switching.

Table 2.10: Counterfactual 1: Firm-led re-optimization

	Baseline	Counterfactual	Δ
$\Delta(CS)$			10.27
% Reallocated ($\tilde{a} = 0$)	0	39.2	39.2
% Switched within firm ($\tilde{a} = 1$)	0.5	0.1	-0.4
% Switched across firms ($\tilde{a} = 1$)	29.5	25.7	-3.8

Consumer nudging

In the previous counterfactual firms reallocated individuals on to the cheapest plan. In this counterfactual I consider an alternative where consumers are encouraged to switch due to reduced switching costs and/or being nudged to being more attentive.

Table 2.11 reports the change in consumer surplus from considering various values of Z and a where $Z \in [0, 1]$ is the proportion of switching costs removed and Table 2.12 reports the proportion of consumers that switch under the counterfactual.

Whereas in the first counterfactual those that switched did not benefit from firm-led re-optimization (i.e. the change in consumer surplus was derived only from the set that did not switch) in this counterfactual the set of switchers are the only consumers to benefit. As the fraction that searches increases consumer surplus unambiguously increases as consumers reallocate on to better matched plans. However, the reduction in switching cost trades off two effects. As Z decreases consumers are more likely to switch on to better plans, thus increasing consumer surplus. But working against this is that for low switching cost values utility from the current plan is reduced, vis a vis the baseline scenario. So whereas the degree of attention expands the choice set, increasing consumer surplus, removing switching costs tends to decrease welfare as the choice set is unaffected, but the utility from the state plan is reduced. This, perverse, result is driven by switching costs entering utility additively and consumer surplus being calculated by summing over the utilities of all products in the choice set. To mitigate the effects of this result Table 2.11 reports the change in consumer surplus when $\kappa = 0.1$. As before choice probabilities are calculated with $\kappa = 1$; it is only welfare that is calculated with the lower value for the switching cost.

Reflecting the impact of how switching costs enter utility, the gains to increasing attention are far greater than that from reducing switching costs. For a given level of inattention reducing switching costs increases the proportion switching by 5-20%. Whilst this increases the match of consumers with plans it does not expand the choice set and hence has little-to-negative effect on consumer surplus. Increasing attention on the other hand results in a significant increase in consumer surplus. With fully attentive consumers and switching costs at their estimated level the average gain to consumers is £30.7 per individual-month. This equates to 120% of the monthly

Table 2.11: Counterfactual 2: consumer nudging - consumer surplus

$\Delta(CS)$	Z				
	0	0.25	0.5	0.75	1
Attention					
0	-14.1	-12.4	-10.6	-8.9	-7.2
0.19	-0.81	0.29	1.54	2.40	0
0.5	13.62	15.12	15.84	15.96	16.28
0.75	22.57	23.25	23.90	23.65	24.73
1	30.07	30.22	30.38	30.55	30.76

tariff cost and is approximately £600m for the full population of those on fixed-term contracts.

Table 2.12: Counterfactual 2: consumer nudging - switching

% Switching	Z				
	0	0.25	0.5	0.75	1
Attention					
0	0	0	0	0	0
0.19	9.3	8.3	8.1	7.5	6.7
0.5	46.3	43.5	40.8	37.5	36.7
0.75	68.7	62.1	59.0	55.6	55.1
1	91.5	81.8	76.8	73.5	72.4

2.10 Conclusion

It is well known that consumers tend to revert to a default situation and that this can leave them worse off than if they were to re-optimize their choices at regular intervals. In this essay I have studied two forms of inertia, switching costs and inattention, that both act to keep consumers on mismatched plans.

Consumers were found to be highly inattentive only considering alternative plans 19% of the time, corresponding to re-evaluating their choice every six months. Coupled with estimated switching costs of £95 it is not surprising that the telecoms regulator in the UK is concerned by the lack of switching. Not only does this lead to consumers making inefficient decisions but it also decreases competition between firms with the switching cost acting to make consumers less price sensitive.

A simple policy proposal, whereby carriers placed consumers on the most cost effective plan each month, was found to increase welfare by £10 per consumer per month. This is approximately 40% of the tariff cost and extrapolating to the whole population equates to £200m per month. An alternative proposal where consumers were nudged to make better decisions had higher welfare gains, but is likely far harder (and costlier) to implement.

It is noted that in neither counterfactual were carriers allowed to reset their prices or adjust the plan offerings in response to changing market shares. This would be an interesting line for future research that would need to recognize that the presence of switching costs makes optimal price setting dynamic.

3

Strategic Effects of Potential Entry

3.1 Introduction

Economists have long been aware that incumbent firms with market power may have incentives to take actions to try to deter new entry (Kaldor (1935) and Bain (1949)). Survey evidence supports the view that managers sometimes act in this way (Smiley (1988)). However, while models of entry deterrence are central to the theoretical Industrial Organization literature (e.g., chapters 8 and 9 of Tirole (1994)), empirical evidence that particular models explain observed firm behavior is limited. In our view, one reason for this is that it is often unclear what the stylized two-period models that dominate the literature predict should happen when firms interact repeatedly as happens in the real world where, for example, a potential entrant may wait for several years before entering.

One, frequently analyzed, situation where incumbent firms repeatedly interact with a potential entrant is in the airline industry. Of particular interest here is the so-called "Southwest Effect". As documented by Goolsbee and Syverson (2008) (GS), incumbent airlines lower prices by as much as 20% on airport-pair routes when

Southwest serves both endpoint airports without (yet) serving the route itself, and as suggested in Bennett and Craun (1993) and Morrison (2001), these price cuts have substantial welfare effects. For example, Morrison estimates that Southwest's presence as a potential competitor lowered consumers' annual expenditure on airfares by $3.3 billion in 1998.

Despite the magnitude of the welfare effects the driver of the "Southwest Effect" remains elusive. In this paper we present evidence that attempts to illuminate why firms lower prices *prior* to Southwest entering (rationalizing why firms cut prices on entry of Southwest is far more straightforward and uncontroversial). In particular, and following the theory literature (see, e.g. Bulow et al. (1985) or Tirole (1994)), we present evidence of entry deterrence and entry accommodation. For an incumbent to deter entry it must credibly commit to post-entry product competition that would be unprofitable for an entrant. The strategies involved are numerous, for example, investment in capacity to lower marginal costs, expanding a product line to contract the residual demand curve or investing in switching costs to 'lock-in' the incumbent's installed base. A slightly more nuanced argument relates to deterring entry using limit pricing (see Milgrom and Roberts (1982)). In this strategy an incumbent firm with private information regarding demand or cost can use its price (or quantity) to signal to a potential entrant that entry would be loss making.

The level of investment required to deter entry may exceed the gains from remaining a dominant incumbent. In this case firms may choose to accommodate entry. This is not to say that firms wait idly for entry to occur, but instead that they can adopt strategies soften post-entry competition, recognizing that entry is inevitable. As with deterrence the strategies available are numerous. But whereas optimal deterrence strategies are derived by communicating to a potential entrant that it will experience negative post-entry profits, those for accommodation stem from an incumbent firm maximizing its own post-entry profits. This can call for counter-intuitive strategies such as raising marginal cost by under-investing in capacity, offering a limited product line or commitments to maintain high prices (e.g. "most favored customer clauses" or retail price maintenance).

Using a set of markets where we think incumbent firms are most likely to use entry deterring or accommodating strategies, we present new reduced-form evidence that point to the price cuts observed around Southwest becoming a potential entrant being driven by entry deterrence. In particular the evidence suggests that firms are using limit pricing to communicate their marginal costs. To do so, we draw on the empirical strategy proposed by Ellison and Ellison (2011) (EE), by showing that price-cutting behavior is more pronounced in markets where, based on exogenous factors, we predict an intermediate probability of entry by Southwest, which is where incumbents' incentives to make costly investments to deter entry should be greatest.

We also preset evidence that the price cuts are not easily rationalized by the leading alternative explanations, including other deterrence mechanisms that might also cause prices to fall, such as a desire to build customer loyalty to increase future demand, or by incumbents increasing their capacities in a way that reduces their marginal costs by lowering their load factors. In contrast, we see that another strategy, increased code-sharing, that is also adopted when Southwest threatens entry (Goetz and Shapiro (2012)), occurs primarily in those markets where Southwest is most likely to enter, suggesting that it may reflect incumbents readying themselves to accommodate entry.

Aside from the literature that has studied the Southwest Effect our work is related to the literature that has tried to provide empirical evidence of strategic investment. A common approach has looked for evidence of different investment strategies amongst firms (e.g., Lieberman (1987)) or effects of incumbent investment on subsequent entry (e.g., Chevalier (1995)) without specifying the exact mechanism involved. Masson and Shaanan (1982) try to provide evidence of limit pricing by pooling annual data on pricing from 37 different industries. Masson and Shaanan (1986) take a similar approach using data from 26 industries to argue that there is more evidence of incumbents using limit pricing than excess capacity to deter entry. While the empirical approach is very different, this conclusion is consistent with our results.[1] Much closer to our approach is Seamans (2013) who, inspired by the approach of EE, argues that the pricing of incumbent cable TV systems is consistent with an MR model of entry deterrence as, in the cross-section, prices vary non-monotonically to the distance to the nearest potential telephone company entrant.[2]

In the context of airlines, Snider (2009) and Williams (2012) provide structural evidence in favor of airlines using capacity investment in order to predate on small new entrants on routes coming out of their hubs. Our evidence suggests that incumbents did not use capacity investment as a strategy to try to deter a much stronger potential entrant, Southwest.

The rest of the paper is organized as follows. Section 4.3 describes our data, and discusses the sample selection that we feel best captures where firms may be using strategic investment to accommodate or deter entry. Section 3.3 provides reduced-form evidence of incumbent firms' strategic investment and discusses which of the mechanisms is most likely linked to the Southwest Effect. Section 4.6 concludes.

3.2 Data and Sample Selection

We now turn to examining the evidence for why incumbent airline carriers cut prices on airline routes when faced by Southwest as a potential entrant. In this section we

[1] Strassmann (1990) used the Masson and Shaanan approach to try to identify evidence of limit pricing in airline markets looking at 92 heavily-traveled routes. She found evidence that high prices attracted entry, but no significant evidence that prices were lowered strategically in order to deter entry.

[2] One difference in our reduced-form approach from the one used by Seamans is that we look directly at whether there is a non-monotonicity of *price changes* with respect to a predicted probability of entry once Southwest becomes a potential entrant.

discuss the empirical setting and existing literature, the data and our selection of a subset of markets that we believe best relates to where firms may be using strategic investment to deter or accommodate entry.

3.2.1 Empirical Application

With its large number of distinct airport-pair or city-pair markets that are usually served by at most a small number of carriers, the deregulated airline industry has provided a natural setting for investigating the economics of entry (Berry (1992)), the sources of market power and the effects of mergers (Borenstein (1989), Borenstein (1990), Kim and Singal (1993), Benkard et al. (2010)), and price discrimination (Borenstein and Rose (1995), Lazarev (2013)), amongst other topics. Several studies (e.g., Morrison and Winston (1987)) show that ticket prices tend to be lower when there are more potential competitors (defined as carriers serving one or both endpoints, but not yet serving the route)[3], but the literature has found that "the most dramatic effects from potential competition arise in the case of Southwest Airlines, which has long been the dominant low cost carrier" (Kwoka and Shumilkina (2010), p. 772). The well-known studies of GS and Morrison (2001) estimate that potential competition from Southwest lowers prices by as much as 33% and 19-28%, respectively.[4] While these estimates are far larger than any estimates, of which we are aware, of potential competition effects in any other industry (Bergman (2002)),

[3] For example, Morrison and Winston (1987) find that an additional potential competitor lowered prices by \$0.0015/passenger mile (1987 dollars) compared with \$0.0044/passenger mile for an actual competitor. Kwoka and Shumilkina (2010) find the largest effect of potential entry involving firms other than Southwest that we have seen in the literature, focusing on the effect of the 1987 merger of US Air and Piedmont. In cases where one of the merging airlines operated and the other was a potential entrant prior to the merger, prices rose by 5-6% relative to a control group where one of the carriers operated and the other one was not present at all.

[4] The fact that incumbent prices fell on at least some routes that Southwest had not yet entered was also frequently noted in the press. For example, "Consider what happened in the two years since Southwest began flying to TF Green Airport in Warwick RI ... competing airlines ... lowered fares - and not only to the cities where Southwest was flying", article by Laurence Zuckerman, 'As Southwest Invades East, Airline Fares Heading South', Oklahoma City Journal Record, February 8, 1999.

no clear rationale for why incumbents lower prices when Southwest is a potential competitor has been provided. GS show that price declines are smaller on routes where Southwest announces its entry before it begins operating at the airport, which they tentatively interpret as evidence in favor of an entry deterrence, rather than an entry accommodation explanation, although the difference from the remaining routes in their sample is not statistically significant. They do show incumbents tend not to increase capacity when lowering prices, and they speculate that incumbents may be trying to increase their customers' loyalty, possibly through frequent-flyer programs, in order to reduce the demand that Southwest might receive post-entry (GS, p. 1629).

Our contribution is to show that for a set of markets where a limit pricing story is plausible alternative explanations cannot generate the sort of price declines that are observed in the data. We do so by providing new evidence showing that the price declines are motivated by deterrence, and by providing new evidence against other explanations for why prices fall, such as a desire to build customer loyalty or because marginal costs, that are a function of load factors, are falling.

In order to identify markets whereby a limit pricing story is plausible we need to consider two factors. First, incumbents must have private information about demand conditions or their own marginal cost. To that end we focus on a set of airport-pair markets where there is a dominant incumbent (the exact definition will be given below).[5] Almost all of these markets involve at least one hub or focus city for the incumbent, and it was in these markets that Bennett and Craun (1993) originally identified the Southwest Effect. On these routes, it is it well-understood that the incumbent's marginal (opportunity) cost of selling a seat to a local passenger will

[5] GS also use airport-pairs, and if we used city-pairs, the number of dominant incumbent markets where Southwest becomes a potential entrant would be small. Morrison (2001) estimates that Southwest has substantially smaller effects on fares when it only serves nearby airports as either an actual or a potential competitor.

depend critically on the number and the profitability of connecting passengers that could also travel on the segment. As can be seen in litigation involving alleged predation by carriers at their hubs (Edlin and Farrell (2004), Elzinga and Mills (2005)), which relied on the incumbent's internal accounting measures, appropriately measuring marginal costs on these routes is very complicated even ex-post, partly because of the vast number of different destinations connecting passengers might be flying to. The incumbent's marginal cost is therefore likely to be opaque to potential entrants, including Southwest, that have to make contemporaneous decisions about whether to enter, as well as being likely to evolve over time as network flows and the options available to connecting travelers change.[6] In addition, the traditional importance of hub routes for legacy airline profitability makes it plausible that incumbents would be willing to sacrifice current profits to try to deter the entry of a carrier known to set very low prices once it enters, on as many of these routes as possible.[7]

A second factor necessary for limit pricing to be plausible is that the potential entrant's decision to enter should be sensitive to what it believes about the incumbent's marginal cost or some other feature of the market that will affect its post-entry profits. Consistent with GS's logic about pre-announced entry, there are clearly going to be some routes, including to Southwest's focus airports such as Las Vegas or Chicago Midway, where Southwest is almost certain to enter immediately, or very soon after, it enters an airport, independent of an incumbent's actions. At the other extreme, there are likely to be some routes (whether due to distance, or market size) that are very unlikely to be entered even if the incumbent's marginal costs are high.

[6] One might object that other carriers can use publicly available data to understand these network flows. However, the Department of Transportation only releases these data with a lag of at least three months, and our theoretical and simulation results hold even if we assume that the incumbent's current marginal cost is revealed to the entrant after it has made its entry decision.

[7] For example, when Southwest entered Philadelphia in 2004, the US Airways CEO David Siegel told employees "Southwest is coming for one reason: they are coming to kill us. They beat us on the West Coast, and they beat us in Baltimore. If they beat us in Philadelphia, they're going to kill us." (Business Travel News, March 25, 2004, "Philadelphia Could be US Airways' Last Stand").

This is recognized by our identification strategy, as our evidence will come from the fact that we observe the largest price declines in a set of markets that a simple entry model predicts are most likely to be marginal for Southwest to enter, which are the markets where its beliefs about the incumbent are most likely to matter.[8]

3.2.2 Data

Most of our data is drawn from the U.S. Department of Transportation's Origin-Destination Survey of Airline Passenger Traffic (Databank 1, DB1), a quarterly 10% sample of domestic tickets, and its T100 database that reports monthly carrier-segment level information on flights, capacity and the number of passengers carried on the segment (which may include connecting passengers). We aggregate the T100 data to the quarterly-level to match the structure of the DB1 data. Our data covers the period from Q1 1993-Q4 2010 (72 quarters).[9]

Following GS, we define a market to be a non-directional airport-pair with quar-

[8] While Southwest, like other carriers, has never fully described its entry strategy, comments from the company indicate that it is sensitive to current market conditions. For example, "It's all based on customer demand. We're always evaluating markets to see if they are overpriced and underserved" (quote by Southwest spokesperson Brandy King, cited in an article 'Southwest to Offer Flights between Sacramento and Orange County, CA' by Clint Swett, Knight Ridder Tribune Business News, 6 Mar 2002). Also, "Southwest does not have any hard and fast criteria dictating when it enters a market. The method is a cautious, reactive approach designed to take advantage of opportunities as they arrive" (description of March 13 2008 comments by Brook Sorem, Southwest's manager of Schedule Planning, reported in an World Airport Week article "What Can Airports Do to Attract Southwest Airlines?", March 24, 1998). Herb Kelleher, one of the founders and longtime Chairman and CEO of Southwest, also admitted to having at least six different strategic plans for how Southwest might develop in the Northeast United States, after its initial entry into Providence, R.I. (from Wall Street Journal article by Scott McCartney, "Turbulence Ahead: Competitors Quake as Southwest is Set to Invade the Northeast", October 23, 1996). The claim that at least some of Southwest's entry decisions in the mid/late-1990s and 2000s were marginal is supported by existing research, such as Boguslaski et al. (2004), which showed that the ability of a simple probit (based on exogenous market characteristics) to predict Southwest's entry decisions declined significantly in the 1990s (explaining only 41% of entry decisions from 1995-2000 compared with almost 60% for the 1990-2000 decade as whole).

[9] There are some changes in reporting requirements and practices over time. For example, prior to 1998 operating and ticketing carriers are not distinguished in DB1, making it impossible to analyze code-sharing in the first part of our data, and prior to 2002 regional affiliates, such as Air Wisconsin operating as United Express, were not required to report T100 data. See footnote 17 for some related comments.

ters as periods. We only consider pairs where, on average, at least 50 DB1 passengers are recorded as making return trips each period, possibly using connecting service, and in everything that follows a one-way trip is counted as half of a round-trip. We exclude pairs where the round-trip distance is less than 300 miles. We define Southwest as having entered a route once it has at least 65 flights per quarter recorded in T100 and carries 150 non-stop passengers on the route in DB1, and we consider it to be a potential entrant once it serves at least one route out of each of the endpoint airports.[10]

Based on our potential entrant definition there are 1,872 markets where Southwest becomes a potential entrant after the first quarter of our data and before Q4 2009, a cutoff that we use so we can see whether Southwest enters the market in the following year, an observed outcome that we will use to estimate which market characteristics make entry more likely. Southwest enters 339 of these markets during the period of our data. We will call these 1,872 markets our "full sample". Most of our analysis will focus on the subset of these markets where there is one carrier that is a *dominant incumbent* before Southwest enters. As we want to identify only carriers that are really committed to a market, rather than just serving it briefly, we use the following rules to identify a dominant carrier:

1. to be considered active in a quarter it must carry at least 150 DB1 non-stop passengers;

2. once it becomes active in a market the carrier must be active in at least 70% of quarters before Southwest enters, and in 80% of those quarters it must account for 80% of direct traffic on the market and at least 50% of total traffic.[11]

[10] While this definition means that we may consider Southwest to have entered a market when its schedule is quite limited, we note that this is actually a more stringent criterion than the one used by GS.

[11] To apply this definition we have to deal with carrier mergers (for example, Northwest was the

Table 3.1: Comparison of the Full and Dominant Incumbent Samples

	Full Sample		Dominant Incumbent Samples			
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Mean endpoint population (m.)	2.373	1.974	2.850	1.894	3.155	2.081
Round-trip distance (miles)	2,548.48	1,327.04	1,251.44	749.58	1,315.1	803.07
Constructed market size measure	27,837	44,541	62,751	66,633	47,975	61,397
Origin or destination is a:						
primary airport	0.161	0.368	0.330	0.473	0.277	0.451
secondary airport	0.301	0.459	0.321	0.469	0.354	0.482
big city	0.587	0.492	0.858	0.350	0.877	0.331
leisure destination	0.093	0.291	0.113	0.318	0.108	0.312
slot controlled airport	0.033	0.179	0.057	0.230	0.092	0.292
Number of markets	1,872		106		65	

We identify 106 markets with a dominant incumbent before Southwest enters, but in some of these markets Southwest enters at the same time as it becomes a potential entrant (i.e., the market is one of the first ones that Southwest enters when it begins serving one of the endpoint airports) and in a few of them the dominant incumbent becomes active only after Southwest is a potential entrant on the route. In 65 markets we observe quarters where the incumbent carrier is dominant both before Southwest becomes a potential entrant and after it is a potential entrant but before it actually entered. It is data from these routes that will identify the effects of the potential entry threat on the price set by a dominant incumbent, although we include the remaining 41 routes in our regressions as they help to pin down the coefficients on the time effects and other controls included in the specification.[12] The 106 and 65 markets are listed in Appendix B.

Table 3.1 provides some statistics for the full sample, and the sub-samples of 106

dominant carrier on the Minneapolis-Oklahoma City route before it merged with Delta in 2008, after which Delta is the dominant carrier). When we define carrier fixed effects we treat the dominant carrier before and after a merger as the same carrier even if the name of the carrier changed.

[12] For example, Southwest began service out of Philadelphia (PHL) in Q3 2004. It already operated at both Chicago Midway (MDW) and Columbus, OH (CMH), and so, under our definitions, it became a potential entrant into both the PHL-MDW (where the dominant incumbent was ATA) and PHL-CMH (where the dominant incumbent was US Airways) markets in Q3 2004. However, it immediately began service on the PHL-MDW route, but did not enter the PHL-CMH market until Q4 2006.

and 65 markets. Relative to the full sample, the dominant incumbent markets tend to be shorter with larger endpoint cities, as measured by either average population or an indicator for whether one of the endpoints meets the “big city”definition of Gerardi and Shapiro (2009).[13] All of the markets in our dominant firm sample are shorter than the longest routes that Southwest flies non-stop (these include long, cross-country routes such as Las Vegas-Providence), so, by this metric, it is plausible that any of our routes could be entered.[14] As only the largest cities have multiple major airports, the dominant incumbent markets are also more likely to involve an airport identified as a primary or secondary airport. On the other hand, the standard deviations show that both samples are quite heterogeneous with respect to these market characteristics. We also construct a variable measuring market size, which we will use as an additional variable for predicting the probability that Southwest enters a market. As explained in Appendix C, this variable is constructed by estimating a generalized gravity equation using the Poisson Pseudo-Maximum Likelihood approach recommended by Silva and Tenreyro (2006), which allows us to capture the fact that the amount of travel on a route varies systematically with distance and the popularity of the particular airports.

Table 3.2 reports, for the dominant incumbent markets, summary statistics for variables that vary over time, such as average prices (in Q4 2009 dollars), yield (average fare divided by route distance, a widely used metric for comparing fares

[13] Gerardi and Shapiro (2009) define the largest 30 MSAs as being big cities, although they exclude some MSAs, such as Orlando, on the basis that are primarily vacation destinations. We also follow them in defining “leisure” destinations, which include cities such as New Orleans and Charleston, SC, as well as Las Vegas and several cities in Florida. We define slot controlled airports as JFK, LaGuardia and Newark in the New York area, Washington National and Chicago O’Hare, although O’Hare is no longer slot controlled. We identify metropolitan areas with more than one major airport using http://en.wikipedia.org/wiki/List_of_cities_with_more_than_one_airport, and identify the primary airport in a city as the one with the most passenger traffic in 2012.

[14] The longest route in the dominant firm sample is Las Vegas-Pittsburgh, which is one of the markets that Southwest enters immediately. Even though some longer routes are flown by only one carrier, they fail to meet our definition of dominance because many people will fly these routes via connecting service on other carriers.

across routes of different lengths) and market shares. Quarters are aggregated into three groups, which we will refer to frequently below: "Phase 1" - before Southwest is a potential entrant; "Phase 2" - when Southwest is a potential entrant but has not yet entered the route; and, "Phase 3" - after Southwest enters (if it enters during the sample). Entered markets will be a selected set of markets which explains why the dominant carrier's average capacity and passenger numbers for the Phase 3 markets are higher than for the other groups. The summary statistics are, however, consistent with Southwest's actual entry into a market reducing prices dramatically, so that an incumbent should be willing to make investments to deter entry if it is likely that they would be effective. They are also consistent with incumbents responding to the threat of entry by lowering prices, suggesting that limit pricing may be one of these investments.[15]

The summary statistics also provide some evidence against an alternative story for why prices fall in Phase 2. Recall that in Phase 2, Southwest serves both endpoint airports so that passengers may be able to travel the route by connecting on Southwest[16], in which case one might argue that Southwest should be viewed as a competitor with an inferior product rather than just a potential entrant. This could provide an alternative explanation for why prices fall. However, from the table we see that Southwest's average market share in Phase 2 is less than 1.5%, compared with the dominant carrier's share of over 80%, while Southwest's fares for these connections are also high compared to its fares when it enters the market with direct service. Therefore, the degree of direct competitive pressure that Southwest exerts

[15] Yields and average fares do not vary in the same proportion across the phases, consistent with the fact that the set of markets that Southwest enters are not random with respect to the length of the route. For this reason we will look at both price metrics in the results below.

[16] Southwest does not always allow customers to buy tickets between any pair of airports that it serves, reflecting the fact that, compared to the legacy carriers, its business model is more focused on point-to-point travel. However, we do not have data on which routes it will sell tickets that involve connections.

Table 3.2: Summary Statistics: Dominant Incumbent Sample

	Phase 1: $t < t_0$ All markets		Phase 1: $t < t_0$ Delayed entry markets		Phase 2: $t_0 \leqslant t < t_e$		Phase 3: $t \geqslant t_e$	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Incumbent Pricing								
Yield (average fare / distance)	0.510	0.317	0.525	0.313	0.439	0.297	0.292	0.143
Average fare	472.19	137.68	509.28	147.97	418.13	123.29	254.76	60.50
Southwest Pricing								
Yield (average fare / distance)	-		-		0.254	0.095	0.234	0.726
Average fare	-		-		363.41	84.38	215.98	59.42
Passenger Shares								
Incumbent	0.815	0.193	0.771	0.208	0.851	0.129	0.460	0.195
Southwest	-		-		0.016	0.030	0.486	0.210
Incumbent Capacity and Traffic								
Capacity (seats performed)	75,592	53,430	68,752	49,729	69,324	48,844	100,038	52,900
Segment passengers (incl. connecting passengers)	45,995	32,828	41,862	30,423	48,413	33,389	71,261	38,987
Load factor	0.612	0.104	0.618	0.107	0.712	0.117	0.714	0.079
Proportion passengers connecting	0.837	0.118	0.849	0.113	0.832	0.125	0.775	0.098
Code-share measure	0.086	0.211	0.124	0.257	0.257	0.352	0.188	0.245
Southwest Capacity and Traffic								
Capacity (seats performed)	-		-		-		87,629	59,887
Segment passengers (incl. connecting passengers)	-		-		-		56,403	38,478
Load factor	-		-		-		0.658	0.080
Proportion passengers connecting	-		-		-		0.702	0.102
Code-share measure	-		-		-		0.018	0.093
Number of markets	106		65		65		54	

on the incumbent's pricing in Phase 2 should be small. In Section 3.3 we will provide additional evidence against this 'actual competition' explanation for why prices fall when entry is threatened.

The last sections of the table show the amount of capacity (measured by seats performed), the total number of passengers carried on the segment, and the load factor (number of passengers carried divided by the number of seats). All numbers are based on data from T100. We also report an estimate of the proportion of passengers traveling the route to make connections.[17] For both incumbents and

[17] The number of connecting passengers is calculated by taking 10 times the number of passengers traveling the route in DB1 from the total number of passengers reported in T100. When one combines data from DB1 and T100, some inconsistencies are introduced, because the DB1 sampling weights are not necessarily the same across routes and some passengers reported as direct in DB1 may be traveling via other airports without a change of plane or on regional affiliates that have not reported T100 data in some of our quarters. Therefore we restrict ourselves to some fairly

Southwest, the majority of passengers carried on these routes are making connections, a point we will return to in Section ??. The entry of Southwest as a potential entrant or an actual route entrant is associated with an increase in the incumbent's load factor and a decline in the proportion of connecting passengers, consistent with the fall in local fares raising local demand. We also report a measure of code-sharing by the incumbent, based on observations after 1998, where DB1 reports both the ticketing and operating carrier. Our measure is the proportion of the carrier's passengers on a route that were ticketed by a different carrier.[18] Goetz and Shapiro (2012) find that incumbents are more likely to code-share with other carriers when Southwest threatens entry, and we also see code-sharing increasing in Phases 2 and 3 in our data.

3.3 Evidence of Strategic Investment in the Dominant Incumbent Sample

In this section we present reduced-form evidence that incumbents may use strategic investment in order to cut prices when Southwest becomes a potential entrant on an airline route. We do so using two different frameworks. First, we extend the analysis in GS by trying to discriminate between several alternative explanations for why prices fall and by focusing on dominant incumbent markets that fit the market structure assumed by most models of strategic investment. Second, we utilize an

broad-brush comments about connecting traffic patterns even though, as our extended model in Section ?? suggests, the fact that there is a lot of connecting traffic on the routes in our sample may play an important role in explaining why there may be significant limit pricing.

[18] A code-sharing arrangement allows specific non-operating (marketing) carriers to sell tickets on a flight operated by another carrier, and the flight itself will usually be given a flight number for each of the code-sharing carriers. Continental and Northwest, and United and US Airways engaged in fairly extensive code-sharing in some quarters during our data. Carriers that are not code-sharing may still sell a ticket on another carrier's flight as part of an 'interlining' agreement. Therefore the fact that we measure a proportion as not being equal to zero is not indicative that a full code-sharing agreement was in place. However, we see much higher proportions for carrier combinations with known code-sharing agreements.

approach suggested by EE to assess whether there is a non-monotonic relationship between the level of strategic investment and the probability of Southwest's entry into a market.

3.3.1 GS Analysis

We start by confirming that dominant incumbents do cut prices significantly when Southwest becomes a potential entrant on a route by serving both endpoint airports but not yet serving the route (Phase 2). To do so we follow GS, who use markets with any number of incumbents, by utilizing the following regression specification:

$$\text{Price Measure}_{j,m,t} = \gamma_{j,m} + \tau_t + \alpha X_{j,m,t} + \ldots$$
$$\sum_{\tau=-8}^{8+} \beta_\tau SWPE_{m,t_0+\tau} + \sum_{\tau=0}^{3+} \beta_\tau SWE_{m,t_e+\tau} + \varepsilon_{j,m,t} \tag{3.1}$$

where $\gamma_{j,m}$ are market-carrier fixed effects and τ_t are quarter fixed effects. Only observations for the dominant incumbent are included in the regression, but the control variables X include the number of other carriers serving the market (separate counts for direct and connecting service) as well as interactions between the jet fuel price[19] and route distance. t_0 is the quarter in which Southwest becomes a potential entrant, so $SWPE_{m,t_0+\tau}$ is an indicator for Southwest being a potential entrant, but not an actual entrant, into market m for quarter $t_0 + \tau$. If Southwest enters it does so at t_e, and $SWE_{m,t_e+\tau}$ is an indicator for Southwest actually serving the market in quarter $t_e + \tau$. We use observations for up to three years (12 quarters) before Southwest becomes a potential entrant, and the β coefficients measure price changes relative to those quarters that are more than eight quarters before Southwest becomes a potential entrant or, if Southwest becomes a potential entrant within the first eight quarters that the dominant carrier is observed in the data, the first quarter that the market is observed. We estimate separate coefficients for the quarters immediately

[19] Specifically, U.S. Gulf Coast Kerosene-Type Jet Fuel Spot Price FOB (in $/gallon).

around the entry events, but aggregate those quarters further away from the event where we have fewer observations. In our analysis markets are weighted equally, but the results are similar if observations are weighted by the average number of passengers carried on the route.

Table 3.3 presents two sets of coefficient estimates, using the log of the average price and the yield as alternative price measures. Average prices fall by 10-14% when Southwest becomes a potential entrant. The average yield in Phase 1 is 0.544, so the yield coefficients imply similar proportional changes. If Southwest enters, average prices decline by an *additional* 30-45%, giving a decline of 45-60% relative to prices at the start of Phase 1. While our Phase 2 price declines are slightly smaller than those identified by GS, our Phase 3 declines are significantly larger, presumably reflecting the fact that dominant incumbents have more market power prior to Southwest's entry than the average incumbent in GS's sample.

3.3.2 EE Analysis

In the context of a fairly general model of strategic investment by an incumbent monopolist, EE argue that, when deterrence incentives are present, they can generate a non-monotonic relationship between the level of investment and the probability of entry. In our setting, their logic would apply in the following way. When Southwest becomes a potential entrant, an incumbent will not be willing to cut prices very much in markets where entry is very unattractive to Southwest, because it is likely only to be sacrificing monopoly profits. In markets where entry is very attractive, the monopolist will also not want to cut prices because it is unlikely that entry can be deterred and it will only be sacrificing the profits that it can make before entry happens. On the other hand, in markets that are marginal for entry, it is possible that entry will be prevented (or delayed) if the incumbent signals that its marginal costs are low enough. EE show how this insight can be developed into a two-stage

Table 3.3: Incumbent Pricing In Response to Southwest's Actual and Potential Entry

		Phase 1		*Phase 2*		*Phase 3*
Fare						
	$t_0 - 8$	-0.047	t_0	-0.105^{***}	t_e	-0.416^{***}
		(0.029)		(0.031)		(0.066)
	$t_0 - 7$	-0.022	$t_0 + 1$	-0.115^{***}	$t_e + 1$	-0.514^{***}
		(0.0307)		(0.034)		(0.069)
	$t_0 - 6$	-0.040	$t_0 + 2$	-0.131^{***}	$t_e + 2$	-0.539^{***}
		(0.034)		(0.032)		(0.077)
	$t_0 - 5$	-0.041	$t_0 + 3$	-0.131^{***}	$t_e + 3$	-0.602^{***}
		(0.034)		(0.032)		(0.080)
	$t_0 - 4$	-0.015	$t_0 + 4$	-0.135^{***}	$t_e + 4$	-0.608^{***}
		(0.033)		(0.034)		(0.082)
	$t_0 - 3$	-0.009	$t_0 + 5$	-0.137^{***}	$t_e + 5$	-0.577^{***}
		(0.029)		(0.038)		(0.084)
	$t_0 - 2$	-0.0761^{**}	$t_0 +$ 6-12	-0.206^{***}	$t_e +$ 6-12	-0.589^{***}
		(0.029)		(0.047)		(0.081)
	$t_0 - 1$	-0.0874^{***}	$t_0 +$ 13+	-0.309^{***}	$t_e +$ 13+	-0.589^{***}
		(0.029)		(0.051)		(0.086)
Yield						
	$t_0 - 8$	-0.027	t_0	-0.060^{***}	t_e	-0.234^{***}
		(0.014)		(0.019)		(0.051)
	$t_0 - 7$	-0.005	$t_0 + 1$	-0.051^{**}	$t_e + 1$	-0.273^{***}
		(0.017)		(0.021)		(0.052)
	$t_0 - 6$	-0.011	$t_0 + 2$	-0.063^{***}	$t_e + 2$	-0.286^{***}
		(0.018)		(0.019)		(0.056)
	$t_0 - 5$	-0.008	$t_0 + 3$	-0.062^{***}	$t_e + 3$	-0.308^{***}
		(0.018)		(0.019)		(0.057)
	$t_0 - 4$	-0.009	$t_0 + 4$	-0.066^{***}	$t_e + 4$	-0.315^{***}
		(0.018)		(0.019)		(0.059)
	$t_0 - 3$	-0.008	$t_0 + 5$	-0.066^{***}	$t_e + 5$	-0.313^{***}
		(0.017)		(0.022)		(0.060)
	$t_0 - 2$	-0.042^{**}	$t_0 +$ 6-12	-0.115^{***}	$t_e +$ 6-12	-0.332^{***}
		(0.018)		(0.027)		(0.059)
	$t_0 - 1$	-0.047^{**}	$t_0 +$ 13+	-0.185^{***}	$t_e +$ 13+	-0.361^{***}
		(0.018)		(0.034)		(0.067)

Notes:. Standard errors clustered by route-carrier are in parentheses. ***, ** and * denote statistical significance at the 1, 5 and 10% levels respectively. Number of observations is 3,904 and the adjusted R^2s are 0.81 ("Fare") and 0.86 ("Yield"). Phases are defined in the text.

empirical strategy for identifying strategic investment in settings where observable and exogenous variables, such as market size, change the attractiveness of entry. In the first stage, a simple model of the entry probability is estimated to construct a single index of the attractiveness of the market to the potential entrant and then, in the second stage, the monotonicity of the relationship between this index and the incumbent's (possibly strategic) investment is examined.

For our first stage, we estimate a probit model of Southwest's entry using the full sample of 1,872 markets where an observation is a route and the dependent variable is equal to one if Southwest entered within four quarters of becoming a potential entrant, where we are implicitly assuming that Southwest will typically choose to enter the most attractive markets from an airport first.[20] The explanatory variables include several measures (and their squares) of market size, including measures of endpoint population and our gravity model-based market size measure; route distance; measures of carrier presence at the endpoint airports; and, an indicator for whether one of the airports is slot constrained. Further details of the variables included and the estimated coefficients are given in Appendix **??**. Consistent with previous research, e.g. Boguslaski et al. (2004), we are able to explain a reasonable degree of variation (pseudo-R^2 0.37) in Southwest's entry decisions in the full sample, with Southwest more likely to enter shorter routes between bigger cities, especially when it already serves a significant number of routes out of the endpoints. For the subset of 65 markets the predicted within-four-quarter entry probabilities vary from 0.01 to 0.9, with the 20^{th}, 40^{th} and 60^{th} and 80^{th} percentiles at 0.02, 0.085, 0.204 and

[20] It would be inappropriate to use a dummy for Southwest ever entering because, in our relatively long sample, different markets are exposed to the possibility of entry for different periods of time. Using a four quarter rule also means that we minimize the truncation problem associated with the end of the sample while still having a significant number of observations. In the data, Southwest enters around 60% of the routes that it will ever enter from an airport in the first quarter that it begins operations, and of the remaining markets that it eventually enters, it enters around 33% in the following three quarters, 18% in its second year, 13% in its third year and the remainder in later years.

0.512.

In the second stage, we only use Phase 1 and 2 observations from the dominant incumbent sample, and we test how the size of the Phase 2 price decline varies with the entry probability using the following market-carrier fixed effects specification:

$$\text{Price Measure}_{j,m,t} = \gamma_{j,m} + \tau_t + \alpha X_{j,m,t} + \dots$$

$$\beta_0 SWPE_{m,t} + \beta_1 \widehat{\rho_m} \times SWPE_{m,t} + \beta_2 \widehat{\rho_m}^2 \times SWPE_{m,t} + \epsilon_{j,m,t} \qquad (3.2)$$

where $\widehat{\rho_m}$ is the predicted probability of entry (within one year) for market m, j is the dominant carrier, $\gamma_{j,m}$ and τ_t are market-carrier and quarter fixed effects, and $X_{j,m,t}$ includes the same controls that were used in the GS specification. $SWPE_{m,t}$ is an indicator for a market-quarter in which Southwest is a potential entrant (i.e., a Phase 2 observation). Standard errors are adjusted to allow for uncertainty in the first-stage estimate $\widehat{\rho_m}$, as well as heteroskedasticity and first-order serial correlation in the residuals.[21] If the incumbent is using a deterrence strategy that causes prices to fall, then we would expect $\widehat{\beta_0} \approx 0$, $\widehat{\beta_1} < 0$ and $\widehat{\beta_2} > 0$. On the other hand, an entry accommodation explanation for falling prices would predict that $\widehat{\beta_0} \approx 0$ and a combination of $\widehat{\beta_1}$ and $\widehat{\beta_2}$ such that prices are expected to fall more in markets where entry is more likely. Obviously given our identification strategy, one might be concerned that there is something unobserved about intermediate probability of entry markets, that will also affect prices. While we cannot assess this directly, we can, of course, assess how observable variables (including some not included in the entry probit) vary with the implied entry probability. To this end, Appendix D presents a 'balance table' where we divide the dominant incumbent markets into three groups based on the implied probability of entry. On most dimensions, the intermediate probability markets lie between the low and high probability markets, and

[21] To do this, we specify the derivatives of the first-stage log-likelihood as an additional set of moment conditions and adapt the methodology outlined by Ho (2006). Regressions using the estimated second-stage residuals indicate that only first-order serial correlation is significant, and allowing for additional periods of serial correlation does not change the standard errors significantly.

on the remaining dimensions the differences between the groups are not statistically significant.

3.4 Explanations for the Observed Price Declines

Why might prices fall in Phase 2? Several explanations are possible. One explanation, 'actual competition', is that once Southwest serves the endpoint airports, its ability to provide connecting service provides enough competition to the dominant incumbent that the static equilibrium response is for prices to fall. The other explanations involve either some type of change in the incumbent's demand (for example, from connecting passengers) that lowers the incumbent's load factor and its marginal costs, and so causes the optimal monopoly price to fall, or some type of strategic response by the incumbent. Strategic responses could be of two types: one type would be that the incumbent is sacrificing current profits to try to increase its profits in the game that will be played once Southwest enters (i.e., 'entry accommodation') and the other is that the incumbent takes actions that will make Southwest perceive that entry will be less profitable ('entry deterrence'). Limit pricing is clearly a deterrence story, given the standard assumption that the game that follows entry is complete information. On the other hand, several strategies might be used for either accommodation or deterrence. For example, the dominant incumbent might change capacity to affect its marginal cost; lower prices to increase future demand by raising customer loyalty; or, sign code-sharing agreements with other carriers. We will now consider each of these explanations.

3.4.1 Incumbent Price Cuts due to Entry Accommodation, Capacity Investment or Limit Pricing

Table 3.4 reports results for several strategic investment variables in the EE analysis of equation (3.2), while Figure 3.1 illustrates the same relationship for the log of average fare (left panel) and average yield (right panel). Consistent with a deterrence explanation, but not an accommodation explanation, on average, the price declines are largest and most statistically significant for intermediate probabilities of Southwest entry, but are smaller, and not necessarily significantly different from zero, for markets where entry probabilities are either high or low. In the regression both the linear and the quadratic terms are statistically significant at the 1% level (this is also true in the log(average price) regression). Appendix Table F.1 shows that the same pattern holds for the 25^{th}, 50^{th} and 75^{th} percentiles of the price distribution, which, as well as indicating the robustness of the result for mean prices, also suggests that incumbents are not targeting particular types of consumers, either to prevent them from making connections on other routes via Southwest (which might lead to targeting of price-sensitive travelers buying the cheapest direct tickets) or to build their future loyalty (which might lead to targeting of business travelers who tend to buy more expensive tickets close to the date of departure), when cutting prices.

Table 3.4: Ellison and Ellison Reduced-Form Analysis: Second Stage Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Yield	Yield	Log Capacity	Log Passengers	Log Load Factor	Prop. Code-share
$SWPE_{m,t}$	0.0139	-0.064	0.079	0.139**	0.0599***	0.0030
	(0.0147)	(0.059)	(0.052)	(0.053)	(0.0169)	(0.0026)
$\widehat{\rho_m} * SWPE_{m,t}$	-0.721***	-0.555**	-0.2436	0.451	0.695***	-0.023
	(0.136)	(0.273)	(0.402)	(0.423)	(0.177)	(0.030)
$\widehat{\rho_m}^2 * SWPE_{m,t}$	0.921***	0.722**	0.2045	-1.009	-1.214***	0.123**
	(0.194)	(0.315)	(0.515)	(0.581)	(0.3118)	(0.049)
Observations	3,622	3,622	3,100	3,100	3,100	2,243

Notes: Heteroskedasticity robust Newey-West standard errors allowing for one period serial correlation and corrected for first-stage approximation error in the entry probabilities in parentheses. Column (2) includes controls for the convenience of connecting on Southwest. The different number of observations reflect differences in the coverage and reporting in the DB1 and T100 data during our sample period. ***, ** and * denote statistical significance at the 1, 5 and 10% levels respectively.

Figure 3.2 plots the estimated change in yield for each of the dominant incumbent markets separately against the estimated entry probability (the figure using log of the average fare looks very similar). These market-specific effects are estimated by replacing the three $SWPE_{m,t}$ terms in specification (3.2) with $SWPE_{m,t}\times$market m dummy interactions, with the plotted points being the point estimates of the coefficients on these interactions. While the effects for individual markets are heterogeneous, which is consistent with a limit pricing model where either incumbents have different levels of marginal cost or perceive different degrees of entry threat that are not measured perfectly by our estimated $\widehat{\rho}$s, it is clear that prices do not tend to decrease in markets with high or very low entry probabilities, while a significant proportion of the markets in between experience quite large Phase 2 price declines.

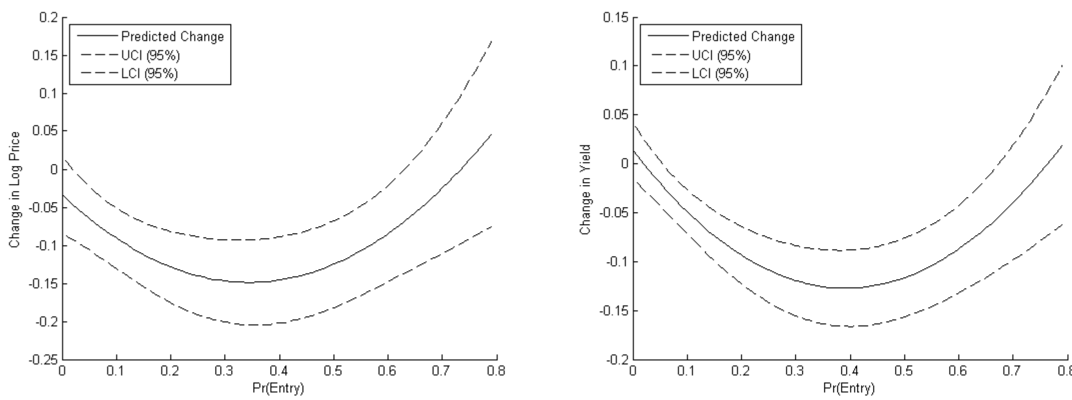

FIGURE 3.1: Predicted incumbent price and yield changes in Phase 2 as a function of Southwest's predicted probability of entry.

Of course, these results are also consistent with dominant incumbents using other strategies, such as adding capacity (as suggested by Snider (2009) and Williams (2012) in the context of possible predation by hub carriers), to try to deter entry, either instead of or in addition to, limit pricing. Capacity investments might cause monopoly prices to decline, by lowering load factors and therefore lowering the marginal cost of carrying additional passengers. We investigate whether capacity is increasing in the intermediate probability markets, by using the log of the incumbent's capacity on the route (measured by T100 'seats performed') as the dependent variable in specification (3.2). In fact, as seen in Table 3.4, column (3) and Figure 3.3 (top-left panel), capacity does not tend to change significantly when Southwest becomes a potential entrant in any type of market, irrespective of the entry probability.

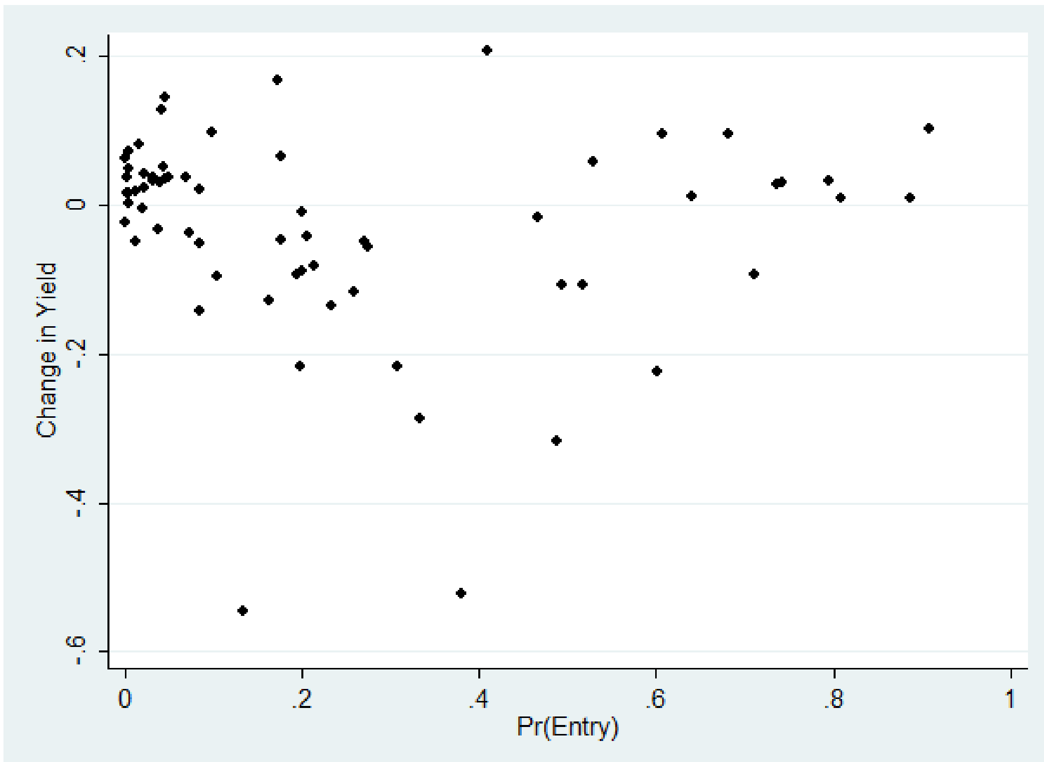

FIGURE 3.2: Market-by-market predicted yield declines in Phase 2 as a function of Southwest's predicted probability of entry.

We also use the log of the incumbent's load factor and the log of the number of passengers flown on the route (measured in T100, where both local and connecting passengers are counted) as dependent variables, as reductions in connecting traffic could also cause costs to fall, independent of capacity changes, on some routes.[22] The results are in columns (4) and (5) of Table 3.4 and the top-right and bottom-left

[22] For example, once Southwest is operating at an airport it may provide more attractive options to customers who would otherwise have connected using the incumbent's route of interest as a segment.

panels of Figure 3.3. We observe that when Southwest becomes a potential entrant, the dominant carrier tends to carry more passengers and have higher load factors in markets with intermediate probabilities of entry, suggesting that, if anything, its marginal costs increase rather than fall.

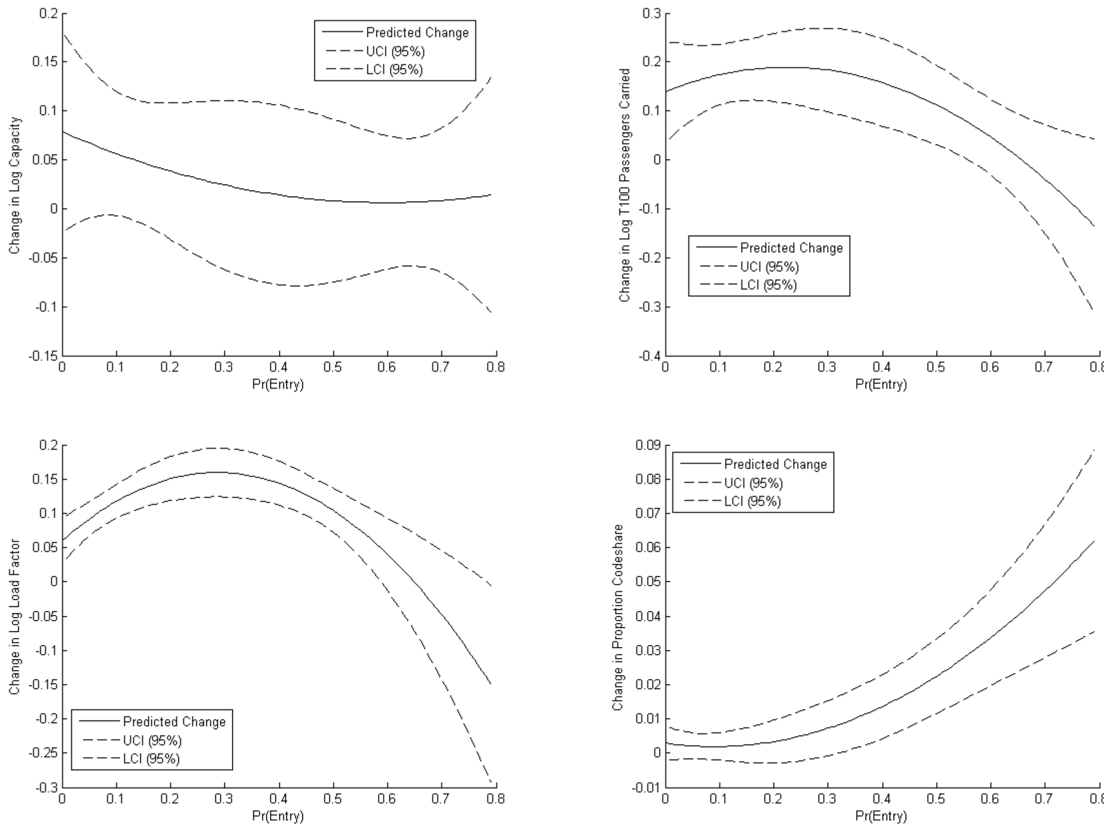

FIGURE 3.3: Predicted incumbent responses in Phase 2 as a function of Southwest's predicted probability of entry. The responses shown are the log of capacity (seats performed) (top-left panel), the log of segment passengers (includes passengers connecting onto other routes) (top-right), the log of the load factor, (bottom-left panel), and proportion of passengers carried that have a different ticketing carrier (code-shared) (bottom-right panel).

Capacity investment is not the only non-limit pricing strategy that carriers might use to try to deter entry. As noted above, code-sharing also increases when entry is threatened. When we use our code-sharing variable as the dependent variable in (3.2), we see that, in contrast to price cuts, dominant carriers increase code-sharing on routes where Southwest's entry is most likely (column (6) of Table 3.4 and the bottom-right panel of Figure 3.3).[23] This suggests that code-sharing, unlike lower prices, is used primarily as a strategy for preemptively accommodating Southwest's entry.

3.4.2 Incumbent Price Cuts due to Actual Competition

One feature of the results in Table 3.3, also found in GS, is that prices start declining two quarters *before* Southwest becomes a potential entrant. While in some settings one would be concerned that a pre-treatment change must reflect some other development that might cause both prices to fall and Southwest to become a potential entrant, in our setting, this pattern reflects the fact that Southwest announces its entry into an airport some months before it begins flights, while our entry variables are defined by the start of actual operations.[24] As Southwest will make decisions about which routes it will serve in future quarters once its arrival into the airport is announced, this pattern is consistent with strategic explanations for why prices fall, including our limit pricing model. At the same time, however, this decline is not consistent with an actual competition story, or at least one where consumers do not substitute journeys intertemporally, because connecting service on Southwest could only provide a substitute for passengers once Southwest begins operating flights.

A second piece of evidence against the actual competition explanation comes from examining what happens to different percentiles of the price distribution. In Tables 3.5-3.7 we report estimates from specification (3.1) where we use the 25^{th}, 50^{th} and 75^{th} percentiles of the fare and yield distribution as dependent variables. The

[23] Results are similar if we instead use a dummy for any of the dominant incumbent's passengers being ticketed by another carrier which is the variable used by Goetz and Shapiro (2012). On average, 15% of routes satisfy this definition of being code-shared in Phase 1 and 35% in Phase 2.

[24] For example, Southwest announced its entry into Philadelphia on October 28, 2003, and began operations on May 9, 2004. It announced entry into Boston on February 19, 2009 and began operations on August 16, 2009 (dates from http://swamedia.com/channels/By-Category/pages/openings-closings, accessed November 10, 2015).

results reveal that when Southwest threatens entry, prices decline significantly across the fare distribution. One might expect that actual competition from connecting service on Southwest would primarily attract price-elastic leisure travelers who would otherwise buy cheaper tickets on the incumbent. In this case, one might expect to see larger price reductions for cheaper seats, whereas what is observed is that all prices fall, with larger declines at higher percentiles.[25]

A third piece of evidence against the claim that our results are driven by actual competition from connecting service on Southwest during Phase 2, we repeat the second-stage EE regressions controlling for the convenience of Southwest connections. We do so by including three additional dummies interacted with $SWPE_{m,t}$ that divide the markets into groups based on the convenience of a connection via one of Soutwest's focus airports (Baltimore, Chicago Midway, Phoenix and Las Vegas).[26] We see that the quadratic and linear coefficients in the yield regression, reported in column (2) of Table 3.8, remain statistically significant, and, in fact, contrary to what would be expected if actual competition was causing prices to fall, the coefficients on the dummy variables (not reported) indicate that prices fall most on the routes where Southwest connections are least convenient.

3.4.3 Incumbent Price Cuts to 'Lock-In' Demand

One interpretation of why incumbents cut prices when entry is threatened is that these price cuts may increase consumer loyalty. This might help the incumbent to

[25] Borenstein and Rose (1995) argue that actual competition tends to increase within-carrier fare dispersion in the airline industry for this reason, whereas here we are observing fare compression. Of course, one might argue that when the price of cheaper tickets fall, a carrier is constrained to lower more expensive business class or more flexible ticket prices in order to maintain incentive compatibility constraints as part of a second degree price discrimination scheme. However, this would not explain why higher prices fall more (both proportionately and in absolute terms) and why prices fall before Southwest actually begins offering connecting service.

[26] For each market, we calculate the total distance that would be involved in traveling via the most convenient of these focus airports and divide this distance by the non-stop round-trip distance. Markets are divided into three equally-sized groups based on this ratio.

Table 3.5: Incumbent Pricing In Response to Southwest's Actual and Potential Entry: 25th Percentile of Prices

	Phase 1		*Phase 2*		*Phase 3*	
Fare						
t_0-8	-0.039	t_0	-0.091^{**}	t_e	-0.424^{***}
	(0.031)		(0.037)		(0.074)
t_0-7	0.002	t_0+1	-0.105^{**}	t_e+1	-0.455^{***}
	(0.033)		(0.039)		(0.073)
t_0-6	-0.019	t_0+2	-0.117^{**}	t_e+2	-0.497^{***}
	(0.037)		(0.041)		(0.076)
t_0-5	-0.029	t_0+3	-0.134^{***}	t_e+3	-0.534^{***}
	(0.036)		(0.036)		(0.079)
t_0-4	0.027	t_0+4	-0.085^{**}	t_e+4	-0.547^{***}
	(0.038)		(0.040)		(0.081)
t_0-3	0.005	t_0+5	-0.099^{**}	t_e+5	-0.523^{***}
	(0.036)		(0.042)		(0.083)
t_0-2	-0.062^{*}	t_0+6-12	-0.170^{***}	t_e+6-12	-0.502^{***}
	(0.034)		(0.049)		(0.082)
t_0-1	-0.055^{*}	t_0+13+	-0.286^{***}	t_e+13+	-0.494^{***}
	(0.031)		(0.056)		(0.089)
Yield						
t_0-8	-0.014	t_0	-0.040^{**}	t_e	-0.169^{***}
	(0.012)		(0.016)		(0.04)
t_0-7	0.016	t_0+1	-0.029^{*}	t_e+1	-0.174^{***}
	(0.016)		(0.015)		(0.040)
t_0-6	-0.001	t_0+2	-0.036^{**}	t_e+2	-0.187^{***}
	(0.017)		(0.017)		(0.043)
t_0-5	-0.005	t_0+3	-0.048^{***}	t_e+3	-0.191^{***}
	(0.013)		(0.015)		(0.042)
t_0-4	0.02	t_0+4	-0.030	t_e+4	-0.195^{***}
	(0.016)		(0.020)		(0.043)
t_0-3	0.01	t_0+5	-0.035^{*}	t_e+5	-0.195^{***}
	(0.016)		(0.018)		(0.045)
t_0-2	-0.019	t_0+6-12	-0.068^{***}	t_e+6-12	-0.213^{***}
	(0.017)		(0.022)		(0.048)
t_0-1	-0.018	t_0+13+	-0.115^{***}	t_e+13+	-0.234^{***}
	(0.017)		(0.033)		(0.053)

Notes: Estimates of specification (3.1) when dependent variable is log of the 25th percentile passenger-weighted fare ("Fare") or this fare divided by the non-stop route distance ("Yield"). The adjusted R^2s are 0.75 ("Fare") and 0.78 ("Yield"). Other notes from Table 3.3 apply here.

Table 3.6: Incumbent Pricing In Response to Southwest's Actual and Potential Entry: 50th Percentile of Prices

	Phase 1		*Phase 2*		*Phase 3*	
Fare						
$t_0 - 8$	-0.029	t_0	-0.121^{***}	t_e	-0.524^{***}
	(0.034)		(0.039)		(0.090)
$t_0 - 7$	-0.020	$t_0 + 1$	-0.131^{***}	$t_e + 1$	-0.615^{***}
	(0.042)		(0.047)		(0.090)
$t_0 - 6$	-0.023	$t_0 + 2$	-0.129^{***}	$t_e + 2$	-0.650^{***}
	(0.04)		(0.042)		(0.097)
$t_0 - 5$	-0.016	$t_0 + 3$	-0.144^{***}	$t_e + 3$	-0.722^{***}
	(0.039)		(0.042)		(0.104)
$t_0 - 4$	-0.034	$t_0 + 4$	-0.163^{***}	$t_e + 4$	-0.685^{***}
	(0.041)		(0.042)		(0.104)
$t_0 - 3$	-0.024	$t_0 + 5$	-0.175^{***}	$t_e + 5$	-0.628^{***}
	(0.0385)		(0.050)		(0.107)
$t_0 - 2$	-0.107^{***}	$t_0 +$ 6-12	-0.249^{***}	$t_e +$ 6-12	-0.615^{***}
	(0.037)		(0.051)		(0.102)
$t_0 - 1$	-0.097^{**}	$t_0 + 13+$	-0.364^{***}	$t_e + 13+$	-0.612^{***}
	(0.0386)		(0.061)		(0.109)
Yield						
$t_0 - 8$	-0.018	t_0	-0.066^{***}	t_e	-0.281^{***}
	(0.017)		(0.024)		(0.065)
$t_0 - 7$	0.007	$t_0 + 1$	-0.0491^{**}	$t_e + 1$	-0.314^{***}
	(0.022)		(0.028)		(0.064)
$t_0 - 6$	0.004	$t_0 + 2$	-0.060^{***}	$t_e + 2$	-0.330^{***}
	(0.023)		(0.025)		(0.068)
$t_0 - 5$	0.011	$t_0 + 3$	-0.070^{***}	$t_e + 3$	-0.352^{***}
	(0.023)		(0.024)		(0.070)
$t_0 - 4$	-0.017	$t_0 + 4$	-0.074^{***}	$t_e + 4$	-0.352^{***}
	(0.024)		(0.024)		(0.072)
$t_0 - 3$	-0.007	$t_0 + 5$	-0.074^{***}	$t_e + 5$	-0.347^{***}
	(0.025)		(0.024)		(0.075)
$t_0 - 2$	-0.057^{**}	$t_0 +$ 6-12	-0.143^{***}	$t_e +$ 6-12	-0.367^{***}
	(0.025)		(0.033)		(0.075)
$t_0 - 1$	-0.053^{**}	$t_0 + 13+$	-0.220^{***}	$t_e + 13+$	-0.400^{***}
	(0.026)		(0.050)		(0.086)

Notes: Estimates of specification (3.1) when dependent variable is log of the median passenger-weighted fare ("Fare") or this fare divided by the non-stop route distance ("Yield"). The adjusted R^2s are 0.73 ("Fare") and 0.82 ("Yield"). Other notes from Table 3.3 apply here.

Table 3.7: Incumbent Pricing In Response to Southwest's Actual and Potential Entry: 75th Percentile of Prices

		Phase 1		*Phase 2*		*Phase 3*
Fare						
	$t_0 - 8$	-0.053	t_0	-0.135^{***}	t_e	-0.476^{***}
		(0.034)		(0.046)		(0.091)
	$t_0 - 7$	-0.063	$t_0 + 1$	-0.170^{***}	$t_e + 1$	-0.597^{***}
		(0.039)		(0.045)		(0.092)
	$t_0 - 6$	-0.064	$t_0 + 2$	-0.193^{***}	$t_e + 2$	-0.615^{***}
		(0.041)		(0.046)		(0.098)
	$t_0 - 5$	-0.055	$t_0 + 3$	-0.176^{***}	$t_e + 3$	-0.697^{***}
		(0.042)		(0.046)		(0.106)
	$t_0 - 4$	-0.042	$t_0 + 4$	-0.194^{***}	$t_e + 4$	-0.707^{***}
		(0.042)		(0.047)		(0.102)
	$t_0 - 3$	-0.0333	$t_0 + 5$	-0.190^{***}	$t_e + 5$	-0.655^{***}
		(0.0377)		(0.052)		(0.106)
	$t_0 - 2$	-0.105^{***}	t_0 + 6-12	-0.242^{***}	t_e + 6-12	-0.672^{***}
		(0.040)		(0.061)		(0.101)
	$t_0 - 1$	-0.111^{***}	t_0 + 13+	-0.388^{***}	t_e + 13+	-0.636^{***}
		(0.039)		(0.069)		(0.109)
Yield						
	$t_0 - 8$	-0.029	t_0	-0.097^{***}	t_e	-0.342^{***}
		(0.022)		(0.03)		(0.079)
	$t_0 - 7$	-0.034	$t_0 + 1$	-0.098^{**}	$t_e + 1$	-0.403^{***}
		(0.023)		(0.033)		(0.077)
	$t_0 - 6$	-0.026	$t_0 + 2$	-0.113^{***}	$t_e + 2$	-0.422^{***}
		(0.025)		(0.032)		(0.083)
	$t_0 - 5$	-0.018	$t_0 + 3$	-0.099^{***}	$t_e + 3$	-0.456^{***}
		(0.026)		(0.032)		(0.086)
	$t_0 - 4$	-0.035	$t_0 + 4$	-0.115^{***}	$t_e + 4$	-0.468^{***}
		(0.025)		(0.033)		(0.086)
	$t_0 - 3$	-0.036	$t_0 + 5$	-0.118^{***}	$t_e + 5$	-0.460^{***}
		(0.023)		(0.037)		(0.087)
	$t_0 - 2$	-0.075^{***}	t_0 + 6-12	-0.179^{***}	t_e + 6-12	-0.488^{***}
		(0.027)		(0.044)		(0.084)
	$t_0 - 1$	-0.076^{***}	t_0 + 13+	-0.289^{***}	t_e + 13+	-0.518^{***}
		(0.027)		(0.051)		(0.097)

Notes: Estimates of specification (3.1) when dependent variable is log of the 75th percentile passenger-weighted fare ("Fare") or this fare divided by the non-stop route distance ("Yield"). The adjusted R^2s are 0.81 ("Fare") and 0.84 ("Yield"). Other notes from Table 3.3 apply here.

Table 3.8: EE Reduced Form Analysis: Second Stage, Quadratic Specification

	(1) Yield	(2) Yield
$SWPE_{m,t}$	0.0139	-0.064
	(0.0147)	(0.0587)
$\widehat{\rho_m} * SWPE_{m,t}$	-0.721***	-0.555**
	(0.1358)	(0.273)
$\widehat{\rho_m}^2 * SWPE_{m,t}$	0.921***	0.722**
	(0.1938)	(0.315)
N	3622	3622

Heteroskedasticity robust Newey-West standard errors allowing for one period serial correlation and corrected for first-stage approximation error in the entry probabilities in parentheses. Column (2) includes controls for the convenience of connecting on Southwest. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

deter entry, by reducing Southwest's expected demand, and/or it might increase the incumbent's expected profits in the duopoly game that follows entry. As we noted in Table 3.7 price declines are larger on the 75th percentile than they are on the 25th, 50th or at the average. Larger declines for more expensive fares might be consistent with explanations where the incumbent lowers fares for frequent business travelers, who are more likely to buy expensive last-minute tickets, in order to try to increase their future demand (for example, by building up their accumulated miles in frequent-flyer loyalty programs). This kind of strategy could be used either to try to deter Southwest's entry, by lowering its expected demand, or to strengthen the incumbent's position should Southwest enter. However, it is not clear why the incumbent would also cut the prices of low-priced tickets, that are usually sold further ahead of the date of departure, and are targeted at infrequent, leisure travelers.

To directly test whether firms are engaging in investments in switching costs we estimate a simple nested logit demand model of the incumbent's demand, where the outside good is 'not flying' and different carriers flying the route are gathered in the single nest. Our market size measure is described in Appendix C. Our estimating equation is the standard one used with aggregate data, following Berry (1994), and given that we are focused here on understanding whether the incumbent can increase its future demand by lowering prices, we estimate the model using only (average) price and share observations for the incumbent. However, as well as the carrier's average price in the current period, we also include prices in previous quarters, and, if there is a significant loyalty effect, then we expect the coefficients on these lagged prices to also be negative. As we describe in Appendix **??**, our instruments for the current average price and the inside share are the one-quarter lagged jet fuel price, the interaction of this price and the non-stop route distance, the carrier's presence at the endpoints and a dummy for whether Southwest has entered the market. When we include price lags, we introduce appropriately lagged values of these variables as additional instruments. Our sample consists of dominant incumbent observations from the dominant incumbent sample markets in all periods, including Phase 2, when entry is threatened.

The estimated coefficients are shown in Table 3.9, where we increase the number of lagged prices included in columns (2)-(5). In the final column we only include the lagged price from the same quarter in the previous year, as there might be some travelers who tend to make trips in particular seasons. The F-statistics in the first-stage regressions (not reported) are all greater than 50. We observe that none of the coefficients on the lagged prices are statistically significant and that they vary in sign (eight are positive (i.e., the 'wrong sign'), and three are negative). In each specification the coefficient on the current price is statistically significant at conventional significance levels.

One might argue that, because of the size of the standard errors, we cannot rule that the true lagged coefficients are negative and 'large enough' to justify reductions

Table 3.9: Nested Logit Demand Estimates for the Incumbent with Lagged Prices: All Phases

	(1)	(2)	(3)	(4)	(5)	(6)
Fare ($100s)	-0.355***	-0.222*	-0.275**	-0.294**	-0.476***	-0.354***
	(0.0398)	(0.119)	(0.132)	(0.126)	(0.114)	(0.0573)
Inside Share	0.687***	0.726***	0.739***	0.692***	0.809***	0.747***
	(0.0939)	(0.0974)	(0.102)	(0.105)	(0.108)	(0.104)
Fare_{t-1}		-0.0994	-0.0419	0.0126	0.185	
		(0.109)	(0.197)	(0.199)	(0.185)	
Fare_{t-2}			0.0246	0.0194	-0.0930	
			(0.105)	(0.184)	(0.190)	
Fare_{t-3}				0.0120	0.0818	
				(0.0960)	(0.193)	
Fare_{t-4}					0.0316	0.0516
					(0.120)	(0.0422)
Observations	5,309	5,071	4,864	4,680	4,515	4,688

Notes: Specification also includes a linear time trend, carrier dummies, a dummy for whether the incumbent is a hub carrier on the route, quarter of year dummies, market characteristics (distance, distance^2, indicators for whether the route includes a leisure destination or is in a city with another major airport) and dummies for the number of competitors offering direct service. The instruments are described in the text. Robust standard errors in parentheses. ***,** ,* denote statistical significance at the 1, 5 and 10% levels respectively.

in current prices, although, of course, one would need to argue that in addition these price cuts are only profitable when entry is threatened to explain why prices fall in Phase 2 in the data. For this reason, we should point out that there seem to us to be at least two other problems with the customer loyalty story. First, it is not obvious that the types of customers, typically business travelers, that are likely to be heavily invested in frequent-flyer programs are either numerous enough or have demand that is elastic enough to be able to rationalize why a carrier would want to cut all of its prices when most of the discounts would likely go to other travelers. It would surely be more profitable to offer frequent-flyers more direct and targeted benefits such as awarding them with additional miles when they take flights, a kind of discount that would not show up in our price data. Second, it is not clear that, given its commitment to charge relatively low maximum prices, that Southwest's entry strategy would be so sensitive to whether the incumbent had been able to 'lock-up' the demand of business travelers to whom the incumbent traditionally charges high prices.

3.5 Conclusion

We have presented reduced-form evidence to explain why incumbent carriers cut prices when Southwest threatens entry on a particular airline route. This is a natural setting to study, given that it provides the largest documented effect of potential competition on prices. We provide new reduced-form evidence that a limit pricing explanation can explain why prices fall, by analyzing where the price declines are greatest, and also providing evidence against alternative explanations, such as prices being cut to boost future demand, Southwest providing actual competition through connecting service or being a side-effect of either entry-deterring capacity investments or changes in load factors.

Consistent with findings elsewhere in the literature (e.g. those of GS) we estimate that prices decline by 10-14% once Southwest becomes a potential entrant on a route (defined as operating out of both endpoints but not flying the route itself). Significantly, the price declines are similar across the fare distribution. Coupled with evidence that firms' lagged prices do not affect current period demand we are able to rule out that the price declines are being used to 'lock-in' future demand. Furthermore, we find that price declines are most pronounced on those routes with intermediate entry probabilities. Yet, we find no evidence that firms are using capacity investment to deter entry on these routes. In fact, we observe that on markets with intermediate entry probabilities (i.e. where the gain from strategically investing should be most profitable) the dominant carrier tends to carry more passengers and have higher load factors, suggesting its marginal costs are *increasing*.

In addition to ruling out various entry deterring strategies we find that incumbent firms are using code-sharing to accommodate entry. Firms enter into these agreements on those routes where Southwest's entry is most likely, suggesting firms are attempting to use code-sharing to soften post-entry competition. In our view this leaves limit pricing as the most prominent explanation for the observed price cuts. For limit pricing to be profitable Southwest's entry decision needs to be sensitive to what it believes about the incumbent's marginal cost and that the signal should communicate the distribution of the incumbent's marginal costs. The combination of price cuts being most pronounced on intermediate probability routes and that the cuts occur across the fare distribution sits neatly with these requirements.

Whilst we consider the evidence to be persuasive that incumbents are using limit pricing to deter entry we recognize that the standard framework for limit pricing involves firms competing for only two periods with fixed marginal costs. This is at odds with the price declines that we observe whereby firms are seen to cut prices repeatedly during the time Southwest remains a potential entrant. It is also unlikely that firms' marginal (and fixed and entry) costs are perfectly persistent. This suggests that it is necessary for us to build a model of dynamic limit pricing that allows for repeated interaction between an incumbent, with private information regarding its serially correlated marginal cost, and a potential entrant. The aim of constructing a model, whereby limit pricing is incentive compatible, would be to provide more persuasive evidence that limit pricing can rationalize the 14-20% price cuts observed in the data.

The next chapter of the thesis develops a dynamic limit pricing model to see whether the price declines observed in the data can be rationalized using the model, calibrated with estimates from the DB1 data on our incumbent sample.

4

Dynamic Limit Pricing with an Application to the Airline Industry

4.1 Introduction

In an environment where incumbent firms have market power and asymmetric information over market conditions incumbents may have an incentive to forestall entry by signaling to a potential entrant that entry would be unprofitable. This classic strategy of limit pricing (Milgrom and Roberts (1982) (MR)) hints at significant welfare gains from potential, as opposed, to actual entry. Yet, whilst the theory literature is well known and often studied, empirical evidence is rare.

In this paper we extend the (MR) model of limit pricing with asymmetric information, to a dynamic setting.[1] We then show that our model provides a plausible

[1] The earlier limit pricing literature assumed that a low pre-entry price might deter entry because potential entrants would view it as implying that low prices would be set post-entry, even if arguments for why this would be rational were not explicitly developed (e.g., Modigliani (1958), Gaskins (1971), Kamien and Schwartz (1971), Baron (1973) and, for a critique, Friedman (1979)). MR addressed this issue by introducing asymmetric information between the incumbent and potential entrant. Matthews and Mirman (1983) and Harrington (1986) provide early developments of the MR framework. We note that we use the term dynamic limit pricing to refer to the fact that, in our model, the incumbent may be able to set its price multiple times before the potential entrant enters. The term dynamic limit pricing has also sometimes been used to refer to the process by which an incumbent facing entry by multiple firms will change its price over time, partly to limit

explanation for why, in the 1990s and 2000s, incumbent airlines often responded to the threat of entry by Southwest by lowering their prices, and then keeping them low, even before entry actually occurred, which is part of the phenomenon commonly known as the "Southwest Effect".[2]

In the two-period MR model, an incumbent faces a potential entrant who is uninformed about some relevant aspect of the market, such as the incumbent's marginal cost. In equilibrium, the incumbent may deter entry by choosing a price that is low enough to credibly signal that the value of this variable is so low that the potential entrant's post-entry profits would not cover its entry costs. However, once the model is extended so that an incumbent can set prices in multiple periods and the potential entrant has multiple opportunities to enter, it is unclear *a priori* whether the model would give a unique prediction about how the incumbent would price (for example, does the incumbent need to set low prices in every period or only in some initial set of periods?), and indeed, in the applied literature, dynamic games with persistent asymmetric information have often been viewed as being too intractable to work with, at least using standard notions of equilibrium (Doraszelski and Pakes (2007), Fershtman and Pakes (2012)). We show that, when we allow the incumbent's private information to be positively serially correlated, but not perfectly persistent, over time, Markov Perfect Bayesian Equilibrium strategies and beliefs on the equilibrium path are unique under an application of the D1 refinement when a set of conditions on the payoffs of the incumbent hold in every period. When the unobserved variable is the incumbent's marginal cost, and this is assumed to evolve exogenously, we show that these conditions will be satisfied under some simple, and quite weak, restrictions

the growth of entrants (Gaskins (1971)).

[2] The term Southwest Effect was taken from the title of a 1993 Department of Transportation study (Bennett and Craun (1993)) which noted that many contemporary pricing trends on short-haul routes could be attributed to the presence of Southwest on a route itself or its presence on routes serving the endpoint airports.

on the static primitives of the model. In this equilibrium, the incumbent engages in limit pricing to perfectly reveal its current marginal cost in every period, so that our model predicts that the incumbent will keep prices low until entry actually occurs, at which point we assume, for simplicity, like MR, that the game changes to be one of complete information.

Having developed the model, we investigate whether it can explain the Southwest Effect. As documented by Goolsbee and Syverson (2008) (GS) and discussed in Chapter **??**, incumbent airlines lower prices by as much as 20% on airport-pair routes when Southwest serves both endpoint airports without (yet) serving the route itself. While this is a natural setting in which to consider limit pricing as it provides the largest documented case of potential competition lowering prices in any industry (Bergman (2002)), to the best of our knowledge, no one has examined whether a limit pricing story can explain what is observed in the data.

We exploit our model's tractability by calibrating it using demand and marginal cost parameters that we estimate using data from our sample markets. We show that our simple model is able to generate the size of price declines that are observed for markets with intermediate probabilities of Southwest entry. We also use our calibrated model to analyze several comparative statics as to which parameters lead to more aggressive limit pricing in equilibrium and to study the welfare effects under limit pricing versus that where the incumbent's marginal cost is observed or where the incumbent cannot credibly signal its cost.

In addition to the literature that has studied the Southwest Effect our work is related to the theory literature on games of asymmetric information. In characterizing what happens in a dynamic, finite horizon version of MR, we recursively apply the results of Mailath (1987), Mailath and von Thadden (2013) and Ramey (1996) in one-shot signaling models. Roddie (2012a) and Roddie (2012b) also take a recursive approach to solving a dynamic game of asymmetric information, focusing on the example of a quantity-setting game between two incumbents, one of whom has a privately-known marginal cost that evolves exogenously. We differ from Roddie not only in considering an entry-deterrence game, but also in the high-level theoretical conditions that we use and that, in the exogenous marginal cost version of our model, we show that these conditions will be satisfied throughout a dynamic game under some sufficient and easy-to-check conditions on static features of the model. Kaya (2009) and, in a limit pricing context, Toxvaerd (2010) consider repeated signaling models where the sender's type is fixed over time. This structure can lead to signaling only in the early periods of a game, whereas, with an evolving type, our model has repeated signaling in equilibrium.[3]

A second connected literature is the broader empirical literature on strategic investment. Recently, Snider (2009) and Williams (2012), both in the context of alleged predation by airlines, have estimated structural Ericson and Pakes (1995)-style dynamic oligopoly models with complete information (up to i.i.d. payoff shocks). One feature of these models is that there are often multiple equilibria. We differ from this literature by considering a finite horizon dynamic model with asymmetric information and explicitly establishing conditions and a refinement under which the Markov Perfect Bayesian equilibrium that we look at is unique.[4] Fershtman and Pakes (2012) consider an alternative way of incorporating persistent asymmetric information in a dynamic game, focusing on infinite horizon games with finite states and actions. They propose an alternative solution concept, Experience Based Equilibrium, under which players have beliefs about expected payoffs from their own

[3] A model where the incumbent's type is fixed would have difficulty in explaining two aspects of our empirical application. First, incumbents not only cut prices when Southwest first appears as a potential entrant, they also keep prices low even if Southwest does not initially enter. Second, and more fundamentally, if the incumbent's type is fixed then Southwest should be able to infer the incumbent's type from how it set prices *before* Southwest became a potential entrant, leaving it unclear what cutting prices once Southwest threatens entry would achieve.

[4] We do not directly estimate our model here, and leave the development of an estimation methodology to future work.

alternative actions, rather than their rivals' types. This approach may greatly reduce the memory required to store agents' beliefs in a game where agents of different types choose the same actions in equilibrium (i.e., pooling or semi-pooling). In contrast, we consider a finite horizon game with continuous actions where we can show that the equilibrium involves full separation, so that equilibrium beliefs can be handled easily.[5] In doing so, we can extend one of the classic two-period models of theoretical Industrial Organization by incorporating both dynamics and calibrate our model to show that it can generate the magnitude of price declines observed in the data.

The rest of the paper is organized as follows. Section 4.2 lays out our model of dynamic limit pricing when marginal costs are exogenous and characterizes the equilibrium. Section 4.3 describes our data, and discusses the potential applicability of our model to explaining the Southwest Effect. Section 4.4 uses data from a set of incumbent monopoly markets to estimate demand and retrieve incumbent's marginal costs. Section 4.5 shows that parameterized versions of our model can generate significant limit pricing with exogenous marginal costs. Section 4.6 concludes. Appendices contain theoretical proofs.

4.2 Model

In this section we develop a model of a dynamic entry deterrence game with serially correlated asymmetric information. To make the exposition straightforward, we focus on a game where the incumbent has a time-varying (constant) marginal cost of carrying passengers that evolves exogenously. This model is, in essence, a direct extension of MR. We develop our equilibrium concept, explain what is required for existence and uniqueness of a fully separating Markov Perfect Bayesian Equilibrium

[5] Borkovsky et al. (2014) contains a more detailed comparison of the EBE approach and the one used here.

(MPBE), and provide some simple conditions on static payoffs and outcomes under which these requirements will be satisfied. Proofs of theoretical propositions are in Appendix G and the way that we solve the model is explained in Appendix H.

4.2.1 A Dynamic Limit Pricing Model with Exogenous Marginal Costs

There is a finite sequence of discrete time periods, $t = 1, ..., T$, two long-lived firms and a common discount factor of $0 < \beta < 1$. We assume finite T so that we can apply backwards induction to prove existence and uniqueness, but, when we give numerical illustrations, T will be large and we will focus on the strategies that are (almost) stationary in the early part of the game.[6] At the start of the game, firm I is an incumbent, who is assumed to remain in the market forever, and firm E is a long-lived potential entrant. Once E enters, it will also remain in the market forever.[7] The marginal costs of the firms are $c_{E,t}$ and $c_{I,t}$. In order to economize on notation, we will assume that $c_{E,t} = c_E$ when presenting the theory but all the results hold, with all strategies conditioned on $c_{E,t}$, when $c_{E,t}$ is also serially correlated but publicly observed (see Gedge et al. (2014) for the full presentation of the theory for this case), which is the model we actually compute in our numerical illustrations. $c_{I,t}$ lies on a compact interval $[\underline{c_I}, \overline{c_I}]$ and evolves, exogenously, according to a first-order Markov process $\psi_I : c_{I,t-1} \rightarrow c_{I,t}$ with full support (i.e., $c_{I,t-1}$ can evolve to any point on the support in the next period). Note, however, that E may have a quite precise prior on $c_{I,t}$ given what it has previously observed. The conditional pdf is denoted $\psi_I(c_{I,t}|c_{I,t-1})$. We make the following assumptions.

[6] One approach in the theoretical literature (e.g., Toxvaerd (2008)) is to show properties of an infinite game by taking the $T \rightarrow \infty$ limit of finite horizon games. We could apply this type of argument in our setting.

[7] While we assume here that the incumbent and an entrant will remain in the market forever, this assumption is not necessary in that there can be a unique limit pricing equilibrium in an extended model where future exit is possible. In our dominant incumbent sample routes, there is only one route where Southwest enters and then exits before the end of our sample, while the incumbent remains in the market for at least two years after Southwest enters in 80% of cases.

Assumption 1. ***Marginal Cost Transitions***

1. *$\psi_I(c_{I,t}|c_{I,t-1})$ is continuous and differentiable (with appropriate one-sided derivatives at the boundaries).*

2. *$\psi_I(c_{I,t}|c_{I,t-1})$ is strictly increasing i.e., a higher type in one period implies higher types in the following period are more likely. Specifically, we will require that for all $c_{I,t-1}$ there is some c' such that $\frac{\partial \psi_I(c_{I,t}|c_{I,t-1})}{\partial c_{I,t-1}}|_{c_{I,t}=c'} = 0$ and $\frac{\partial \psi_I(c_{I,t}|c_{I,t-1})}{\partial c_{I,t-1}} < 0$ for all $c_{I,t} < c'$ and $\frac{\partial \psi_I(c_{I,t}|c_{I,t-1})}{\partial c_{I,t-1}} > 0$ for all $c_{I,t} > c'$. Obviously it will also be the case that $\int_{\underline{c_I}}^{\overline{c_I}} \frac{\partial \psi_I(c_{I,t}|c_{I,t-1})}{\partial c_{I,t-1}} dc_{I,t} = 0$.*

To enter in period t, E has to pay a private information sunk entry cost, κ_t, which is an i.i.d. draw from a commonly-known time-invariant distribution $G(\kappa)$ (density $g(\kappa)$) with support $[\underline{\kappa} = 0, \overline{\kappa}]$.[8]

Assumption 2. ***Entry Cost Distribution***

1. *$\overline{\kappa}$ is large enough so that, whatever the beliefs of the potential entrant, there is always some probability that it does not enter because the entry cost is too high.*

2. *$G(\cdot)$ is continuous and differentiable and the density $g(\kappa) > 0$ for all $\kappa \in [0, \overline{\kappa}]$.*

Demand is assumed to be common knowledge and fixed, although it would be straightforward to extend the model to allow for time-varying demand observed by both firms.

[8] In MR's presentation, E's entry cost is publicly observed but its marginal cost is private information, although the reverse assumption would generate the same results. In our setting, it is important that what is privately known by the potential entrant is not serially correlated, as otherwise, I would need to make inferences about its value, which would greatly complicate the solution of the model. Given that we assume that I's marginal cost is serially correlated, it seems appropriate, as well as consistent with most of the literature on dynamic entry models, to assume that it is E's entry cost that is i.i.d. In Section 4.5 we parameterize a model where c_E is serially correlated, to match the data, but observed.

Pre-Entry Stage Game

Before E has entered, so that I is a monopolist, E does not observe $c_{I,t}$. E does observe the whole history of the game to that point. The timing of the game in each of these periods is as follows:

1. I sets a price $p_{I,t}$, and receives flow profit

$$\pi_I^M(p_{I,t}, c_{I,t}) = q^M(p_{I,t})(p_{I,t} - c_{I,t}) \tag{4.1}$$

   where $q^M(p_{I,t})$ is the demand function of a monopolist. Define

$$p_I^{\text{static monopoly}}(c_I) \equiv \text{argmax}_{p_I} q^M(p_I)(p_I - c_I) \tag{4.2}$$

   The incumbent can choose a price from the compact interval $[\underline{p}, \overline{p}]$.[9]

2. E observes $p_{I,t}$ and κ_t, and then decides whether to enter (paying κ_t if it does so). If it enters, it is active at the start of the following period.

3. I's marginal cost evolves according to ψ_I.

Assumption 3. ***Monopoly Payoffs***

1. $q^M(p_I)$*, the demand function of a monopolist, is strictly monotonically decreasing in* p_I*, continuous and differentiable.*

2. $\pi_I^M(p_I, c_I)$ *has a unique optimum in price and for any* $p_I \in [\underline{p}, \overline{p}]$ *where* $\frac{\partial^2 \pi_I^M(p_I,c_I)}{\partial p_I^2} > 0$*,* $\exists k > 0$ *such that* $\left|\frac{\partial \pi_I^M(p_I,c_I)}{\partial p_I}\right| > k$ *for all* c_I*.* These assumptions are consistent, for example, with strict quasi-concavity of the profit function.

[9] All of our theoretical results would hold when the monopolist sets a quantity. The choice of strategic variable in the duopoly game that follows entry may matter, as will be explained below.

3. $\overline{p} \geqslant p_I^{\text{static monopoly}}(\overline{c_I})$ *and* $\underline{p}$ *is low enough such that no firm would choose it (for any t) even if this would prevent E from entering whereas any higher price would induce E to enter with certainty.*[10]

Post-Entry Stage Game

We assume that once E enters, marginal costs, which continue to evolve as before, are observed so there is no scope for further signaling, and we assume that a unique equilibrium in the static duopoly game is played. Static per-period equilibrium profits are $\pi_I^D(c_{I,t})$ and $\pi_E^D(c_{I,t})$, and outputs $q_I^D(c_{I,t})$ and $q_E^D(c_{I,t})$. The choice variables of the firms, which could be prices or quantities, are denoted $a_{I,t}$ and $a_{E,t}$.

Assumption 4. ***Duopoly Payoffs and Output***

1. $\pi_I^D(c_I), \pi_E^D(c_I) \geqslant 0$ *for all* c_I. This assumption also rationalizes why neither firm exits.

2. $\pi_I^D(c_I)$ *and* $\pi_E^D(c_I)$ *are continuous and differentiable in their arguments; and* $\pi_I^D(c_I)$ *(*$\pi_E^D(c_I)$*) is monotonically decreasing (increasing) in* c_I.

3. $\pi_I^D(c_I) < \pi_I^M(p_I^{\text{static monopoly}}(c_I), c_I)$ *for all* c_I.

4. $q_I^D(c_I) - q^M(p_I^{\text{static monopoly}}(c_I)) - \frac{\partial \pi_I^D(c_I)}{\partial a_E}\frac{\partial a_E^*}{\partial c_I} < 0$ *for all* c_I*, where* a_E^* *is the equilibrium price or quantity choice of the entrant in the duopoly game.*

The fourth condition implies that a decrease in marginal cost is more valuable to a monopolist than a duopolist, and it is important in showing a single-crossing condition on the payoffs of an incumbent monopolist who can signal its costs.[11]

[10] For some parameters, although not for our chosen parameters, this could require $\underline{p} < 0$. The purpose of this restriction is to ensure that the action space is large enough to allow all types to separate.

[11] Note that because demand is decreasing in price, if this condition holds when a monopolist incumbent sets the static monopoly price then it will also hold if it sets a lower limit price, a fact that is used in the proof.

The condition is easier to satisfy when the duopolists compete in prices (strategic complements), as $\frac{\partial \pi_I^D(c_I)}{\partial a_E}\frac{\partial a_E^*}{\partial c_I} > 0$ in this case, and when c_E is low relative to $\underline{c_I}$ (i.e., the potential entrant is always relatively efficient).[12] This makes sense in our empirical setting as Southwest is viewed as having had significantly lower costs than legacy carriers during our sample period.

Equilibrium

We assume that there is a unique Nash equilibrium in the post-entry complete information duopoly game.[13] Our interest is in characterizing pre-entry play. Our basic equilibrium concept is MPBE (Roddie (2012a), Toxvaerd (2008)), which requires, for each period:

- a time-specific pricing strategy for I, as a function of its marginal cost $\varsigma_{I,t} : c_{I,t} \rightarrow p_{I,t}$;
- a time-specific entry rule for E, $\sigma_{E,t}$, as a function of its beliefs about I's marginal cost, its own marginal cost and its own entry cost draw; and,
- a specification of E's beliefs about I's marginal costs given all possible histories of the game.

In equilibrium, E's entry rule should be optimal given its beliefs, those beliefs should be consistent with the evolution of I's marginal costs and I's strategy on the equilibrium path, and I's pricing rule must be optimal given what E will infer from I's price and how E will react based on these inferences.

[12] In his presentation of the two-period MR model, Tirole (1994) suggests a condition that a static monopolist produces more than a duopolist with the same marginal cost is reasonable. However, it will not hold in all models, such as one with homogeneous products and simultaneous Bertrand competition when the entrant has the higher marginal cost but it is below the incumbent's monopoly price.

[13] Existence and uniqueness of the post-entry equilibrium will depend on the particular form of demand assumed, and will hold for the common demand specifications (e.g., linear, logit, nested logit) with single product firms and linear marginal costs.

The following theorem contains our main theoretical result for this model.

Theorem 1. *Consider the following strategies and beliefs:*

In the last period, $t = T$, a monopolist incumbent will set the static monopoly price, and the potential entrant will not enter whatever price the incumbent sets.

In all periods $t < T$:

(i) E's entry strategy will be to enter if and only if its entry cost κ_t is lower than a threshold $\kappa_t^*(\widehat{c}_{I,t})$, where $\widehat{c}_{I,t}$ is E's point belief about I's marginal cost and

$$\kappa_t^*(\widehat{c}_{I,t}) = \beta[\mathbb{E}_t(\phi_{t+1}^E|\widehat{c}_{I,t}) - \mathbb{E}_t(V_{t+1}^E|\widehat{c}_{I,t})] \tag{4.3}$$

where $\mathbb{E}_t(V_{t+1}^E|\widehat{c}_{I,t})$ is E's expected value, at time t, of being a potential entrant in period $t+1$ (i.e., if it does not enter now) given equilibrium behavior at $t+1$, and $\mathbb{E}_t(\phi_{t+1}^E|\widehat{c}_{I,t})$ is its expected value of being a duopolist in period $t+1$ (which assumes it has entered prior to $t+1$).[14] *The threshold $\kappa_t^*(\widehat{c}_{I,t})$ is strictly increasing in $\widehat{c}_{I,t}$;*

(ii) I's pricing strategy, $\varsigma_{I,t} : c_{I,t} \to p_{I,t}^*$, will be the solution to a differential equation

$$\frac{\partial p_{I,t}^*}{\partial c_{I,t}} = \frac{\beta g(\kappa_t^*(c_{I,t}))\frac{\partial \kappa_t^*(c_{I,t})}{\partial c_{I,t}}\{\mathbb{E}_t[V_{t+1}^I|c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I|c_{I,t}]\}}{q^M(p_{I,t}) + \frac{\partial q^M(p_{I,t})}{\partial p_{I,t}}(p_{I,t} - c_{I,t})} \tag{4.4}$$

and an upper boundary condition $p_{I,t}^*(\overline{c_I}) = p^{\text{static monopoly}}(\overline{c_I})$. $\mathbb{E}_t[V_{t+1}^I|c_{I,t}]$ is I's expected value of being a monopolist at the start of period $t+1$ given current (t period) costs and equilibrium behavior at $t+1$. $\mathbb{E}_t[\phi_{t+1}^I|c_{I,t}]$ is its expected value of being a duopolist in period $t+1$;

(iii) E's beliefs on the equilibrium path: observing a price $p_{I,t}$, E believes that I's marginal cost is $\varsigma_{I,t}^{-1}(p_{I,t})$.

This equilibrium exists, and these strategies form the unique MPBE strategies and equilibrium-path beliefs consistent with a recursive application of the D1 refinement.

[14] We define values at the beginning of each stage. See the discussion in Appendix G for more details.

For completeness, we assume that if E observes a price which is not in the range of $\varsigma_{I,t}(c_{I,t})$ then it believes that the incumbent has marginal cost $\overline{c_I}$.

Proof. See Appendix G. □

Note that these strategies constitute the Riley Equilibrium (Riley (1979)) where the incentive compatibility constraints consistent with full separation are satisfied at minimum cost to I.

Our proof applies well-known results from the literature on one-shot signaling models. Mailath and von Thadden (2013)[15] provide conditions on a signaler's payoffs[16] under which there will only be one separating equilibrium, with the strategy characterized by a differential equation and a boundary condition. The key conditions are type monotonicity (a price cut is more costly for an incumbent with higher marginal costs), belief monotonicity (the incumbent always benefits when the entrant believes that he has lower marginal costs) and a single-crossing condition (a lower cost incumbent is always willing to cut the current price slightly more in order to differentiate itself from a higher cost type). The D1 refinement (Cho and Sobel (1990), Ramey (1996)), which restricts the inferences that the receiver can make when observing off-the-equilibrium-path actions[17], can be used to eliminate pooling equilibria given a single-crossing condition as long as any pool does not involve firms choosing the lowest possible price ($\underline{p}$).[18] We apply these results recursively, by which

[15] Mailath and von Thadden (2013) provide a generalization of Mailath (1987), expanding the set of models to which the results apply.

[16] The signaler's payoff function can be written as $\Pi^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})$ where $\widehat{c}_{I,t}$ is E's point belief about the incumbent's marginal cost when taking its period t entry decision. An alternative way of writing the payoff function that is used when ruling out pooling equilibria is $\Pi^{I,t}(c_{I,t}, \kappa'_t, p_{I,t})$ where κ'_t is the time-specific entry cost threshold used by the potential entrant.

[17] Specifically, D1 requires the receiver to place zero posterior weight on a signaler having a type θ_1 if there is another type θ_2 who would have a strictly greater incentive to deviate from the putative equilibrium for any set of post-signal beliefs that would give θ_1 an incentive to deviate.

[18] Applying D1 in a setting with repeated signaling is potentially complicated by the possibility that an off-the-equilibrium-path signal in one period could change how the receiver interprets signals in

we mean that, starting at the end of the game, we work backwards solving for equilibrium pricing and entry strategies in period $t+1$, and then using these strategies to show the continuation payoffs in period t satisfy the requirements for existence and uniqueness, meaning that we can apply these results to derive period t strategies.

The more novel part of our results is that we can show that our assumptions on static monopoly and duopoly quantities and payoffs are sufficient for the Mailath and von Thadden and Ramey requirements to be met in *every period* of the game, which makes this model particularly tractable as uniqueness can be demonstrated before the model is solved.[19] When we consider a model with endogenous marginal costs that depend on capacity investments, we can no longer rely on simple static conditions, but we are able to verify uniqueness by checking the conditions of Mailath and von Thadden (2013) and Ramey (1996) in every period for all possible capacity levels when we solve the model recursively.

4.3 Data and Sample Selection

In previous work (see Chapter 3) we analyzed several possible explanations for why we observe the Southwest Effect. Of these we felt the most persuasive explanation was that incumbent firms were using low prices to signal a low, but opaque, marginal cost. This conclusion followed from observing the largest price cuts on markets where the probability of entry was on the margin and that price cuts were observed across the distribution of fares. The markets used in that analysis are the same as is to be used in this paper and although we will provide a quick overview of the sample markets the reader is referred to Chapter 3 for more details on sample selection and

future periods. We follow Roddie (2012a) in using a recursive interpretation of D1, where we work backwards through the game, applying the refinement in each period, under the assumption that if an out-of-equilibrium action was observed, the players would expect refined equilibrium strategies to be used in subsequent periods.

[19] The static conditions are sufficient, not necessary, so our equilibrium may exist even if the conditions are violated.

reduced-form evidence.

4.3.1 Empirical Application

In contrast to the existing literature on the Southwest Effect, our contribution is to show that a limit pricing story provides an empirically plausible explanation for why incumbents lower prices when entry is threatened. We do so by showing that parameterized versions of our dynamic version of MR's limit pricing model with serially correlated exogenous marginal costs can generate price declines of the size observed in the data. It should be noted that our assumptions regarding other features of the market, such as Southwest's marginal costs, are observed is strong, and is made largely for simplicity and to avoid, like MR do, a model with two-way learning, it can be rationalized by the fact that Southwest operates a simpler point-to-point network with a homogeneous fleet of Boeing 737s, suggesting that some factors that make a legacy carrier's marginal costs at its hub opaque are likely to be less important for the potential entrant.

While we believe that our model is informative about the Southwest Effect, we note several features of airline markets that our model does not capture. For example, a carrier sells tickets on the same route at many different prices, whereas we assume that the incumbent sets a single price each period. We do show empirically, however, that the incumbent lowers prices in a similar way across the fare distribution. Our model also misses the fact that a potential entrant might be able to infer some information about marginal costs from prices set on other routes, and it also abstracts away from the fact that incumbents might be concerned about potential entrants other than Southwest, as several of these exist for most of the routes that we consider. While several heterogeneous potential entrants could certainly complicate the pricing game, it is plausible that, because of its low prices once it enters, an incumbent would be particularly willing to make short-run sacrifices to deter or delay

Table 4.1: Comparison of the Full and Dominant Incumbent Samples

	Full Sample		Dominant Incumbent Samples			
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Mean endpoint population (m.)	2.373	1.974	2.850	1.894	3.155	2.081
Round-trip distance (miles)	2,548.48	1,327.04	1,251.44	749.58	1,315.1	803.07
Constructed market size measure	27,837	44,541	62,751	66,633	47,975	61,397
Origin or destination is a:						
primary airport	0.161	0.368	0.330	0.473	0.277	0.451
secondary airport	0.301	0.459	0.321	0.469	0.354	0.482
big city	0.587	0.492	0.858	0.350	0.877	0.331
leisure destination	0.093	0.291	0.113	0.318	0.108	0.312
slot controlled airport	0.033	0.179	0.057	0.230	0.092	0.292
Number of markets	1,872		106		65	

the entry of Southwest, which was the largest low-cost carrier during our sample.

4.3.2 Data

Our data is drawn from the U.S. Department of Transportation's Origin-Destination Survey of Airline passenger Traffic (Databank 1, DB1), a quarterly 10% sample of domestic tickets and its T100 database that reports monthly carrier-segment level information on flights, capacity and the number of passengers carried on the segment (which may include connecting passengers). We aggregate the T100 data to the quarterly-level to match the structure of the DB1 data.

Using the criteria laid out in Chapter 3 there are 1,872 markets where Southwest is categorized as a potential entrant and of these 106 are served by a dominant incumbent prior to Southwest entering. In 65 markets we observe quarters where the incumbent carrier is dominant both before Southwest becomes a potential entrant and after it is a potential entrant but before it actually entered. For the remaining 41 markets entry occurs contemporaneously with establishing a presence at both endpoints. The 106 and 65 markets are listed in Appendix B

Table 4.1, which reproduces Table 3.1, provides some statistics for the full sample, and the sub-samples of 106 and 65 markets. Relative to the full sample, the dominant incumbent markets tend to be shorter with larger endpoint cities, as measured by either average population or an indicator for whether one of the endpoints meets the “big city”definition of Gerardi and Shapiro (2009). All of the markets in our dominant firm sample are shorter than the longest routes that Southwest flies non-stop (these include long, cross-country routes such as Las Vegas-Providence), so, by this metric, it is plausible that any of our routes could be entered. We also construct a variable measuring market size, which we will use when estimating demand in Section 4.5. As explained in Appendix C, this variable is constructed by estimating a generalized gravity equation using the Poisson Pseudo-Maximum Likelihood approach recommended by Silva and Tenreyro (2006), which allows us to capture the fact that the amount of travel on a route varies systematically with distance and the popularity of the particular airports.

Table 4.2 which reproduces Table 3.2, reports for the dominant incumbent markets, summary statistics for variables that vary over time, such as average prices (in Q4 2009 dollars), yield (average fare divided by route distance, a widely used metric for comparing fares across routes of different lengths) and market shares. Quarters are aggregated into three groups, which we will refer to frequently below: “Phase 1” - before Southwest is a potential entrant; “Phase 2” - when Southwest is a potential entrant but has not yet entered the route; and, “Phase 3” - after Southwest enters (if it enters during the sample). Entered markets will obviously be a selected set of markets which explains why the dominant carrier’s average capacity and passenger numbers for the Phase 3 markets are higher than for the other groups. The summary statistics are, however, consistent with Southwest’s actual entry into a market reducing prices dramatically, so that an incumbent should be willing to make investments to deter entry if it is likely that they would be effective. They are also consistent with incumbents responding to the threat of entry by lowering prices, suggesting

Table 4.2: Summary Statistics: Dominant Incumbent Sample

	Phase 1: $t < t_0$ All markets		Phase 1: $t < t_0$ Delayed entry markets		Phase 2: $t_0 \leqslant t < t_e$		Phase 3: $t \geqslant t_e$	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Incumbent Pricing								
Yield (average fare / distance)	0.510	0.317	0.525	0.313	0.439	0.297	0.292	0.143
Average fare	472.19	137.68	509.28	147.97	418.13	123.29	254.76	60.50
Southwest Pricing								
Yield (average fare / distance)		-		-	0.254	0.095	0.234	0.726
Average fare		-		-	363.41	84.38	215.98	59.42
Passenger Shares								
Incumbent	0.815	0.193	0.771	0.208	0.851	0.129	0.460	0.195
Southwest		-		-	0.016	0.030	0.486	0.210
Incumbent Capacity and Traffic								
Capacity (seats performed)	75,592	53,430	68,752	49,729	69,324	48,844	100,038	52,900
Segment passengers (incl. connecting passengers)	45,995	32,828	41,862	30,423	48,413	33,389	71,261	38,987
Load factor	0.612	0.104	0.618	0.107	0.712	0.117	0.714	0.079
Proportion passengers connecting	0.837	0.118	0.849	0.113	0.832	0.125	0.775	0.098
Code-share measure	0.086	0.211	0.124	0.257	0.257	0.352	0.188	0.245
Southwest Capacity and Traffic								
Capacity (seats performed)		-		-		-	87,629	59,887
Segment passengers (incl. connecting passengers)		-		-		-	56,403	38,478
Load factor		-		-		-	0.658	0.080
Proportion passengers connecting		-		-		-	0.702	0.102
Code-share measure		-		-		-	0.018	0.093
Number of markets	106		65		65		54	

that limit pricing may be one of these investments.[20]

4.4 Can A Dynamic Limit Pricing Model Generate Large Price Declines?

While the reduced-form evidence presented in Chapter 3 is consistent with dominant incumbents using limit pricing to try to deter entry, this raises the question of whether our model can generate the magnitude of declines that are observed. We address this question by parameterizing the model presented in Section 4.2 where marginal costs are exogenous. We estimate many of the demand and cost parameters using

[20] Yields and average fares do not vary in the same proportion across the phases, consistent with the fact that the set of markets that Southwest enters are not random with respect to the length of the route. For this reason we will look at both price metrics in the results below.

data from the dominant incumbent markets.

4.4.1 Demand Estimation

In order to calibrate the model outlined in Section 4.2 we need to estimate demand for flying, retrieve marginal costs of firms and use these to identify the degree of persistence in marginal costs and specify a functional form for the entry cost distribution. We estimate demand using the dominant incumbent sample for Phases 1 and 3 (i.e., before Southwest becomes a potential entrant, and after Southwest enters, if it enters), so that we do not use observations where we believe that limit pricing may be taking place.[21] Markets are non-directional, and we use our gravity model-based definition of market size (Appendix C), with travel on carriers other than the dominant incumbent and Southwest included in the outside good.[22] We do, however, control for the number of other carriers that fly any passengers non-stop in our specification of utility.

Viewing each carrier in the market as offering a single product, we assume the standard nested logit indirect utility specification with a single 'fly/do not fly' level of nesting (e.g., Berry (1994)):

$$u_{i,j,m,t} = \mu_j + \tau_1 T_t + \tau_{2-4} Q_t + \gamma X_{j,m,t} - \alpha p_{j,m,t} + \xi_{j,m,t} + \zeta_{i,m,t}^{FLY} + (1-\lambda)\varepsilon_{i,j,m,t}$$
$$\equiv \theta_{j,m,t} - \alpha p_{j,m,t} + \xi_{j,m,t} + \zeta_{i,m,t}^{FLY} + (1-\lambda)\varepsilon_{i,j,m,t}$$

where μ_j is a carrier j fixed effect, T_t is a time trend, and Q_t are quarter-of-year dummies. $p_{j,m,t}$ is the passenger-weighted average round-trip fare for carrier j on market m in quarter t and $\xi_{j,m,t}$ is an unobserved (to the econometrician) quality characteristic. $X_{j,m,t}$ includes an indicator for whether one of the endpoints is a hub

[21] We also restrict ourselves to Phase 1 observations where the dominant incumbent has at least 50 direct DB1 passengers and Phase 3 observations where the formerly dominant incumbent and Southwest have 50 DB1 passengers, although these restrictions have little impact on the size of our sample or the demand estimates.

[22] We use non-directional markets because entry decisions are non-directional and in our model we are assuming that incumbents set one price for each market.

for carrier j, a set of market characteristics (distance, distance^2, and indicators for whether one of the route's endpoint cities has another major airport or is a leisure destination) and a set of dummies for the number of other firms serving the market non-stop.

We estimate the model using the standard estimating equation for a nested logit model with aggregate data (Berry (1994)):

$$\log\left(\frac{s_{j,m,t}}{s_{0,m,t}}\right) = \mu_j + \tau_1 T_t + \tau_{2-4} Q_t + \gamma X_{j,m,t} - \alpha p_{j,m,t} + \lambda \log(\overline{s}_{j,m,t|FLY}) + \xi_{j,m,t}$$

where $\overline{s}_{j,m,t|FLY}$ is carrier j's share of passengers flying the route on the incumbent or Southwest and $s_{j,m,t}$ is firm j's market share.

Table **??** presents OLS and 2SLS estimates of the demand model. In the latter case we instrument for $p_{j,m,t}$ and $\overline{s}_{j,m,t|FLY}$ using the one-period lagged price of jet fuel, the interaction of the lagged jet fuel price and non-stop route distance, each carrier's average presence at the endpoint airports in that quarter[23], and, for the incumbent, whether Southwest has entered the market and, for Southwest, whether the route involves a hub for the incumbent. Controlling for endogeneity increases the estimated price elasticity of demand (the average elasticity implied by the column (2) estimates is -2.4) and, consistent with previous research, consumers are estimated to prefer traveling on a carrier with a hub at one of the endpoints. Based on the 2SLS results, we parameterize the model using $\widehat{\alpha} = -0.408$ and $\widehat{\lambda} = 0.762$, and assume a market size of 58,777, which is the mean size (during Phase 2) of the 65 markets for which we have Phase 2 observations. We set θ_j equal to 0.33 and 0.30 for the incumbent and potential entrant respectively and fix the $\xi_{j,t}$s to be equal to zero so that, ignoring price, carrier quality does not vary over time. The choice of θ for the incumbent matches the average value implied by the estimates for incumbent

[23] A carrier's presence at an airport is defined as being equal to its share of originating traffic (calculated using DB1) at the airport.

carriers in Phase 3, while the value for the potential entrant allows us to match the average difference in incumbent and Southwest qualities in Phase 3 estimated using a route-quarter fixed effects regression (not reported).

Table 4.3: Nested Logit Demand Estimates

	OLS	2SLS
Price (\$00s, $\widehat{\alpha}$)	-0.327***	-0.408***
	(0.012)	(0.037)
Inside Share ($\widehat{\lambda}$)	0.799***	0.762***
	(0.032)	(0.069)
Hub Carrier	0.184*	0.242***
	(0.030)	(0.034)
Selected Carrier Dummies		
American	-0.112**	-0.139**
	(0.054)	(0.060)
Continental	0.174**	0.263**
	(0.086)	(0.101)
Delta	-0.184***	-0.212***
	(0.043)	(0.045)
Northwest	0.296***	0.451***
	(0.052)	(0.086)
United	-0.358***	-0.330***
	(0.075)	(0.079)
US Airways	-0.027	0.048
	(0.047)	(0.058)
Southwest	-0.012	-0.082
	(0.041)	(0.050)
Observations	5,778	5,778
R^2	0.312	-

Notes: Specification also includes a linear time trend, quarter of year dummies, market characteristics (distance, distance^2, indicators for whether the route includes a leisure destination or is in a city with another major airport) and dummies for the number of competitors offering direct service. The instruments used for 2SLS are described in the text. Robust standard errors in parentheses. ***,**,* denote statistical significance at the 1, 5 and 10% levels respectively.

4.4.2 Marginal Costs

We specify that the marginal costs of firm j exist on a support $[\underline{c_j}, \overline{c_j}]$ and evolve according to a stationary AR(1) process: $mc_{j,t} = \rho^{AR} mc_{j,t-1} + (1-\rho^{AR})\left(\frac{\underline{c_j}+\overline{c_j}}{2}\right) + \varepsilon_{j,t}$. The $\varepsilon_{j,t}$ innovation is drawn from a normal distribution that is truncated so that marginal costs remain on their support and the untruncated distribution has mean zero and standard deviation σ_ε. In the long-run each carrier's expected marginal cost is equal to the mid-point of its support.

To estimate the ρ^{AR} and the average difference in marginal costs, we use the 2SLS demand estimates and back out the marginal cost for each carrier-route-quarter observation assuming that pricing in Phases 1 and 3 is characterized by the standard static monopoly/Bertrand Nash first-order conditions (recall that Phase 2 observations, where the incumbent may be limit pricing, are not used for estimation). To make comparisons across routes, we transform these marginal costs to \$s-per-mile of the non-stop route. The median and mean marginal costs calculated in this way are \$0.13/mile and \$0.16/mile. A route-quarter fixed effects regression using observations from Phase 3 (not reported) indicates that, on average, Southwest's marginal cost was \$0.055/mile lower than the incumbent's (difference significant at the 1% level).[24]

We estimate ρ^{AR} by regressing the implied per-mile marginal costs on the one-period lagged value, controlling for observed route characteristics (such as distance,

---

[24] The average total operating cost per available seat mile (CASM) reported by legacy carriers in 2010 (Dept. of Transportation Form 41, as reported by the MIT Airline Data project, http://web.mit.edu/airlinedata/www/default.html) was \$0.148. Between 1995 and 2010, the average difference in reported CASMs between legacy carriers and Southwest was \$0.037. An alternative measure, operating costs per equivalent seat mile (CESM), which adjusts for the fact that different airlines fly routes of different lengths, as we are doing when we include route fixed effects, gives an average difference of \$0.061. Therefore, while none of these statistics are trying to measure what economists would define as marginal costs, they provide some evidence that both the level of our marginal cost estimates and the estimated difference between the marginal costs of legacy carriers and Southwest have roughly the correct magnitude.

market size and the presence of slot constraints at either endpoint), carrier dummies, a full set of quarter dummies and, as a measure of a component of costs that is presumably observed by carriers, the one-quarter lagged jet fuel price interacted with route distance. Column (1) of Table 4.4 shows the estimates when we pool observations for all carriers. As the implied marginal costs are likely to be measured with error (partly because market shares and average prices are based on the limited sample of passengers included in the DB1 data), in column (2) we instrument for the lagged marginal cost with the third through fifth lags of marginal cost. The estimated persistence of marginal costs increases significantly. In the third and fourth columns, we provide 2SLS estimates for the incumbent carriers and Southwest separately. In both cases $\widehat{\rho^{AR}} \approx 0.97$ and we use this value in the calibration. We set σ_{ε}, the standard deviation of innovations to marginal cost for our representative 1,200 mile route, equal to \$36.[25]

Table 4.4: Marginal Cost Evolution Estimates

	(1) OLS All Carriers	(2) 2SLS All Carriers	(3) 2SLS Southwest
$\widehat{\text{MC per mile}}_{j,m,t-1}$	0.847***	0.974***	0.975***
	(0.1038)	(0.0123)	(0.0889)
Observations	5,432	4,544	1,603
R^2	0.813	-	-

Notes: The dependent variable is $\widehat{\text{MC per mile}}_{j,m,t}$, carrier j's computed marginal cost per mile in market m in quarter t. The specification also includes market characteristics (market size, average population, distance and a dummy for whether one of the airports is slot constrained), quarter dummies, carrier dummies and the lagged price of jet fuel interacted with route distance. In columns (2)-(4) we use the third through fifth lags of marginal cost per mile to instrument for lagged marginal costs. Robust standard errors, corrected for the uncertainty in the demand estimates, are in parentheses. ***, **, * denote statistical significance at the 1, 5 and 10% levels respectively.

In our parameterization, we consider a market with a round-trip distance of

[25] The distribution of estimated innovations has fatter tails than a normal. Our choice of standard deviation allows us to match the interquartile range of cost innovations based on the IV estimates in column (2) of Table **??**.

1,200 miles, close to the median for the dominant incumbent markets. Examples in our sample that are close to this length include Los Angeles-Salt Lake City and Minneapolis-Tulsa, and this distance implies an average marginal cost advantage for Southwest of close to \$70. We choose supports of $[\underline{c_I}, \overline{c_I}] = [\$160, \$280]$ and $[\underline{c_E}, \overline{c_E}] = [\$90, \$210]$. The width of these supports is chosen so that our assumption that $q_I^D(c_{I,t}, c_{E,t}) - q^M(p_I^{\text{static monopoly}}(c_{I,t})) - \frac{\partial \pi_I^D(c_{I,t}, c_{E,t})}{\partial a_{E,t}} \frac{\partial a^*_{E,t}}{\partial c_{I,t}} < 0$, which is the key sufficient condition for existence and uniqueness of the equilibrium we are looking at, holds for all possible costs.

4.4.3 Entry Costs

In our baseline specification we assume that E's entry costs are drawn from a truncated normal (support of [\$0, \$100 million]) where the untruncated distribution has a mean of \$55.4 million and a standard deviation of \$2 million. The mean and standard deviation parameters were selected, based on a coarse grid search, so that the average degree to which the incumbent shades prices below the static monopoly price when strategies are approximately stationary at the start of the game is similar to the size of the price cuts observed in markets with intermediate probabilities of entry when Southwest becomes a potential entrant. We still get quantitatively significant limit pricing if we consider much lower mean entry costs.

4.5 Calibration of the Model

We parameterize a model with the structure outlined in Section 4.2 using our estimates from Section 4.4. Table 4.5 reports the parameters used in the calibration.

Equilibrium Strategies

Given the assumed distribution for entry costs, we solve the model using the method described in Appendix H, assuming that $T = 200$. For this T, entry probabilities and

Table 4.5: Parameterization

ρ^I_{MC}	0.97
ρ^E_{MC}	0.97
$[\underline{c}_I, \overline{c}_I]$	$[\$160, \$280]$
$[\underline{c}_E, \overline{c}_E]$	$[\$90, \$210]$
β	0.98
κ	$LN(\$55.4m, \$2m)$

pricing strategies are stationary (to four decimal places) at the start of the game.

Figure 4.1 shows the incumbent's pricing functions at $t = 1$, for several values of c_E, together with Southwest's entry probabilities, when the (untruncated) entry cost distribution has a mean of \$55.4 million and a standard deviation of \$2 million. In equilibrium, there is substantial shading below the monopoly price for all $c_I < \overline{c_I}$ and, given the distribution of marginal costs for each firm, on average, prices are \$80.54 or 16.1% below the monopoly price, illustrating that our model can generate significant limit pricing in equilibrium. While these statistics measure limit pricing at the start of the game, it is worth noting that limit pricing can be more pronounced in later periods. For example, in the periods leading up to the time that entry effectively becomes blockaded because of the entry cost, at around $t = 150$ for our parameters, prices may be lowered by as much as 43.5% ($t = 136$), as keeping the entrant out for an additional period may substantially increase the probability that the incumbent will remain a monopolist in the future. This feature may have some relevance for the initial reaction of incumbent airlines faced with the threat of entry by Southwest if, as we noted in Section 3.3 when discussing the example of MetroJet, incumbents are aware that they may be able to slowly make operational changes that lower their marginal costs and will reduce the threat of entry in the future if they are able to deter it for the first few years when entry is threatened.

The mean entry costs used in these simulations may seem large, but they should

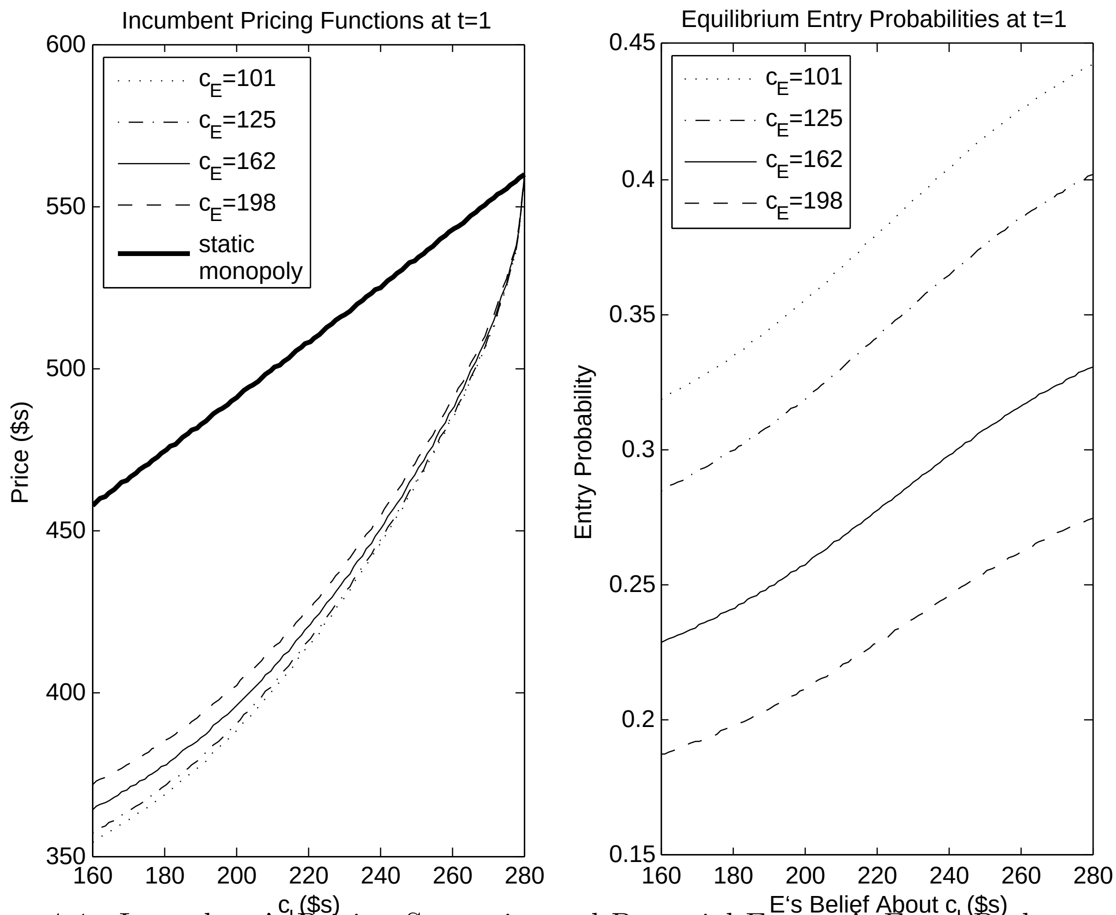

FIGURE 4.1: Incumbent's Pricing Strategies and Potential Entrant's Entry Probabilities at $t = 1$ for Different Values of c_E.

be interpreted as including the present discounted value of recurring fixed costs (for example, gate leases), that will be incurred after entry, as these have not been included elsewhere in the model, as well as the opportunity cost implied by the possibility that Southwest might have used the planes or gates for other routes.[26] We also note that one can still get significant limit pricing with lower mean entry costs: for example, if we assume mean entry costs that are only \$11.1 million (i.e.,

[26] In the case of Southwest, there is documentary evidence that these opportunity costs considerations were significant. For example, in 2011, after the end of our sample, when it reduced its service out of Philadelphia, a Southwest spokesman argued that "This is a matter of rightsizing the market for us. We felt we could reallocate those aircraft to be more productive." (Linda Loyd, "Southwest Airlines to drop Philadelphia-Pittsburgh service", McClatchy-Tribune Business News, July 27 2011). Southwest network expansions were also often timed to coincide with the final delivery of new aircraft that expanded its capacity.

80% smaller) we still generate average shading that is 8.9% of the monopoly price, compared with 16.1% under our baseline assumption.

As in the two-period MR model with a fully separating equilibrium, limit pricing is unambiguously welfare increasing (as long as limit prices are above marginal costs) in our model relative to a model where the potential entrant can observe the incumbent's marginal cost (full information), as entry decisions are the same on the equilibrium path but prices are lower before entry occurs because of limit pricing. For our baseline parameters, the gains in welfare that occur before entry occurs are quite large in percentage terms. Based on the steady-state distribution of marginal costs for this single representative market, the 16.1% shading increases expected consumer surplus by $843,000 per quarter (26.5% of the full information consumer surplus), while reducing the incumbent's profits by $154,000 (6.4%).[27]

Shading is substantial in this example because E's entry decision is relatively sensitive to its beliefs about the incumbent's marginal cost, and, consistent with the insights in EE, the average level of the entry probability is neither very high nor very low. Based on solving many games with different parameters, the degree of shading increases when there is greater serial correlation in the incumbent's marginal cost or the variance of the entry cost distribution falls.[28] Greater serial correlation leads to more shading in equilibrium for two reasons. First, from the perspective of the entrant, the incumbent's marginal cost becomes a better predictor of the entrant's profits if it enters, so the entry decision becomes more sensitive to its beliefs. Second, from the perspective of a low-cost incumbent, it also implies that if entry is deterred

[27] One can do additional welfare comparisons using the model. For example, Gedge et al. (2014) compare welfare under limit pricing and in an alternative model where the incumbent is unable to either observe or infer $c_{I,t}$ because pricing is too opaque, so that there is monopoly pricing prior to entry, but the probability of entry may increase.

[28] For example, if $\rho = 0.99$ and the standard deviation of marginal cost innovations and entry costs are $5 and $1.5 million respectively, the average degree of shading at the beginning of the game is 29%.

in the current period it is also likely to be deterred, in equilibrium, in subsequent periods. Both of these effects increase a low-cost incumbent's incentive to invest in entry deterrence by reducing the current price. A lower variance of the entry cost distribution makes the entry decision more sensitive to beliefs about the incumbent, which also increases the incumbent's incentive to limit price. On the other hand, if we either increase or decrease market size substantially then, because the average attractiveness of entry changes in the same direction, and away from the intermediate values that maximize deterrence incentives, the degree of shading falls.[29]

4.6 Conclusion

We have presented a theoretical framework for analyzing a classic form of strategic behavior, entry deterrence by setting a low price, in a dynamic setting. We show that under a standard refinement, our model has a unique Markov Perfect Bayesian Equilibrium in which the incumbent's pricing policy perfectly reveals its true type in each period. Our characterization of the equilibrium makes it straightforward to compute equilibrium pricing strategies, and we predict that an incumbent could keep prices low for a sustained period of time before entry occurs. The resulting tractability stands in contrast to the widely-held belief in the applied literature that dynamic games with persistent asymmetric information are too intractable to be used in empirical work, at least when using standard equilibrium concepts. While we do make some restrictive assumptions (for example, that there is one firm with a single piece of private information), these assumptions allow us to extend one of the most widely-cited and important two-period entry deterrence models in the literature

[29] When we move to a market size of 10,000 people (between the 5^{th} and 10^{th} percentiles of observed market sizes), the equilibrium entry probabilities are always tiny (less than 1e-10) and the incumbent's strategy involves essentially no shading. On the other hand, when we move to a market size of 200,000 (between the 90^{th} and 95^{th} percentiles), the equilibrium entry probabilities at the start of the game are greater than 0.86, and the average degree of shading is less than 4% of the monopoly price.

(Milgrom and Roberts (1982)) to allow for both dynamics and a rich, endogenous cost structure that is likely to reflect the cost structure of most industries that have dominant incumbents.

We exploit tractability to study whether our model of dynamic limit pricing can explain why incumbent carriers cut prices when Southwest becomes a potential entrant on a particular airline route. This is a natural setting to study, given that it provides the largest documented effect of potential competition on prices. We show that parameterized versions of our model can generate large price declines of the magnitude observed in the Southwest Effect. We believe that the evidence in favor of a limit pricing explanation is particularly strong when we consider the quarters when entry is first threatened, before the incumbent will have time to make operational changes that may help it to lower the probability of entry in the long-run if it can be deterred initially.

While we have explored one type of asymmetric information model and one application, we believe that our approach could be used to explore other empirical settings where asymmetric information models may explain firm behavior. For example, it is often claimed that predatory pricing is motivated by incumbents wanting to signal information on their costs or their intentions to both the current competitor and potential future competitors, and it would be interesting to compare how well this type of signaling story compares quantitatively against non-informational models of predation where the dominant incumbent makes observable investments (for instance, in capacity (Snider (2009), Williams (2012)), or learning-by-doing (Besanko et al. (2014))) that commit it to lower future costs. We would also like to explore whether there are assumptions under which a model with several incumbents could also have a tractable equilibrium with significant limit pricing behavior. This would allow us to expand our analysis in this paper to a broader set of industries and markets.

Appendix A

Drawing from the Poisson Distribution

It is important that the draws for the number of tasks is held constant when estimating usage parameters. As these parameters include the mean for the Poisson distribution each time the coefficients change so too will the mean and hence the number of tasks. This would severely inhibit the ability of the solver to maximize the objective function.

To get around this I make use of the relationship between the Exponential and Poisson distributions (as shown by Hendel (1999)). Essentially, if the Poisson describes the number of tasks that arrive per period the exponential describes the duration between tasks. Let y be the sum of K exponential random variables with density:

$$g(y) = \frac{y^{k-1}e^{y}}{\Gamma(k)} \tag{A.1}$$

and corresponding CDF:

$$\Pr(y < \lambda) = \int_0^{\lambda} \frac{y^{k-1}e^{y}}{\Gamma(k)} = 1 - \sum_{j=1}^{k-1} \frac{e^{-\lambda}\lambda^{j}}{j!} \tag{A.2}$$

where we can see the similarity to the Poisson CDF, i.e. if n is a random variable distributed Poisson with mean λ, then:

$$\Pr(n = j) = \frac{e^{-\lambda}\lambda^j}{j!} \tag{A.3}$$

so that we can rewrite equation (A.2) using (A.3) as:

$$\Pr(y < \lambda) = \sum_{j=k}^{\infty} \Pr(n = j) \tag{A.4}$$

where y is the sum of k exponentially distributed random variables and n is distributed Poisson. We can bound λ by looking at the probability that λ is between two contiguous y (each being a sum of y_x exponentials). This is given by:

$$\Pr(y_k < \lambda < y_{k+1}) = \int_0^{\lambda} \int_{\lambda - y_k} e^{-y_{k+1}} dy_{k+1} \frac{y_k^{k--1} e^{-y_k}}{\Gamma(k)} dy_k = \frac{e^{-\lambda}\lambda^k}{k!} = \Pr(n = k) \tag{A.5}$$

Therefore, in order to draw from a Poisson with mean λ I can draw exponentially distributed random variables until their partial sum exceeds λ. The number of draws before reaching λ is distributed Poisson.

This process still requires me to draw from the exponential distribution which would lead to the draws changing as λ changes (see equation (A.2)). In order to keep draws constant I therefore draw from a Uniform distribution and apply the inverse transformation of the exponential CDF.

Appendix B

List of Dominant Incumbent Markets

In the following list (*) identifies markets in the subset of 65 markets where Southwest is observed for at least some quarters as a potential, but not an actual, entrant. Carrier names reflect those at the end of the sample (so, for example, Northwest routes are listed under Delta).

American (AA): Nashville-Raleigh, Burbank-San Jose, Colorado Springs-St Louis(*), Las Vegas-San Jose, Los Angeles-San Jose(*), Reno-San Jose(*), Louisville-St Louis, San Jose-Orange County(*), St. Louis-Tampa

Alaska (AS): Boise-Portland, Boise-Seattle, Eugene-Seattle(*), Spokane-Portland(*), Spokane-Seattle, Oakland-Portland, Oakland-Seattle, Oakland-Orange Country(*), Palm Springs-Seattle(*), Palm Springs-San Francisco(*)

Continental (CO): Baltimore-Houston(Bush)(*), Cleveland-Palm Beach(*), Houston-Jackson, MS(*), Houston-Jacksonville(*), Houston-Orlando(*), Houston-Omaha(*), Houston-Palm-Beach(*), Houston-Raleigh(*), Houston-Seattle(*), Houston-Orange County(*), Houston-Tampa(*), Houston-Tulsa(*), Orlando-Omaha(*)

Delta (DL): Albany-Detroit(*), Albany-Minneapolis(*), Hartford-Minneapolis(*), Boise- Minneapolis(*), Boise-Salt Lake City, Buffalo-Detroit(*), Colorado Springs-Salt Lake City, Detroit-Milwaukee(*), Detroit-Norfolk, VA(*), Fresno-Reno(*), Fort Lauderdale-Minneapolis(*), Spokane-Minneapolis(*), Spokane-Salt Lake City, Jacksonville-LaGuardia(*), Los Angeles-Salt Lake City, LaGuardia-New Orleans(*), LaGuardia-Southwest Florida(*), Kansas City-Salt Lake City(*), Minneapolis-New Orleans(*), Minneapolis-Oklahoma City(*), Minneapolis-Omaha(*), Minneapolis-Providence(*), Minneapolis-Orange County(*), Minneapolis-Tulsa(*), Oakland-Salt Lake City, Portland-Salt Lake City, Reno-Salt Lake City(*), San Diego-Salt Lake City, Seattle-Salt Lake City(*), San Jose-Salt Lake City, Salt Lake City-Sacramento, Salt Lake City-Orange County(*)

United (UA): Hartford-Washington Dulles(*), Nashville-Washington Dulles(*),Boise-San Francisco(*), Eugene-San Francisco(*), Washington Dulles-Indianapolis(*), Washington Dulles- Jacksonville(*), Washington Dulles-LaGuardia(*), Washington Dulles-Raleigh(*), Washington Dulles-Tampa

US Airways (US): Albany-Baltimore, Hartford-Baltimore, Hartford-Philadelphia(*), Buffalo-Baltimore, Buffalo-LaGuardia(*), Buffalo-Philadelphia(*), Baltimore-Jacksonville, Baltimore-Orlando, Baltimore-Norfolk, Baltimore-Palm Beach, Baltimore-Providence, Baltimore-Tampa, Columbus-Philadelphia(*), Colorado Springs-Phoenix(*), Las Vegas-Omaha, Las Vegas-Pittsburgh, Las Vegas-Tulsa, LaGuardia-Pittsburgh(*), Manchester-Philadelphia(*), New Orleans- Philadelphia(*), Norfolk-Philadelphia(*), Omaha-Phoenix, Philadelphia-Pittsburgh, Philadelphia-Providence, Phoenix-Orange County(*), Sacramento-Orange County(*)

Other Carriers: Midwest Airlines (YX): Columbus-Milwaukee(*), Orlando-Milwaukee; Airtran (FL): Baltimore-Milwaukee; Midway Airlines (JI): Jacksonville-Raleigh(*); ATA (TZ): Chicago Midway-Philadelphia, Chicago Midway-Southwest Florida.

Appendix C

Construction of Market Size

A simple approach to defining the size of an airline market is to assume that it is proportional to the arithmetic or geometric average population of the endpoint cities (e.g., Berry and Jia (2010)). However, the number of passengers traveling on a route also varies systematically with distance, time and the number of people who use the particular airports concerned.[1] Recognizing this fact, like Benkard et al. (2010) amongst others, we try to create a better measure of market size, that we use when estimating demand in Section 4.5 (see also Appendices **??** and **??**) and also as one of the variables, in addition to average endpoint population, that can predict the probability of entry by Southwest in Section 3.3.

We estimate a generalized gravity equation using our full sample, where the expected number of passengers traveling on a route is a function of time, distance and the number of originating and final destination passengers at both of the endpoint airports as well as interactions between these variables and distance. The originating

[1] This can reflect either the fact that customers in some cities may be able to choose between multiple airports, which may be more or less convenient, but also that some destinations, such as vacation destinations, receive more visitors than would be expected based on their populations.

and destination variables are measured in the first quarter of our data (Q1 1993) in order to avoid potential problems arising from the fact that they will be affected by the airport entry of Southwest, and incumbents' responses to Southwest entry. We specify

$$\mathbb{E}\left[\text{Passengers}_{o,d,t}\right] = \exp\left\{\begin{array}{c} \beta_0 + \beta_1 Q_t + \beta_2 \log(\text{distance}_{o,d}) + \beta_2 \log(\text{distance}^2_{o,d}) + ... \\ \sum_{j=\{o,d\}} \beta_{3,j} \log(\text{originating}_{j,1993}) + \beta_{4,j} \log(\text{originating}^2_{j,1993}) + ... \\ \sum_{j=\{o,d\}} \beta_{5,j} \log(\text{destination}_{j,1993}) + \beta_{6,j} \log(\text{destination}^2_{j,1993}) + ... \\ \text{interactions between } \log(\text{distance}) \\ \text{and originating and destination variables} \end{array}\right\}$$

where o is the origin airport, d is the destination airport and Q_t are quarter dummies. $\text{Passengers}_{o,d,t}$ is the number of DB1 passengers with itineraries in either direction on the route, independent of whether they use direct or connecting service.[2] The specification is estimated using the Poisson Pseudo-Maximum Likelihood estimator, as suggested by Silva and Tenreyro (2006), because estimates from a log-linearized regression will be inconsistent when the residuals are heteroskedastic. The estimates on several coefficients are shown in Table C.1.

With the estimates in hand, we calculate the predicted value of the number of passengers for each market-quarter and then form our estimate of market size by multiplying this estimate by 3.5, so that, on average, the market share of all carriers combined (as a share of the potential market) is between 25% and 40%. Based on this measure, the median-sized route in our 106 dominant incumbent sample is Salt Lake City-Orange County where Delta is the dominant incumbent (6,806 DB1 people, or 68,060 people accounting for the fact DB1 is a 10% sample).

[2] A return passenger counts as 1, and a one-way only passenger counts as 0.5.

Table C.1: Selected Coefficients from the Gravity Equation Used to Estimate Market Size

	DB1 Passengers
log(Distance)	19.07***
	(0.313)
log(Distance)^2	-1.102***
	(0.0218)
log(Final Destination$_{o,1993}$)	0.028***
	(0.003)
log(Final Destination$^2_{o,1993}$)	-0.0017***
	(0.00017)
log(Final Destination$_{d,1993}$)	30.92***
	(0.148)
log(Final Destination$^2_{d,1993}$)	-1.655***
	(0.0076)
log(Originating$_{o,1993}$)	-21.21***
	(0.176)
log(Originating$^2_{o,1993}$)	1.274***
	(0.0091)
log(Originating$_{d,1993}$)	0.422***
	(0.0042)
log(Originating$^2_{d,1993}$)	-0.0221***
	(0.0002)
Observations	166,072
Pseudo-R^2	0.818

Note: *** denotes statistical significance at the 1% level.

Appendix D

Balance Table

This Appendix provides a 'balance table' for the dominant firm sample, where we divide markets into three groups based on the probability of Southwest entry implied by the estimated EE first-stage probit (see Appendix E). For each market, we first calculate the mean of the variable across Phase 1 observations (i.e., before Southwest is a potential entrant), and the reported means are averages across these market-level observations. Standard deviations are in parentheses and the right-hand columns present p-values from tests that the means of the variables are the same across the three groups.

In most cases, the mean values for the intermediate probability of entry markets lie between those for the low and high probability markets. In other cases, for example, average endpoint population, we cannot reject the hypothesis that the means for the three different groups are the same. In the case of the load factor, the value for intermediate entry probability markets is slightly lower than for the high entry probability markets, but the size of the difference is quite small, and we cannot reject the hypothesis that the population means for the intermediate and

Table D.1: Balance Table for Dominant Firm Sample

Variable Description	Market Probability of WN Entry Low	Intermediate	High	p-value for 2-Sided Test of Equality of Means Low and Int.	Int. and High	All Three
Market entered by	0.200	0.457	0.861	0.022**	0.000***	0.000***
Southwest (dummy)	(0.406)	(0.505)	(0.351)			
Incumbent is a legacy	0.914	0.771	0.889	0.035**	0.192	0.191
carrier (dummy)	(0.284)	(0.426)	(0.318)			
Non-stop Distance	1,745.9	1,139.3	879.7	0.001***	0.083*	0.000***
(roundtrip)	(743.3)	(730.2)	(481.3)			
Market Size	2,408.4	3,458.1	7,707.5	0.035**	0.000***	0.000***
	(1,559.5)	(2,413.6)	(4,987.6)			
Average endpoint	2,793,204	2,466,516	3,279,128	0.441	0.084*	0.213
city population	(1,729,275)	(1,797,930)	(2,093,279)			
One or both endpoint airport is hub	0.829	0.600	0.583	0.480	0.888	0.051*
for dominant incumbent	(0.382)	(0.497)	(0.500)			
One or both endpoint cities is	0.486	0.571	0.694	0.401	0.289	0.203
multi-airport market	(0.507)	(0.502)	(0.467)			
One or both endpoints is a	0.057	0.057	0.222	1.000	0.046**	0.039**
leisure destination	(0.236)	(0.236)	(0.421)			
Phase 1 route HHI	0.792	0.904	0.944	0.001***	0.127	0.000***
	(0.138)	(0.128)	(0.097)			
Phase 1 proportion of traffic	0.848	0.845	0.817	0.906	0.348	0.562
making connections	(0.112)	(0.098)	(0.141)			
Phase 1 load factor	0.661	0.583	0.591	0.003***	0.743	0.002***
	(0.097)	(0.107)	(0.093)			
Direct Fare ($)	526.59	450.09	433.84	0.022**	0.594	0.008***
	(147.00)	(136.63)	(117.80)			
Number of markets	35	35	36			

high probability markets are equal.

Appendix E

EE Estimation

As outlined in Section 3.3, the EE approach is implemented in two stages. The first stage, which we describe in more detail here, involves the estimation of a probit model using the full sample of 1,872 markets to predict the probability of entry by Southwest at the route-level. There is one observation per market, and the dependent variable ($\text{Entry4}_{m,t}$) is equal to 1 if Southwest enters market m within four quarters of becoming a potential entrant at time t (recall that Southwest becomes a potential entrant when it has operations at both endpoints).

$$\Pr(Entry4_{m,t}|X, t) = \Phi(\tau_t + \alpha X_{m,t})$$

where τ_t contains a full set of quarter dummies. The explanatory variables X_m contain the following market characteristics:

- Distance: round-trip distance between the endpoint airports (also Distance^2);
- Long Distance: a dummy that is equal to 1 for markets with a round-trip distance greater than 2,000 miles;

- Average Pop.: geometric average population for the endpoint MSAs (also Average Pop.^{2});
- Market Size: our estimated market size based on our gravity model described in Appendix C (also Market Size^2). The size is measured in the quarter when Southwest becomes a potential entrant;
- Slot: a dummy that is equal to 1 if either endpoint airport is a slot-controlled airport;
- Leisure Destination: a dummy that is equal to 1 if either endpoint city is a leisure destination as defined by Gerardi and Shapiro (2009);
- Big City: a dummy that is equal to 1 if either endpoint city is a large city, following the population-based definition of Gerardi and Shapiro (2009);
- Southwest Alternate Airport: a dummy equal to 1 in cases where Southwest already serves one of the endpoint airports from an airport that is in the same city as the other endpoint airport;
- HHI: the HHI, based on passenger numbers, for the route in the quarter that Southwest became a potential entrant.

For each of the endpoints separately, we also include:

- Primary Airport: a dummy equal to 1 for the largest airport (measured by passenger traffic in 2012) in a multiple airport city;
- Secondary Airport: a dummy equal to 1 for an airport other than the largest in a multiple airport city;

- Incumbent Presence: the average proportion of all passenger originations accounted for by the incumbents on route m at the airport in the quarter Southwest became a potential entrant (also Incumbent Presence^2);

- Southwest Presence: the proportion of all passenger originations accounted for by Southwest at the airport in the quarter it became a potential entrant (also Southwest Presence^2).

The results are reported in Table E.1.

Table E.1: Probit Model of Southwest's Entry

	Entry by Southwest Within Four Quarters	
Distance	-0.668***	
	(0.204)	
$Distance^2$	0.0385	
	(0.0344)	
Long Distance	-0.0414	
	(0.181)	
Average Pop.	-0.0952	
	(0.109)	
Average $Pop.^2$	0.0117*	
	(0.00637)	
Market Size	0.237***	
	(0.0404)	
Market $Size^2$	-0.00745***	
	(0.00165)	
Slot	-1.801***	
	(0.543)	
Leisure Destination	1.003***	
	(0.174)	
Big City	-0.0134	
	(0.146)	
Southwest Alternate Airport	-0.185	
	(0.195)	
HHI	0.541***	
	(0.202)	
Airport-Specific Variables	Origin	Destination
Primary Airport	0.688***	0.558***
	(0.254)	(0.211)
Secondary Airport (origin)	0.542**	0.0383
	(0.237)	(0.233)
Incumbent Presence	2.174	-4.076
	(1.743)	(1.683)
Incumbent $Presence^2$	-2.085	6.253
	(1.683)	(7.521)
Southwest Presence	2.455**	0.187
	(0.983)	(1.045)
Southwest $Presence^2$	-2.245**	-0.0427
	(0.940)	(1.101)
Observations	1,872	
Pseudo-R^2	0.372	

Appendix F

EE Analysis: Percentiles

We can also repeat the EE analysis, which in Section 3.4 used average prices and yields, using the percentiles. Corresponding to the column (1) of Table 3.4, which showed results for the average yield, Table F.1 shows the results for the percentiles of the yield distribution.

Table F.1: Second Stage Ellison and Ellison Analysis with Percentiles of the Yield Distribution

	(1) 25^{th} percentile	(2) 50^{th} percentile	(3) 75^{th} percentile
$SWPE_{m,t}$	0.0255	0.0439	-0.0023
	(0.0283)	(0.0375)	(0.0501)
$\widehat{\rho_m} \times SWPE_{m,t}$	-0.676**	-0.954***	-0.939**
	(0.259)	(0.329)	(0.359)
$\widehat{\rho_m}^2 \times SWPE_{m,t}$	0.835***	1.157***	1.212**
	(0.316)	(0.428)	(0.500)
Observations	3,622	3,622	3,622

Notes: Notes from Table 3.4 apply here.

Appendix G

Proof of Theorem 1

In this Appendix, we prove that the strategies described in Theorem 1 form a fully separating Markov Perfect Bayesian Equilibrium that is unique under a recursive application of the D1 Refinement. The proof uses induction and makes extensive use of theoretical results for one-shot signaling games from Mailath and von Thadden (2013) and Ramey (1996). Readers should consult Gedge et al. (2014) for a version of the proof assuming that the marginal cost of the entrant is time-varying but observed, which is the model solved in Section 4.5. It is essentially identical to the current proof with additional notation.

G.0.1 Notation and the Definition of Values

At many points in the proof we will make use of notation indicating expectations of a firm's value in a future period, e.g., $\mathbb{E}_t[V^E_{t+1}|\widehat{c}_{I,t}]$. We will use several conventions.

1. $\phi^E_t(c_{I,t})$ denotes E's expected present discounted future value when it is a duopolist at the beginning of period t, and I's marginal cost is $c_{I,t}$. Under our assumption that duopolists use unique static Nash equilibrium strategies in a complete information game, $\phi^E_t(c_{I,t})$ is uniquely defined.

2. $\phi_t^I(c_{I,t})$ denotes I's expected present discounted future value when it is a duopolist at the beginning of period t, and its marginal cost is $c_{I,t}$. Under our assumption that duopolists use unique static Nash equilibrium strategies in a complete information game, $\phi_t^I(c_{I,t})$ is uniquely defined.

3. $V_t^I(c_{I,t})$ denotes I's expected present discounted future value when it is an incumbent monopolist at the beginning of period t, and its marginal cost is $c_{I,t}$. κ_t is not known when the value is defined, so that the value is the expectation over the different possible values of κ_t. This value will be dependent on the pricing strategy that I will use in period t, E's period t entry strategy and the strategies of both firms in future periods.

4. $V_t^E(c_{I,t})$ denotes E's expected present discounted future value when it is a potential entrant at the beginning of period t, and I's marginal cost is $c_{I,t}$. Of course, E does not know $c_{I,t}$ at the moment when this value is being defined (i.e., prior to I choosing a price) but defining values in this way is convenient because it both defines the value of both firms at the same moment each period (the beginning) and economizes on the amount of notation. κ_t is not known when the value is defined, so that the value is the expectation over the different possible values of κ_t.

When we write ϕ_t^E, ϕ_t^I, V_t^E or V_t^I to economize on notation, their dependence on $c_{I,t}$, or the entrant's beliefs about $c_{I,t}$, should be understood. For example, $\mathbb{E}_t[V_{t+1}^E|\widehat{c}_{I,t}]$ is the expected value of E as a potential entrant at the start of period $t+1$ given a belief that $c_{I,t}$ is exactly $\widehat{c}_{I,t}$. As in this example, when E has point beliefs we will denote the believed value as $\widehat{c}_{I,t}$. If E does not have a point belief, we will denote their density as $q(\widetilde{c_{I,t}})$ and assume that only values on the interval $[\underline{c_I}, \overline{c_I}]$ can have positive density.

G.0.2 Useful Lemmas

We will make frequent use of several results:

Lemma 2. *Suppose that $f(x)$ is a strictly positive function, $g(x|w)$ is a strictly positive conditional pdf on $x, w \in [\underline{x}, \overline{x}]$. Further suppose that (i) for a given value of w $\exists x' \in (\underline{x}, \overline{x})$ such that $\frac{\partial g(x'|w)}{\partial w} = 0$, $\frac{\partial g(x|w)}{\partial w} < 0$ for $\forall x < x'$ and $\frac{\partial g(x|w)}{\partial w} > 0$ for $\forall x > x'$; and, (ii) $k \equiv \int_{\underline{x}}^{\overline{x}} f(x) \frac{\partial g(x|w)}{\partial w} dx$. If $\forall x$, $\frac{\partial f(x)}{\partial x} > 0$ then $k > 0$. On the other hand, if $\forall x$, $\frac{\partial f(x)}{\partial x} < 0$ then $k < 0$.*

Proof.

$$
\begin{aligned}
k &\equiv \int_{\underline{x}}^{\overline{x}} f(x) \frac{\partial g(x|w)}{\partial w} dx \\
&= \int_{\underline{x}}^{x'} f(x) \frac{\partial g(x|w)}{\partial w} dx + \int_{x'}^{\overline{x}} f(x) \frac{\partial g(x|w)}{\partial w} dx \\
&> f(x') \left\{ \int_{\underline{x}}^{x'} \frac{\partial g(x|w)}{\partial w} dx + \int_{x'}^{\overline{x}} \frac{\partial g(x|w)}{\partial w} dx \right\} = 0 \text{ if } \frac{\partial f(x)}{\partial x} > 0 \\
\text{or } &< f(x') \left\{ \int_{\underline{x}}^{x'} \frac{\partial g(x|w)}{\partial w} dx + \int_{x'}^{\overline{x}} \frac{\partial g(x|w)}{\partial w} dx \right\} = 0 \text{ if } \frac{\partial f(x)}{\partial x} < 0
\end{aligned}
$$

□

There are several useful corollaries of Lemma 2.

Corollary 3. *Suppose that $\phi^E_{t+1}(c_{I,t+1}) > V^E_{t+1}(c_{I,t+1})$,*

$\frac{\partial\{\phi^E_{t+1}(c_{I,t+1})-V^E_{t+1}(c_{I,t+1})\}}{\partial c_{I,t+1}} > 0$ *for all* $c_{I,t+1}$ *and* $\frac{\partial\psi_I(c_{I,t+1}|\widehat{c}_{I,t})}{\partial\widehat{c}_{I,t}}$ *satisfies Assumption 1, then*

$$\frac{\partial\mathbb{E}_t[\phi^E_{t+1}|\widehat{c}_{I,t}]}{\partial\widehat{c}_{I,t}} - \frac{\partial\mathbb{E}_t[V^E_{t+1}|\widehat{c}_{I,t}]}{\partial\widehat{c}_{I,t}} = \int_{\underline{c_I}}^{\overline{c_I}} \left[\phi^E_{t+1}(c_{I,t+1}) - V^E_{t+1}(c_{I,t+1})\right] \frac{\partial\psi_I(c_{I,t+1}|\widehat{c}_{I,t})}{\partial\widehat{c}_{I,t}} dc_{I,t+1} > 0.$$

Corollary 4. *Suppose that* $V^I_{t+1}(c_{I,t+1}) > \phi^I_{t+1}(c_{I,t+1})$,
$\frac{\partial\{V^I_{t+1}(c_{I,t+1})-\phi^I_{t+1}(c_{I,t+1})\}}{\partial c_{I,t+1}} < 0$ *for all* $c_{I,t+1}$ *and* $\frac{\partial\psi_I(c_{I,t+1}|c_{I,t})}{\partial c_{I,t}}$ *satisfies Assumption 1, then*

$$\frac{\partial\mathbb{E}_t[V^I_{t+1}|c_{I,t}]}{\partial c_{I,t}} - \frac{\partial\mathbb{E}_t[\phi^E_{t+1}|c_{I,t}]}{\partial c_{I,t}} = \int_{\underline{c_I}}^{\overline{c_I}} \left[V^I_{t+1}(c_{I,t+1}) - \phi^I_{t+1}(c_{I,t+1})\right] \frac{\partial\psi_I(c_{I,t+1}|c_{I,t})}{\partial c_{I,t}} dc_{I,t+1} < 0.$$

A further, very straightforward, result that will be referred to frequently is:

Lemma 5. *(a) Suppose that* $\phi^E_{t+1}(c_{I,t+1}) > V^E_{t+1}(c_{I,t+1})$ *for all* $c_{I,t+1}$ *and* ψ_I *satisfies Assumption 1, then*

$$\mathbb{E}_t[\phi^E_{t+1}|q(\widetilde{c_{I,t}})] - \mathbb{E}_t[V^E_{t+1}|q(\widetilde{c_{I,t}})] =$$

$$\int_{\underline{c_I}}^{\overline{c_I}}\int_{\underline{c_I}}^{\overline{c_I}} \left\{ \begin{array}{c} \left[\phi^E_{t+1}(c_{I,t+1}) - V^E_{t+1}(c_{I,t+1})\right] \times ... \\ \psi_I(c_{I,t+1}|\widetilde{c_{I,t}})q(\widetilde{c_{I,t}}) \end{array} \right\} dc_{I,t+1} d\widetilde{c_{I,t}} > 0$$

including the case where E has a point belief about I's marginal cost as a special case; and,
(b) suppose that $V^I_{t+1}(c_{I,t+1}) > \phi^I_{t+1}(c_{I,t+1})$ *for all* $(c_{I,t+1})$ *and* ψ_I *satisfies Assumption 1, then*

$$\mathbb{E}_t[V^I_{t+1}|c_{I,t}] - \mathbb{E}_t[\phi^I_{t+1}|c_{I,t}] =$$

$$\int_{\underline{c_I}}^{\overline{c_I}} \left[V^I_{t+1}(c_{I,t+1}) - \phi^I_{t+1}(c_{I,t+1})\right] \psi_I(c_{I,t+1}|c_{I,t}) dc_{I,t+1} > 0$$

Proof. Follows immediately from the assumptions as $\psi_I(c_{I,t+1}|\widehat{c}_{I,t}) > 0$ for all costs on $\left[\underline{c_I}, \overline{c_I}\right]$. □

G.0.3 Outline

Our proof uses induction. We first show that if the value functions of both firms satisfy several properties at the start of period $t+1$ then, together with our Assumptions 1-4, it follows that the unique equilibrium strategies in period t satisfying the D1 refinement will be those described in Theorem 1. We then show that this result implies that the value functions at the start of period t will have the same set of properties. Finally, we show that the value functions at the start of the last period satisfy these properties, which is straightforward.

G.0.4 Proof for Period t Given Value Function Properties at $t+1$

We will assume that the entrant's value functions as defined at the start of period $t+1$ have the following properties:

E1^{t+1}. $\phi^E_{t+1}(c_{I,t+1}) > V^E_{t+1}(c_{I,t+1})$; and

E2^{t+1}. $\phi^E_{t+1}(c_{I,t+1})$ and $V^E_{t+1}(c_{I,t+1})$ are uniquely defined functions of $c_{I,t+1}$, and do not depend on κ_t or any earlier values of κ;

E3^{t+1}. $\phi^E_{t+1}(c_{I,t+1})$ and $V^E_{t+1}(c_{I,t+1})$ are continuous and differentiable in their arguments; and

E4^{t+1}. $\frac{\partial[\phi^E_{t+1}(c_{I,t+1})]}{\partial c_{I,t+1}} > \frac{\partial[V^E_{t+1}(c_{I,t+1})]}{\partial c_{I,t+1}}$

Potential Entrant Strategy in Period t

E will compare its expected continuation value if it enters, $\mathbb{E}_t[\phi^E_{t+1}|\widehat{c}_{I,t}]$ if it has a point belief and otherwise $\mathbb{E}_t[\phi^E_{t+1}|q(\widetilde{c_{I,t}})]$, less its entry cost, κ_t, with its expected continuation value if it does not enter, $\mathbb{E}_t[V^E_{t+1}|\widehat{c}_{I,t}]$ or $\mathbb{E}_t[V^E_{t+1}|q(\widetilde{c_{I,t}})]$. By E2^{t+1} these continuation values do not depend on κ_t or earlier entry costs, so that E's optimal entry strategy will be a period-specific threshold rule in its entry cost. Specifically, E will enter if and only if

$$\kappa_t < \kappa^*_t(\widehat{c}_{I,t}) = \beta\left\{\mathbb{E}_t[\phi^E_{t+1}|\widehat{c}_{I,t}] - \mathbb{E}_t[V^E_{t+1}|\widehat{c}_{I,t}]\right\}$$

if E has a point belief $\widehat{c}_{I,t}$; and otherwise its entry strategy will be to enter if and only if

$$\kappa_t < \kappa_t^*(q(\widetilde{c_{I,t}})) = \beta \left\{ \mathbb{E}_t[\phi_{t+1}^E | q(\widetilde{c_{I,t}})] - \mathbb{E}_t[V_{t+1}^E | q(\widetilde{c_{I,t}})] \right\}$$

To derive the incumbent's strategy we also need to show that the threshold has certain properties. Specifically, we need it to be the case that $\kappa_t^* > \underline{\kappa} = 0$ and $\kappa_t^* < \overline{\kappa}$; and, that if E has a point belief, its threshold κ_t^* is continuous and differentiable and strictly increasing in $\widehat{c}_{I,t}$. $\kappa_t^* > \underline{\kappa} = 0$ follows from combining E1^{t+1} and Lemma 5 (a). $\kappa_t^*(\widehat{c}_{I,t})$ will be continuous and differentiable if $\phi_{t+1}^E(c_{I,t+1})$ and $V_{t+1}^E(c_{I,t+1})$ are continuous and differentiable (E3^{t+1}), and ψ_I is continuous and differentiable (Assumption 1). $\kappa_t^*(\widehat{c}_{I,t})$ is strictly increasing in $\widehat{c}_{I,t}$ if $\frac{\partial \mathbb{E}_t[\phi_{t+1}^E | \widehat{c}_{I,t}]}{\partial \widehat{c}_{I,t}} - \frac{\partial \mathbb{E}_{t-1}[V_{t+1}^E | \widehat{c}_{I,t}]}{\partial \widehat{c}_{I,t}} > 0$, which follows from E4^{t+1} and Corollary 3.

Incumbent Strategy in Period t

Existence of a Unique Separating Signaling Strategy To show the existence of a unique separating strategy for the incumbent we will rely on Theorem 1 of Mailath and von Thadden (2013), which is a useful generalization of the results in Mailath (1987). This theorem imposes conditions on the incumbent's 'signaling payoff function' $\Pi^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})$ where, in this application, the first argument is the incumbent's marginal cost, the second argument is E's (point) belief about the I's marginal cost, and $p_{I,t}$ is the price that I sets.

Theorem 6. [Based on Mailath and von Thadden (2013)] *If (MT-i)* $\Pi^{I,t}(c_{I,t}, c_{I,t}, p_{I,t})$ *has a unique optimum in* $p_{I,t}$, *and for any* $p_{I,t} \in [\underline{p}, \overline{p}]$ *where* $\Pi_{33}^{I,t}(c_{I,t}, c_{I,t}, p_{I,t}) > 0$, *there* $\exists k > 0$ *such that* $\left|\Pi_3^{I,t}(c_{I,t}, c_{I,t}, p_{I,t})\right| > k$ *for all* $c_{I,t}$*; (MT-ii)* $\Pi_{13}^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t}) \neq 0$ *for all* $(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})$*; (MT-iii)* $\Pi_2^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t}) \neq 0$ *for all* $(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})$*; (MT-iv)* $\frac{\Pi_3^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})}{\Pi_2^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})}$ *is a monotone function of* $c_{I,t}$ *for all* $\widehat{c}_{I,t}$ *and all* $p_{I,t}$ *below the static* *monopoly price; (MT-v)* $\overline{p} \geqslant p^{\text{static monopoly}}(\overline{c_I})$ *and* $\Pi^{I,t}(\underline{c_{I,t}}, \underline{c_{I,t}}, \underline{p}) < \max_p \Pi^{I,t}(\underline{c_{I,t}}, \overline{c_{I,t}}, p)$*, then* I*'s period* t *unique separating pricing strategy is differentiable on the interior of* $[\underline{c_I}, \overline{c_I}]$ *and satisfies the differential equation*

$$\frac{\partial p^*_{I,t}}{\partial c_{I,t}} = -\frac{\Pi_2^{I,t}}{\Pi_3^{I,t}}$$

with boundary condition that $p^*_{I,t}(\overline{c_I}) = p^{\text{static monopoly}}(\overline{c_I})$*.*

We now show that the conditions (MT-i)-(MT-v) hold assuming that

I1^{t+1}. $V^I_{t+1}(c_{I,t+1}) > \phi^I_{t+1}(c_{I,t+1})$;

I2^{t+1}. $V^I_{t+1}(c_{I,t+1})$ and $\phi^I_{t+1}(c_{I,t+1})$ are continuous and differentiable; and,

I3^{t+1}. $\frac{\partial V^I_{t+1}(c_{I,t+1})}{\partial c_{I,t+1}} < \frac{\partial \phi^I_t(c_{I,t+1})}{\partial c_{I,t+1}}$

as well as the conditions on E's period t entry threshold that were derived above.

Condition (MT-v) is simply a condition on the support of prices, with the second part requiring that $\underline{p}$ is so low that I would always prefer to set some higher price even if this resulted in E having the worst (i.e., highest) possible beliefs about I's marginal cost whereas setting price $\underline{p}$ would have resulted in E having the best (i.e., lowest) possible beliefs. This is implied by Assumption 3.

The signaling payoff function is defined as

$$\Pi^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t}) = q^M(p_{I,t})(p_{I,t} - c_{I,t}) + ...$$

$$\beta((1 - G(\kappa^*_t(\widehat{c}_{I,t})))\mathbb{E}_t[V^I_{t+1}|c_{I,t}] + G(\kappa^*_t(\widehat{c}_{I,t}))\mathbb{E}_t[\phi^I_{t+1}|c_{I,t}])$$

where $G(\kappa^*_t(\widehat{c}_{I,t}))$ is the probability that E enters given its entry strategy.

Condition (MT-i): $\Pi^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})$ only depends on $p_{I,t}$ through the static monopoly profit function $\pi^M_{I,t} = q^M(p_{I,t})(p_{I,t} - c_{I,t})$. The assumptions on the monopoly profit function in Assumption 3 therefore imply that (MT-i) is satisfied.

Condition (MT-ii): Differentiation of $\Pi^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})$ gives

$$\Pi_{13}^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t}) = -\frac{\partial q^M(p_{I,t})}{\partial p_{I,t}} \tag{G.1}$$

$\Pi_{13}^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t}) \neq 0$ for all $(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})$ because monopoly demand is strictly downward sloping on $[\underline{p}, \overline{p}]$ (Assumption 3).

Condition (MT-iii): Differentiating $\Pi^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})$ gives

$$\Pi_2^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t}) = -\beta g(\kappa_t^*(\widehat{c}_{I,t}))\frac{\partial \kappa_t^*(\widehat{c_{I,t}})}{\partial \widehat{c}_{I,t}} \left\{\mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]\right\} \tag{G.2}$$

$\Pi_2^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t}) \neq 0$ for all $(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})$ as $g(\kappa_t^*(\widehat{c}_{I,t})) > 0$ (which is true given Assumption 2 and the previous result that $\underline{\kappa} < \kappa_t^*(\widehat{c}_{I,t}) < \overline{\kappa}$), $\frac{\partial \kappa_t^*(\widehat{c}_{I,t})}{\partial \widehat{c}_{I,t}} > 0$ for all $\widehat{c}_{I,t}$ (true given the previous result on the monotonicity of E's entry threshold rule in perceived incumbent marginal cost), and $\mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}] > 0$ (assumption $\mathrm{I1}^{t+1}$ and Lemma 5(b)).

Condition (MT-iv): Using equations (G.1) and (G.2) we have

$$\frac{\Pi_3^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})}{\Pi_2^{I,t}(c_{I,t}, \widehat{c}_{I,t}, p_{I,t})} = \frac{\left[q^M(p_{I,t}) + \frac{\partial q^M(p_{I,t})}{\partial p_{I,t}}(p_{I,t} - c_{I,t})\right]}{\left(-\beta g(\kappa_t^*(\widehat{c}_{I,t}))\frac{\partial \kappa_t^*(\widehat{c}_{I,t})}{\partial \widehat{c}_{I,t}} \left\{\mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]\right\}\right)}$$

Differentiation with respect to $c_{I,t}$ gives

$$\frac{\partial \frac{\Pi_3^{I,t}(c_{I,t},\widehat{c}_{I,t},p_{I,t})}{\Pi_2^{I,t}(c_{I,t},\widehat{c}_{I,t},p_{I,t})}}{\partial c_{I,t}} = \frac{\frac{\partial q^M(p_{I,t})}{\partial p_{I,t}}}{\left(\beta g(\kappa_t^*(\widehat{c}_{I,t}))\frac{\partial \kappa_t^*(\widehat{c}_{I,t})}{\partial \widehat{c}_{I,t}}\left\{\mathbb{E}_t[V_{t+1}^I] - \mathbb{E}_t[\phi_{t+1}^I]\right\}\right)} + ...$$

$$\frac{\left[q(p_{I,t}) + \frac{\partial q^M(p_{I,t})}{\partial p_{I,t}}(p_{I,t} - c_{I,t})\right]\frac{\partial\left\{\mathbb{E}_t[V_{t+1}] - \mathbb{E}_t[\phi_{t+1}^I]\right\}}{\partial c_{I,t}}\left(\beta g(\kappa_t^*(\widehat{c}_{I,t}))\frac{\partial \kappa_t^*(\widehat{c}_{I,t})}{\partial \widehat{c}_{I,t}}\right)}{\left(\beta g(\kappa_t^*(\widehat{c}_{I,t}))\frac{\partial \kappa_t^*(\widehat{c_{I,t}})}{\partial \widehat{c}_{I,t}}\left\{\mathbb{E}_t[V_{t+1}^I] - \mathbb{E}_t[\phi_{t+1}^I]\right\}\right)^2}$$

where $\mathbb{E}_t[V_{t+1}^I|c_{I,t}]$ and $\mathbb{E}_t[\phi_{t+1}^I|c_{I,t}]$ have been written as $\mathbb{E}_t[V_{t+1}^I]$ and $\mathbb{E}_t[\phi_{t+1}^I]$ to save space.

Sufficient conditions for $\frac{\partial \frac{\Pi_3^{I,t}(c_{I,t},\widehat{c}_{I,t},p_{I,t})}{\Pi_2^{I,t}(c_{I,t},\widehat{c}_{I,t},p_{I,t})}}{\partial c_{I,t}}$ to be < 0 (implying $\frac{\Pi_3^{I,t}(c_{I,t},\widehat{c}_{I,t},p_{I,t})}{\Pi_2^{I,t}(c_{I,t},\widehat{c}_{I,t},p_{I,t})}$ is monotonic in $c_{I,t}$) are: $\left\{\mathbb{E}_t[V_{t+1}^I|c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I|c_{I,t}]\right\} > 0$ (follows from assumption I1^{t+1} and Lemma 5(b));

$\frac{\partial\left\{\mathbb{E}_t[V_{t+1}^I|c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I|c_{I,t}]\right\}}{\partial c_{I,t}} < 0$ (assumption I3^{t+1} and Corollary 4);

$\left[q(p_{I,t}) + \frac{\partial q^M(p_{I,t})}{\partial p_{I,t}}(p_{I,t} - c_{I,t})\right] \geqslant 0$ for all prices below the monopoly price (implied by strict quasi-concavity of the profit function); $g(\kappa_t^*(\widehat{c}_{I,t})) > 0$ (Assumption 2 and the previous result that $\underline{\kappa} < \kappa_t^*(\widehat{c}_{I,t}) < \overline{\kappa}$); $\frac{\partial \kappa_t^*(\widehat{c}_{I,t})}{\partial \widehat{c}_{I,t}} > 0$ (proved above); and, $\frac{\partial q^M(p_{I,t})}{\partial p_{I,t}} < 0$ (Assumption 3).

Uniqueness of the Separating Strategy under the D1 Refinement The Mailath and von Thadden theorem allows us to show that there is only one fully separating strategy, but it does not show that there can be no pooling equilibria. To show this, we use the D1 Refinement and Theorem 3 of Ramey (1996).

Theorem 7. [Based on Ramey (1996)] *Take I's signaling payoff* $\Pi^{I,t}(c_{I,t}, \kappa_t', p_{I,t})$ *where* κ_t' *is E's entry threshold. If conditions (R-i)* $\Pi_2^{I,t}(c_{I,t}, \kappa_t', p_{I,t}) \neq 0$ *for all* $(c_{I,t}, \kappa_t', p_{I,t})$*; (R-ii)* $\frac{\Pi_3^{I,t}(c_{I,t},\kappa_t',p_{I,t})}{\Pi_2^{I,t}(c_{I,t},\kappa_t',p_{I,t})}$ *is a monotone function of* $c_{I,t}$ *for all* κ_t'*; and (R-*

iii) $\overline{p} \geqslant p^{\text{static monopoly}}(\overline{c_I})$ *and* $\Pi^{I,t}(\underline{c_{I,t}}, \overline{\kappa}, \underline{p}) < \max_p \Pi^{I,t}(\underline{c_{I,t}}, \underline{\kappa}, p)$ *for all* t*, then an equilibrium satisfying the D1 refinement will be fully separating.*

The signaling payoff function in this theorem is defined based on E's threshold, not its point belief, to allow for the fact that, with pooling, E's beliefs may not be a point. (R-iii) is a condition on the support of prices, as it says that I would always prefer to use some price above $\underline{p}$ even if doing this led to certain entry when setting $\underline{p}$ would prevent entry from happening. Once again, it is implied by Assumption 3. Essentially replicating the proofs of (MT-iii) and (MT-iv) above, we now show that conditions (R-i) and (R-ii) hold.

Condition (R-i): $\Pi_2^{I,t}(c_{I,t}, \kappa_t, p_{I,t}) = -\beta g(\kappa_t)\left\{\mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]\right\}$. This will not be equal to zero if $g(\cdot) > 0$ (true given Assumption 2 and the condition that an equilibrium level of κ_t' will satisfy $\underline{\kappa} < \kappa_t' < \overline{\kappa}$), and $\left\{\mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]\right\} > 0$ (follows from assumption $\mathrm{I1}^{t+1}$ and Lemma 5(b)).

Condition (R-ii): as before, we have

$$\frac{\Pi_3^{I,t}(c_{I,t}, \kappa_t, p_{I,t})}{\Pi_2^{I,t}(c_{I,t}, \kappa_t, p_{I,t})} = \frac{\left[q^M(p_{I,t}) + \frac{\partial q^M(p_{I,t})}{\partial p_{I,t}}(p_{I,t} - c_{I,t})\right]}{\left(-\beta g(\kappa_t)\left\{\mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]\right\}\right)}$$

Differentiation with respect to $c_{I,t}$ yields

$$\frac{\partial \frac{\Pi_3^{I,t}(c_{I,t},\kappa_t,p_{I,t})}{\Pi_2^{I,t}(c_{I,t},\kappa_t,p_{I,t})}}{\partial c_{I,t}} = \frac{\frac{\partial q^M(p_{I,t})}{\partial p_{I,t}}}{\beta g(\kappa_t)\mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]} + \ldots$$

$$\frac{\left[q(p_{I,t}) + \frac{\partial q^M(p_{I,t})}{\partial p_{I,t}}(p_{I,t} - c_{I,t})\right] \frac{\partial\left\{\mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]\right\}}{\partial c_{I,t}} \left(\beta g(\kappa_t)\right)}{\left(\beta g(\kappa_t)\left\{\mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]\right\}\right)^2}$$

Sufficient conditions for $\frac{\partial \frac{\Pi_3^{I,t}(c_{I,t},\kappa_t,p_{I,t})}{\Pi_2^{I,t}(c_{I,t},\kappa_t,p_{I,t})}}{\partial c_{I,t}}$ to be < 0 (implying $\frac{\Pi_3^{I,t}(c_{I,t},\kappa_t,p_{I,t})}{\Pi_2^{I,t}(c_{I,t},\kappa_t,p_{I,t})}$ monotonic in $c_{I,t}$) are: $\left\{\mathbb{E}_t[V_{t+1}^I|c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I|c_{I,t}]\right\} > 0$ (follows from assumption I1^{t+1} and Lemma 5(b));

$\frac{\partial\left\{\mathbb{E}_t[V_{t+1}^I|c_{I,t}]-\mathbb{E}_t[\phi_{t+1}^I|c_{I,t}]\right\}}{\partial c_{I,t}} < 0$ (assumption I3^{t+1} and Corollary 4);

$\left[q(p_{I,t}) + \frac{\partial q^M(p_{I,t})}{\partial p_{I,t}}(p_{I,t} - c_{I,t})\right] \geqslant 0$ for all prices below the monopoly price (implied by quasi-concavity of the profit function); $g(\kappa_t^*(\widehat{c}_{I,t})) > 0$ (Assumption 2 and the previous result that $\underline{\kappa} < \kappa_t^*(\widehat{c}_{I,t}) < \overline{\kappa}$); and, $\frac{\partial q^M(p_{I,t})}{\partial p_{I,t}} < 0$ (Assumption 3).

Properties of the Potential Entrant's Value Functions for Period t

We now show that, given these strategies (in particular the fact that I's pricing strategy is fully revealing), which depend on the assumed properties of value functions in period $t+1$, that the value functions at the start of period t will have these same properties. For the potential entrant we have to prove:

E1^t. $\phi_t^E(c_{I,t}) > V_t^E(c_{I,t})$; and

E2^t. $\phi_t^E(c_{I,t})$ and $V_t^E(c_{I,t})$ are uniquely defined functions of $c_{I,t}$, and do not depend on κ_{t-1} or any earlier values of κ;

E3^t. $\phi_t^E(c_{I,t})$ and $V_t^E(c_{I,t})$ are continuous and differentiable in both arguments; and

E4^t. $\frac{\partial \phi_t^E(c_{I,t})}{\partial c_{I,t}} > \frac{\partial V_t^E(c_{I,t})}{\partial c_{I,t}}$

From the above, we have that

$$\phi_t^E(c_{I,t}) = \pi_E^D(c_{I,t}) + \beta \int_{\underline{c_I}}^{\overline{c_I}} \phi_{t+1}^E(c_{I,t+1})\psi_I(c_{I,t+1}|c_{I,t})dc_{I,t+1} \tag{G.3}$$

$$V_t^E(c_{I,t}) = \int_0^{\kappa^*(c_{I,t})}\int_{\underline{c_I}}^{\overline{c_I}} \left\{\beta\phi_{t+1}^E(c_{I,t+1})\psi_I(c_{I,t+1}|c_{I,t}) - \kappa\right\} g(\kappa)dc_{I,t+1}d\kappa + ... \tag{G.4}$$

$$\int_{\kappa^*(c_{I,t})}^{\overline{\kappa}}\int_{\underline{c_I}}^{\overline{c_I}} \beta V_{t+1}^E(c_{I,t+1})\psi_I(c_{I,t+1}|c_{I,t})g(\kappa)dc_{I,t+1}d\kappa$$

where we are exploiting the fact that the entrant has correct beliefs about I's marginal cost when taking its entry decision in equilibrium.

Continuity and differentiability of (G.3) and (G.4) follows from ϕ_{t+1}^E and V_{t+1}^E being continuous and differentiable (E3^{t+1}), $\psi_I(c_{I,t+1}|c_{I,t})$ being continuous and differentiable (Assumption 1) and $\kappa^*(c_{I,t})$ being continuous and differentiable as shown above. The fact that both (G.3) and (G.4) are uniquely defined and do not depend on κ_{t-1} or any earlier values of κ follows from inspection of these equations and, in particular, the fact that I's signaling strategy perfectly reveals its current cost so that E's entry threshold in period t does not depend on earlier information. As $\phi_{t+1}^E(c_{I,t+1}) > V_{t+1}^E(c_{I,t+1})$, (G.4) implies

$$V_t^E(c_{I,t}) < \beta \int_{\underline{c_I}}^{\overline{c_I}} \phi_{t+1}^E(c_{I,t+1})\psi_I(c_{I,t+1}|c_{I,t})dc_{I,t+1},$$

and therefore,

$$\phi_t^E(c_{I,t}) - V_t^E(c_{I,t}) > \pi_E^D(c_{I,t}) > 0$$

by our assumption on duopoly profits, so that $\phi_t^E(c_{I,t}) > V_t^E(c_{I,t})$.

To show that $\frac{\partial[\phi_t^E(c_{I,t})]}{\partial c_{I,t}} > \frac{\partial[V_t^E(c_{I,t})]}{\partial c_{I,t}}$, it is convenient to write

$$\phi_t^E(c_{I,t}) - V_t^E(c_{I,t}) = \pi_E^D(c_{I,t}) + \int_0^{\overline{\kappa}} \min\{\kappa, \mathbb{E}_t[\phi_{t+1}^E|c_{I,t}] - \mathbb{E}_t[V_{t+1}^E|c_{I,t}]\}g(\kappa)d\kappa$$

so that

$$\frac{\partial[\phi_t^E(c_{I,t})]}{\partial c_{I,t}} - \frac{\partial[V_t^E(c_{I,t})]}{\partial c_{I,t}} = \frac{\partial\pi_E^D(c_{I,t})}{\partial c_{I,t}} + ...$$

$$\beta \frac{\partial \int_0^{\overline{\kappa}} \min\{\kappa, \mathbb{E}_t[\phi_{t+1}^E|c_{I,t}] - \mathbb{E}_t[V_{t+1}^E|c_{I,t}]\}g(\kappa)d\kappa}{\partial c_{I,t}} > 0$$

where the inequality follows from $\frac{\partial\pi_E^D(c_{I,t})}{\partial c_{I,t}} > 0$ (Assumption 4), $0 < \kappa^* < \overline{\kappa}$ and $\frac{\partial\mathbb{E}_t[\phi_{t+1}^E|c_{I,t}]}{\partial c_{I,t}} - \frac{\partial\mathbb{E}_t[V_{t+1}^E|c_{I,t}]}{\partial c_{I,t}} > 0$ (E4^{t+1} and Corollary 3).

Properties of the Incumbent's Value Functions for Period t

For the incumbent we have to prove:

I1^t. $V_t^I(c_{I,t}) > \phi_t^I(c_{I,t})$;

I2^t. $V_t^I(c_{I,t})$ and $\phi_t^I(c_{I,t})$ are continuous and differentiable; and,

I3^t. $\frac{\partial V_t^I(c_{I,t})}{\partial c_{I,t}} < \frac{\partial\phi_t^I(c_{I,t})}{\partial c_{I,t}}$.

Condition I1^t:

$$V_t^I(c_{I,t}) = \max_{p_{I,t}} q^M(p_{I,t})(p_{I,t} - c_{I,t}) + ... \tag{G.5}$$

$$\beta \left[\begin{array}{c} (1 - G(\kappa_t^*(\varsigma_{I,t}^{-1}(p_{I,t}))))\mathbb{E}_t[V_{t+1}^I|c_{I,t}] \\ +G(\kappa_t^*(\varsigma_{I,t}^{-1}(p_{I,t})))\mathbb{E}_t[\phi_{t+1}^I|c_{I,t}] \end{array} \right]$$

$$\phi_t^I(c_{I,t}) = \pi_I^D(c_{I,t}) + \beta\mathbb{E}_t[\phi_{t+1}^I|c_{I,t}] \tag{G.6}$$

Now, given I1^{t+1} and Lemma 5(b),

$$\beta \left[\begin{array}{c} (1 - G(\kappa_t^*(\varsigma_{I,t}^{-1}(p_{I,t}))))\mathbb{E}_t[V_{t+1}^I|c_{I,t}]+ \\ G(\kappa_t^*(\varsigma_{I,t}^{-1}(p_{I,t})))\mathbb{E}_t[\phi_{t+1}^I|c_{I,t}] \end{array} \right] > \beta\mathbb{E}_t[\phi_{t+1}^I|c_{I,t}]$$

for any $p_{I,t}$ (including the static monopoly price). But, as $q^M(p_{I,t})(p_{I,t} - c_{I,t}) > \pi_I^D(c_{I,t})$ (Assumption 4) when the static monopoly price is chosen, it follows that $V_t^I(c_{I,t}) > \phi_t^I(c_{I,t})$ when a possibly different price is chosen by the incumbent.

Condition I2^t: continuity and differentiability of $V_t^I(c_{I,t})$ and $\phi_t^I(c_{I,t})$ follows from expressions (G.5) and (G.6), and the continuity and differentiability of the static and duopoly profit functions, the incumbent's equilibrium pricing function, the entry threshold function, $\kappa_t^*(c_{I,t})$, the cdf of entry costs G, the cost transition conditional probability function ψ_I, and the following period value functions $V_{t+1}^I(c_{I,t+1})$ and $\phi_{t+1}^I(c_{I,t+1})$ (I2^{t+1}).

Condition I3^t:

$$\frac{\partial V_t^I(c_{I,t})}{\partial c_{I,t}} = \frac{\partial \pi^M(p^*, c_{I,t})}{\partial c_{I,t}} + \beta \frac{\partial \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]}{\partial c_{I,t}} - ...$$

$$\beta \frac{\partial \kappa^*(c_{I,t})}{\partial c_{I,t}} g(\kappa_t^*(c_{I,t})) \left\{ \mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}] \right\} + ...$$

$$\beta(1 - G(\kappa^*(c_{I,t}))) \left[\frac{\partial \mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]}{\partial c_{I,t}} \right]$$

$\frac{\partial \pi^M(p^*, c_{I,t})}{\partial c_{I,t}} = -q^M(p^*) + \frac{\partial p^*(c_{I,t})}{\partial c_{I,t}} \left\{ q^M(p^*) + \frac{\partial q^M(p^*)}{\partial p}(p^* - c_{I,t}) \right\}$. But from the unique equilibrium strategy of the incumbent (recall that $V_t^I(c_{I,t})$ is the value to being an incumbent at the beginning of period t allowing for equilibrium play in that period),

$$\frac{\partial p^*}{\partial c_{I,t}} \left\{ q^M(p^*) + \frac{\partial q^M(p^*)}{\partial p}(p^* - c_{I,t}) \right\} = \beta g(\kappa_t^*(c_{I,t})) \frac{\partial \kappa_t^*}{\partial c_{I,t}} \left\{ \mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}] \right\}$$

so

$$\frac{\partial V_t^I(c_{I,t})}{\partial c_{I,t}} = -q^M(p^*) + \beta \frac{\partial \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]}{\partial c_{I,t}} + ...$$

$$\beta(1 - G(\kappa^*(c_{I,t}))) \left[\frac{\partial \mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]}{\partial c_{I,t}} \right]$$

and

$$\frac{\partial \phi_t^I(c_{I,t})}{\partial c_{I,t}} = \frac{\partial \pi^D(c_{I,t})}{\partial c_{I,t}} + \beta \frac{\partial \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]}{\partial c_{I,t}}$$

$$= -q_I^D(c_{I,t}) + \frac{\partial \pi_I^D}{\partial a_E^D} \frac{\partial a_E^D}{\partial c_{I,t}} + \beta \frac{\partial \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]}{\partial c_{I,t}} < 0$$

where the inequality follows from the assumption that $\frac{\partial \pi^D(c_{I,t})}{\partial c_{I,t}} < 0$ (Assumption 4). Therefore,

$$\frac{\partial V_t^I(c_{I,t})}{\partial c_{I,t}} - \frac{\partial \phi_t^I(c_{I,t})}{\partial c_{I,t}} = q_I^D(c_{I,t}) - q^M(p^*(c_{I,t})) - \frac{\partial \pi_I^D}{\partial a_E^D} \frac{\partial a_E^D}{\partial c_{I,t}} + ...$$

$$\beta(1 - G(\kappa^*(c_{I,t}))) \left[\frac{\partial \{\mathbb{E}_t[V_{t+1}^I | c_{I,t}] - \mathbb{E}_t[\phi_{t+1}^I | c_{I,t}]\}}{\partial c_{I,t}} \right] < 0$$

where the inequality follows from Assumption 4, as $q^M(p^*(c_{I,t})) > q^M(p^{\text{static monopoly}})$ because the limit price will be below the static monopoly price and demand slopes downwards (Assumption 3), and I3^{t+1} and Corollary 4.

G.0.5 Proof for Period T

We now turn to showing that the value functions defined at the start of period T have the required properties. Of course, this is trivial because the game ends after period T so that if I is a monopolist in period T then it should just set the static monopoly price, and E should not enter for any positive entry cost. Therefore, $\phi_T^E(c_{I,T}) = \pi_E^D(c_{I,T})$, $V_T^E(c_{I,T}) = 0$, $\phi_T^I(c_{I,T}) = \pi_I^D(c_{I,T})$ and $V_T^I(c_{I,T}) = q(p^{\text{static monopoly}}(c_{I,T}))(p^{\text{static monopoly}}(c_{I,T}) - c_{I,T})$. Under our assumptions $\phi_T^E(c_{I,T}) > V_T^E(c_{I,T})$, $V_T^I(c_{I,T}) > \phi_T^I(c_{I,T})$, $\frac{\partial \phi_T^E}{\partial c_{I,T}} > \frac{\partial V_T^E}{\partial c_{I,T}} = 0$, $\frac{\partial V_T^I(c_{I,T})}{\partial c_{I,T}} < \frac{\partial \phi_T^I(c_{I,T})}{\partial c_{I,T}} < 0$.[1]

[1] $\frac{\partial V_T^I(c_{I,T})}{\partial c_{I,T}} - \frac{\partial \phi_T^I(c_{I,T})}{\partial c_{I,T}} = q_I^D(c_{I,T}) - q^M(p^{\text{static monopoly}}(c_{I,T})) - \frac{\partial \pi_I^D}{\partial a_E^D} \frac{\partial a_E^{*D}}{\partial c_{I,T}} < 0$ by Assumption 4.

Appendix H

Solution Method for the Dynamic Limit Pricing Model

In this Appendix we provide a step-by-step guide to how we solve the dynamic model with constant marginal costs that evolve exogenously.

Before solving the dynamic model we specify a 100-point grid for c_I and c_E, and for each of the cost combinations, we solve for each firm's single-period profits in the specified complete information duopoly stage game that follows entry and the incumbent's profits when it prices as a static monopolist. We specify a 1,000-point price grid on which we compute $\frac{\partial \pi_I^M(p_I,c_I)}{\partial p_I} = q^M(p_I) + \frac{\partial q^M(p_I)}{\partial p_I}(p_I - c_I)$ at each of the 1,000 prices for each level of c_I. We also verify that the sufficient condition for single-crossing

$$\left(q_I^D(c_I, c_E) - q^M(p_I^{\text{static monopoly}}(c_I)) - \frac{\partial \pi_I^D(c_I, c_E, p_E)}{\partial p_E} \frac{\partial p_E^*}{\partial c_I} < 0 \right)$$

for all c_I at all points in the cost grid. With these values computed, we solve the dynamic game.

Step 1. Consider period T. Calculate the incumbent's static monopoly profits for each discretized value of c_I (in this period it does not face the threat of entry and so it prices as a static monopolist). The incumbent's static monopoly and duopoly profits define V_T^I and ϕ_T^I for each value on the grid. For the potential entrant $V_T^E = 0$ and ϕ_T^E is the static duopoly profit.

Step 2. Consider period $T-1$.

(a) For a given value of $c_{E,T-1}$, use the assumed form of the transition processes for marginal costs to calculate the value of $\mathbb{E}_{T-1}[\phi_T^E|c_{I,T-1}, c_{E,T-1}]$ for each value of $c_{I,T-1}$.

As $\mathbb{E}_{T-1}[V_T^E|c_{I,T-1}, c_{E,T-1}] = 0$ and $\kappa_{T-1}^*(c_{I,T-1}, c_{E,T-1}) = \beta\mathbb{E}_{T-1}[\phi_T^E|c_{I,T-1}, c_{E,T-1}]$, we compute $g(\kappa_{T-1}^*)$ and $\frac{\partial \kappa_{T-1}^*(c_{I,T-1}, c_{E,T-1})}{\partial c_{I,T-1}}$ for each of these values.

(b) We solve for the pricing strategy of the incumbent as a function of its marginal cost, by solving the differential equation starting from the boundary solution that the firm with the highest marginal cost sets the static monopoly price

$$\frac{\partial p_{I,T-1}^*}{\partial c_{I,T-1}} = \frac{\beta g(\kappa_{T-1}^*)\frac{\partial \kappa_{T-1}^*(c_{I,T-1}, c_{E,T-1})}{\partial c_{I,t}}\{\mathbb{E}_{T-1}[V_T^I|c_{I,T-1}, c_{E,T-1}] - \mathbb{E}_{T-1}[\phi_T^I|c_{I,T-1}, c_{E,T-1}]\}}{\frac{\partial \pi_I^M(p_{I,T-1}, c_{I,T-1})}{\partial p_{I,T-1}}}$$

This is done using `ode113` in MATLAB. As we solve the differential equation we interpolate, using cubic splines, the values of $g(\kappa_{T-1}^*)$, $\frac{\partial \kappa_{T-1}^*(c_{I,T-1}, c_{E,T-1})}{\partial c_{I,t}}$, $\{\mathbb{E}_{T-1}[V_T^I|c_{I,T-1}, c_{E,T-1}] - \mathbb{E}_{T-1}[\phi_T^I|c_{I,T-1}, c_{E,T-1}]\}$ and $\frac{\partial \pi_I^M(p_{I,T-1}, c_{I,T-1})}{\partial p_{I,T-1}}$ from the relevant grid points.

(c) Given the entry and pricing strategies we can calculate $V_{T-1}^j(c_{I,T-1}, c_{E,T-1})$ and $\phi_{T-1}^j(c_{I,T-1}, c_{E,T-1})$ for both firms (i.e., the values of each firm as a monopolist/potential entrant/duopolist) as appropriate given the cost state.

Step 3. Consider period $T-2$. Here we proceed using the same steps as in Step 2, except that $\kappa_{T-2}^*(c_{I,T-2}, c_{E,T-2}) = \beta\{\mathbb{E}_{T-2}[\phi_{T-1}^E|c_{I,T-2}, c_{E,T-2}] - \mathbb{E}_{T-2}[V_{T-1}^E|c_{I,T-2}, c_{E,T-2}]\}$.

Step 4. Repeat Step 3 for all previous periods.

Bibliography

Bain, J. (1949), "A note on pricing in monopoly and oligopoly," *The American Economic Review*, 39, 448–464.

Baron, D. P. (1973), "Limit Pricing, Potential Entry, and Barriers to Entry," *American Economic Review*, 63, 666–674.

Becker, G. S. and Murphy, K. M. (1988), "A theory of rational addiction," *The journal of political economy*, pp. 675–700.

Benkard, L., Bodoh-Creed, A., and Lazarev, J. (2010), "Simulating the Dynamic Effects of Horizontal Mergers: U.S. Airlines," Tech. rep., Stanford University.

Bennett, R. D. and Craun, J. M. (1993), *The Airline Deregulation Evolution Continues: The Southwest Effect*, Office of Aviation Analysis, U.S. Department of Transportation, Washington D.C.

Bergman, M. (2002), "Potential Competition: Theory, Empirical Evidence and Legal Practice," Tech. rep., Swedish Competition Authority.

Berry, S. and Jia, P. (2010), "Tracing the Woes: An Empirical Analysis of the Airline Industry," *American Economic Journal: Microeconomics*, 2, 1–43.

Berry, S. T. (1992), "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60, 889–917.

Berry, S. T. (1994), "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242–262.

Besanko, D., Doraszelski, U., and Kryukov, Y. (2014), "The Economics of Predation: What Drives Pricing When There is Learning-by-Doing?" *American Economic Review*, 104, 868–897.

Boguslaski, C., Ito, H., and Lee, D. (2004), "Entry Patterns in the Southwest Airlines Route System," *Review of Industrial Organization*, 25, 317–350.

Borenstein, S. (1989), "Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry," *RAND Journal of Economics*, 20, 344–365.

Borenstein, S. (1990), "Airline Mergers, Airport Dominance, and Market Power," *American Economic Review*, 80, 400–404.

Borenstein, S. (2009), "To what electricity price do consumers respond? Residential demand elasticity under increasing-block pricing," *Preliminary Draft April*, 30.

Borenstein, S. and Rose, N. (1995), "Competition and Price Dispersion in the U.S. Airline Industry," *Journal of Political Economy*, 102, 653–683.

Borkovsky, R. N., Ellickson, P. B., Gordon, B. R., Aguirregabiria, V., Gardete, P., Grieco, P., Gureckis, T., Ho, T.-H., Mathevet, L., and Sweeting, A. (2014), "Multiplicity of Equilibria and Information Structures in Empirical Games: Challenges and Prospects," *Marketing Letters*, 26, 115–125.

Bulow, J. I., Geanakoplos, J. D., and Klemperer, P. D. (1985), "Multimarket oligopoly: Strategic substitutes and complements," *Journal of Political economy*, 93, 488–511.

Chetty, R. and Saez, E. (2013), "Teaching the tax code: Earnings responses to an experiment with EITC recipients," *American Economic Journal: Applied Economics*, 5, 1–31.

Chetty, R., Looney, A., and Kroft, K. (2009), "Salience and Taxation: Theory and Evidence," *American Economic Review*, 99, 1145–1177.

Chevalier, J. (1995), "Capital Structure and Product Market Competition: Empirical Evidence from the Supermarket Industry," *American Economic Review*, 85, 415–435.

Ching, A., Erdem, T., and Keane, M. (2009), "The price consideration model of brand choice," *Journal of Applied Econometrics*, 24, 393–420.

Cho, I. and Sobel, J. (1990), "Strategic Stability and Uniqueness in Signaling Games," *Journal of Economic Theory*, 50, 381–413.

Crawford, G. S., Tosini, N., and Waehrer, K. (2011), "The Impact of'Rollover'Contracts on Switching Costs in the UK Voice Market: Evidence from Disaggregate Customer Billing Data," .

DellaVigna, S. and Malmendier, U. (2006), "Paying not to go to the gym," *The American Economic Review*, pp. 694–719.

Doraszelski, U. and Pakes, A. (2007), "A Framework for Applied Dynamic Analysis in I.O." in *Handbook of Industrial Organization*, eds. M. Armstrong and R. H. Porter, vol. 3, pp. 1887–1966, Elsevier.

Dubé, J.-P., Hitsch, G. J., and Rossi, P. E. (2009), "Do switching costs make markets less competitive?" *Journal of Marketing research*, 46, 435–445.

Dubé, J.-P., Hitsch, G. J., and Rossi, P. E. (2010), "State dependence and alternative explanations for consumer inertia," *The RAND Journal of Economics*, 41, 417–445.

Dubin, J. A. and McFadden, D. L. (1984), "An econometric analysis of residential electric appliance holdings and consumption," *Econometrica: Journal of the Econometric Society*, pp. 345–362.

Edlin, A. and Farrell, J. (2004), "The American Airlines Case: A Chance to Clarify Predation Policy," in *Antitrust Revolution: Economics, Competition and Policy*, eds. J. E. Kwoka and L. J. White, Oxford University Press.

Einav, L. and Finkelstein, A. (2011), "Selection in Insurance Markets: Theory and Empirics in Pictures," *The Journal of Economic Perspectives*, 25, 115–138.

Einav, L., Finkelstein, A., and Levin, J. (2010), "Beyond Testing: Empirical Models of Insurance Markets," *Annu. Rev. Econ*, 2, 311–36.

Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., and Cullen, M. R. (2013), "Selection on Moral Hazard in Health Insurance," *The American Economic Review*, 103, 178–219.

Einav, L., Finkelstein, A., and Schrimpf, P. (2015), "The Response of Drug Expenditure to Non-Linear Contract Design: Evidence from Medicare Part D." *The quarterly journal of economics*, 130, 841.

Ellison, G. and Ellison, S. (2011), "Strategic entry deterrence and the behavior of pharmaceutical incumbents prior to patent expiration," *American Economic Journal: Microeconomics*, 3, 1–36.

Elzinga, K. and Mills, D. (2005), "Predatory Pricing in the Airline Industry: Spirit Airlines v. Northwest Airlines," in *The Antitrust Revolution*, eds. J. E. Kwoka and L. J. White, Oxford University Press.

Ericson, R. and Pakes, A. (1995), "Markov-perfect industry dynamics: A framework for empirical work," *The Review of Economic Studies*, 62, 53–82.

Farrell, J. and Klemperer, P. (2007), "Coordination and lock-in: Competition with switching costs and network effects," *Handbook of industrial organization*, 3, 1967–2072.

Fershtman, C. and Pakes, A. (2012), "Dynamic Games with Asymmetric Information: A Framework for Empirical Work," *Quarterly Journal of Economics*, 127, 1611–1661.

Friedman, J. (1979), “On Entry Preventing Behavior,” in *Applied Game Theory*, eds. S. J. Brams, A. Schotter, and G. Schwodiauer, Physica-Verlag.

Gaskins, D. (1971), “Dynamic Limit Pricing, Barriers to Entry and Rational Firms,” *Journal of Economic Theory*, 3, 306–322.

Gedge, C., Roberts, J. W., and Sweeting, A. (2014), “A Model of Dynamic Limit Pricing with an Application to the Airline Industry,” *NBER Working Paper, No. 20293*.

Gerardi, K. S. and Shapiro, A. H. (2009), “Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry,” *Journal of Political Economy*, 117, 1–37.

Goettler, R. L. and Clay, K. (2011), “Tariff Choice with Consumer Learning and Switching Costs,” *Journal of Marketing Research*, 48, 633–652.

Goetz, C. F. and Shapiro, A. H. (2012), “Strategic Alliance as a Response to the Threat of Entry: Evidence from Airline Codesharing,” *International Journal of Industrial Organization*, 30, 735–747.

Goolsbee, A. and Syverson, C. (2008), “How do incumbents respond to the threat of entry? Evidence from the major airlines,” *The Quarterly Journal of Economics*, 123, 1611–1633.

Grubb, M. and Osborne, M. (2015), “Cellular service demand: Biased beliefs, learning, and bill shock,” *American Economic Review*, 105, 234–271.

Handel, B. R. (2013), “Adverse selection and inertia in health insurance markets: When nudging hurts,” *The American Economic Review*, 103, 2643–2682.

Hansen, L. P. (1982), “Large sample properties of generalized method of moments estimators,” *Econometrica: Journal of the Econometric Society*, pp. 1029–1054.

Harrington, J. E. (1986), “Limit Pricing when the Potential Entrant is Uncertain of its Cost Function,” *Econometrica*, 54, 429–437.

Hartmann, W. R. and Viard, V. B. (2008), “Do frequency reward programs create switching costs? A dynamic structural analysis of demand in a reward program,” *Quantitative Marketing and Economics*, 6, 109–137.

Hausman, J. (1997), “Valuation of New Goods Under Perfect and Imperfect Competition,” in *The Economics of New Goods*, eds. R. Gordon and T. Bresnahan, pp. 207–248, University of Chicago Press, Chicago.

Heckman, J. J. (1981), “Heterogeneity and state dependence,” in *Studies in labor markets*, pp. 91–140, University of Chicago Press.

Hendel, I. (1999), "Estimating multiple-discrete choice models: An application to computerization returns," *The Review of Economic Studies*, 66, 423–446.

Ho, K. (2006), "The Welfare Effects of Restricted Hospital Choice in the US Medical Care Market," *Journal of Applied Econometrics*, 21, 1039–1079.

Ho, K., Hogan, J., and Morton, F. S. (2015), "The impact of consumer inattention on insurer pricing in the Medicare Part D program," Tech. rep., National Bureau of Economic Research.

Honka, E. (2014), "Quantifying search and switching costs in the US auto insurance industry," *The Rand Journal of Economics*, 45, 847–884.

Hortaçsu, A., Madanizadeh, S. A., and Puller, S. L. (2015), "Power to choose? An analysis of consumer inertia in the residential electricity market," Tech. rep., National Bureau of Economic Research.

Ito, K. (2012), "Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing," Tech. rep., National Bureau of Economic Research.

Ito, K. (2014), "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing," *American Economic Review*, 104, 537–63.

Kaldor, N. (1935), "Market Imperfection and Excess Capacity," *Economica*, 2, 33–50.

Kamien, M. I. and Schwartz, N. L. (1971), "Limit Pricing and Uncertain Entry," *Econometrica*, 39, 441–454.

Kaya, A. (2009), "Repeated signaling games," *Games and Economic Behavior*, 66, 841–854.

Kim, E. H. and Singal, V. (1993), "Mergers and Market Power: Evidence from the Airline Industry," *American Economic Review*, 83, 549–69.

Klemperer, P. (1995), "Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade," *The Review of Economic Studies*, 62, 515–539.

Kwoka, J. and Shumilkina, E. (2010), "The Price Effect of Eliminating Potential Competition: Evidence from an Airline Merger," *The Journal of Industrial Economics*, 58, 767–793.

Lazarev, J. (2013), "The Welfare Effects of Intertemporal Price Discrimination: An Empirical Analysis of Airline Pricing in US Monopoly Markets," Tech. rep., New York University.

Lieberman, M. B. (1987), "Excess Capacity as a Barrier to Entry: An Empirical Appraisal," *Journal of Industrial Economics*, 35, 607–627.

Liebman, J. B. and Zeckhauser, R. J. (2004), "Schmeduling," Tech. rep., Working Paper.

Mailath, G. (1987), "Incentive compatibility in signaling games with a continuum of types," *Econometrica*, 55, 1349–1365.

Mailath, G. J. and von Thadden, E.-L. (2013), "Incentive Compatibility and Differentiability: New Results and Classic Applications," *Journal of Economic Theory*, 148, 1841–1861.

Marshall, G. (2015), "Hassle Costs and Price Discrimination: An Empirical Welfare Analysis," *American Economic Journal: Applied Economics*, 7, 123–146.

Masson, R. T. and Shaanan, J. (1982), "Stochastic-Dynamic Limiting Pricing: An Empirical Test," *Review of Economics and Statistics*, 64, 413–422.

Masson, R. T. and Shaanan, J. (1986), "Excess Capacity and Limit Pricing: An Empirical Test," *Economica*, 53, 365–378.

Matthews, S. A. and Mirman, L. J. (1983), "Equilibrium Limit Pricing: The Effects of Private Information and Stochastic Demand," *Econometrica*, 51, 981–996.

McFadden, D. (1989), "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica: Journal of the Econometric Society*, pp. 995–1026.

Milgrom, P. and Roberts, J. (1982), "Limit pricing and entry under incomplete information: An equilibrium analysis," *Econometrica: Journal of the Econometric Society*, pp. 443–459.

Modigliani, F. (1958), "New Developments on the Oligopoly Front," *Journal of Political Economy*, 66, 215–232.

Morrison, S. (2001), "Actual, Adjacent, and Potential Competition: Estimating the Full Effect of Southwest Airlines," *Journal of Transport Economics and Policy*, 35, 239–256.

Morrison, S. A. and Winston, C. (1987), "Empirical Implications and Tests of the Contestability Hypothesis," *Journal of Law and Economics*, 30, 53–66.

OfCom (2012a), "Assessment of Future Mobile Competition and Award of 800MHz and 2.6GHz," .

OfCom (2012b), "Communications Market Report," .

OfCom (2013), "The Communications Market Report," .

OfCom (2015), “Consumer switching in mobile - Plain English summary,” Tech. rep., Office of Communicaitons, UK.

OfCom (2016), “OfCom mobile switching: quantitative research,” Tech. rep., Office of Communicaitons, UK.

Osborne, M. (2011), “Consumer learning, switching costs, and heterogeneity: A structural examination,” *Quantitative Marketing and Economics*, 9, 25–70.

Pakes, A. and Pollard, D. (1989), “Simulation and the asymptotics of optimization estimators,” *Econometrica: Journal of the Econometric Society*, pp. 1027–1057.

Petrin, A. and Train, K. (2010), “A control function approach to endogeneity in consumer choice models,” *Journal of Marketing Research*, 47, 3–13.

Ramey, G. (1996), “D1 signaling equilibria with multiple signals and a continuum of types,” *Journal of Economic Theory*, 69, 508–531.

Reiss, P. C. and White, M. W. (2005), “Household electricity demand, revisited,” *The Review of Economic Studies*, 72, 853–883.

Riley, J. G. (1979), “Informational Equilibrium,” *Econometrica*, 47, 331–359.

Roddie, C. (2012a), “Signaling and Reputation in Repeated Games, I: Finite games,” .

Roddie, C. (2012b), “Signaling and Reputation in Repeated Games, II: Stackelberg limit properties,” .

Saez, E. (2010), “Do taxpayers bunch at kink points?” *American Economic Journal: Economic Policy*, pp. 180–212.

Seamans, R. C. (2013), “Threat of Entry, Asymmetric Information, and Pricing,” *Strategic Management Journal*, 34, 426–444.

Silva, J. S. and Tenreyro, S. (2006), “The Log of Gravity,” *Review of Economics and Statistics*, 88, 641–658.

Sims, C. A. (2003), “Implications of rational inattention,” *Journal of monetary Economics*, 50, 665–690.

Small, K. A. and Rosen, H. S. (1981), “Applied Welfare Economics with Discrete Choice Models,” *Econometrica: Journal of the Econometric Society*, pp. 105–130.

Smiley, R. (1988), “Empirical Evidence on Strategic Entry Deterrence,” *International Journal of Industrial Organization*, 6, 167–180.

Snider, C. (2009), "Predatory Incentives and Predation Policy: The American Airlines Case," Tech. rep., UCLA.

Strassmann, D. L. (1990), "Potential Competition in the Deregulated Airlines," *Review of Economics and Statistics*, 72, 696–702.

Thaler, R. H. and Sunstein, C. (2008), *en Sunstein, CR (2008) Nudge: Improving decisions about health, wealth, and happiness*, New Haven: Yale University Press.

Tirole, J. (1994), *The theory of industrial organization*, MIT press.

Toxvaerd, F. (2008), "Strategic Merger Waves: A Theory of Musical Chairs," *Journal of Economic Theory*, 140, 1–26.

Toxvaerd, F. (2010), "Dynamic Limit Pricing," *Working Paper*.

Train, K. E. (2009), *Discrete choice methods with simulation*, Cambridge university press.

Williams, J. (2012), "Capacity Investments, Exclusionary Behavior, and Welfare: A Dynamic Model of Competition in the Airline Industry," Tech. rep., University of Georgia.

Wooldridge, J. M. (2010), *Econometric analysis of cross section and panel data*, MIT press.

Biography

Christopher David Gedge, more commonly known as Chris Gedge, was born in Hastings, England in 1981. He attended University of Cambridge, where he earned a B.A. in Economics. Prior to this he worked at KPMG as an auditor, specializing in financial services. Chris has an M.A. in Economics and a Ph.D. in Economics from Duke University. From summer 2016 Chris will look after his one year old son, swapping MATLAB, STATA and LaTeXfor Monkey Music, Boppin' Bunnies and the soft play centres of South East London.