The London School of Economics and Political Science

Essays in Applied Econometrics

Cristian Huse

A thesis submitted to the Department of Economics
of the London School of Economics and Political Science
for the degree of Doctor of Philosophy

London, October 2008

Contents

List of Tables

Abstract

This thesis is divided in three essays.

The first essay examines the reactions by incumbent airlines to the threat and actual entry of the low-cost carrier Gol in the Brazilian domestic air transport market. By estimating the reactions in prices, quantities and supply variables, it investigates the plausibility of theories of entry deterrence and accommodation.

The second essay proposes and implements a parsimonious three-factor model of the term structure whose dynamics is driven uniquely by observable state variables. The method allows comparing alternative views on the way state variables – macroeconomic variables, in particular - influence the yield curve dynamics, avoids curse of dimensionality problems commonly appearing in traditional models, and provides more reliable inference by using both the cross-sectional and the time series dimension of the data. I conduct in- and out-of-sample studies using a comprehensive set of US data. I show that even a parsimonious model where the level, slope and curvature factors of the term structure are driven by, respectively, measures of inflation, monetary policy and economic activity consistently outperforms the (latent-variable) benchmark model out-of-sample, when considering the five NBER-dated recessions of the last three decades.

In the third essay I empirically evaluate the incentives to tacitly collude in differentiated product markets. Tacit collusion plays an important role in merger policy: competition agencies sometimes block mergers on the grounds that they will generate 'coordinated effects', an increased likelihood of collusion. I thus propose an approach to coordinated effects merger simulation in markets where multi-product firms operate in differentiated product markets. To the best of my knowledge, this is the first full empirical implementation of a coordinated effects merger simulation model in a differentiated product market. I use the model to study the network server market and, specifically, examine the effect of the merger between HP and Compaq on their and their rivals' collective incentives and ability to sustain tacit collusion. The results suggest that the incentives to collude in the network server market are substantial, but actively decreased following the merger between HP and Compaq. In addition to exploring the incentives for collusion on one market I also examine the impact of (i) multi-market contact on firms' incentive and ability to sustain tacit coordination and (ii) a competitive fringe of smaller players who co-exist with a subset of the larger players in an industry who tacitly collude. By taking the economic theory of tacit collusion seriously in an empirical example, I show that the intuition many economists have for the effect of mergers on the incentives to tacitly collude is actually wrong.

To my family, with love and gratitude

"The PhD is a skin-thickening process."

Myself

"Jamais desviaremos nossos olhos dos elevados picos que almejamos conquistar."

Lema, Patrulha Urso

"The journey is the destination."

Dan Eldon

Acknowledgments

The first thing that comes to my mind when submitting the last piece of work as part of my education, is how badly behaved I was in kindergarten: how I would start fights, kick the nuns who took care of me and, as an obvious result of a number of other acts, how I was almost expelled from there. Thanks to their generosity, I was allowed to stay and followed an 'alternative' program according to which I would help picking the vegetables they planted over there, among other activities which were part of an unusually early youth recovery programme.

But things started much earlier than that, and since then I have been privileged to have such special parents raising me, stimulating me, giving me the best they could afford – if not more – and leading by example. Even if during most of my life I disagreed with them when it comes to actions, I am proud to have agreed about the principles. For that very reason, they are the first and most special ones deserving a word of appraisal. And, by transitivity, this goes back to grandparents, grand-grandparents, grand-grandparents...

After kindergarten came school, but for some reason I really started to get interested in what was being taught at 8th grade, thanks to an outstanding Maths teacher - he was broad, deep and charismatic.

A couple of years later, in one of my Summer holidays during high school, I was convinced to work as an intern and lived another defining moment. After suffering for one month in the heat of Rio, I didn't need to be told about the joys of studying anymore. After that, high school, was great, and to those teachers I owe much more than a couple of lines. I was genuinely impressed with those guys and, became really enthusiastic about studying more and more. They made me think, they gave me inspiration, they made me day-dream about the future.

The initial college years were – to say the least – disappointing, and I had to persuade a lot to keep on track. To start with, I wasn't even sure about which career to follow, so I started two courses at once. At some point I had to learn Calculus, and I still remember how tough it was – what exactly was the aim of all those f's, x's and y's after all ? But I started to like those things, and was lucky enough to have another set of inspiring people teaching me Statistics and Econometrics, and this partly explains why I am here today. Before heading to London I had another stop, and there this interest in Econometrics not only grew, but got also broader.

Initially, London was hell. No money, no known faces, darkness, a cold room in student accommodation (the heater was made in Norway) and I found myself getting lost in the middle of the crowd again. Thanks to a number of people, I started to regain confidence in myself; in particular, Marcia, my MSc tutor, and Margaret, thanks to whom I landed in the FMG.

"The PhD is a skin thickening process". This was something I remember repeating hundreds of times when talking to Alberto. Man, it's over, I am about to be able to say that I have also made it ! Thanks !

Thanks are also due to a number of people 'gravitating' around me during all these years, in one way or another. Especially to Sridhar, Paolo, Beatriz, Moacir, Carsten, Sarquis, Xuewen... – Gee, how I miss those people telling me not to leave late for the airport (over the years, I became a specialist in missing flights), to call me to go for lunch or to help me with some spreadsheet, database or even this very file ? Thanks are due also to the younger generation, especially Carlão and Cláudia.

Thanks are also due to David for providing me "housing", and also to a number of faculty whom I learned to admire during the time spent here. Thanks are due to Peter for the amount of time spent with me in front of a computer screen and all I've learnt from him (which means so little to him, but is such a massive amount of info for me !), to John for giving me the opportunity to work for – and with – him. Finally, I should thank Marcelo for so carefully reading early bits of what is now being submitted as my dissertation.

A special word goes to the IT staff that has always assisted me so well. Kuldip, Alex and, especially, Louise, your help in the moments of desperation is so appreciated. Same holds for the support staff at the LSE: Maria, Sooraya, Polly, Olivia, Pat, Linda, Emma...

I should also not forget Oliver, John and for David, who also helped me with financial support, and here goes my appreciation to them.

My junior coauthors have also been so important in making me a better researcher ! Alberto (again) and Alessandro, I really owe you this, and have been learning loads from you, guys.

My teaching experience – also outside the LSE – makes me feel lucky to have given me the opportunity to meet a number of persons that made me think I am not too bad as a person after all. Bernardo, Flavio, Leonardo, Luiz, Annette, Hans, Ona and Øystein and others did more for me keeping in touch than I ever did for you.

There are also many friends made during this journey that have always been around, despite the distance, ranging from the one I met in kindergarten to so many others I made over the years. Being far away has been tough, and I wish I could be more present, write more emails, show up for birthday parties and so on. You know who you are, and you know how special you are, despite the distance. The friends I made here – outside school – deserve a special word of appreciation due to their wonderful company, their kindness and generosity: Pilar, Anson, Bee, Crison, Ovi, Wendy and so many others, it has been *my honour* to share so many wonderful moments with you...

My last word goes to someone who has been with me in spirit and thought for a long time, but only recently *in locus*. Patricia, thank you so much for helping me carry such a heavy load and still be able to laugh with me along the way.

Introduction

This thesis proposes empirical strategies to answer three empirically motivated questions.

The first essay investigates the reactions by incumbents to the threat of entry and to actual entry of a new competitor. I estimate the response in prices, quantities, and a number of supply variables for a panel of routes disaggregated at the airline-route level. Besides examining the existence of preemptive responses, I investigate whether I can detect entry deterrence and accommodation by the incumbents using information on aircraft utilization, number of flights and measures of flight schedule, accounting for the asymmetry in responses arising due to product differentiation, and controlling for time-varying unobservables at both the market and the carrier level. The results show evidence of preemptive behaviour by the incumbents in the form of fare cuts. The incumbents do respond to both potential and actual entry, and the former is at least as important as the latter. Following the entry of Gol, the incumbents do not however sustain these responses, suggesting accommodation. There is also evidence of network adjustments in the form of a rescheduling of flights, in what can be interpreted as an attempt by the incumbents to avoid head-to-head competition with the entrant by redesigning the flight schedules.

The second essay proposes and implements a parsimonious three-factor model of the term structure whose dynamics is driven uniquely by observable state variables. The motivation behind the paper is that latent variables are well-suited when one is mostly concerned about fitting models to data, but they lack an economic interpretation which is of interest when, for instance, conducting policy experiments. Following this reasoning, the model can be thought of as 'Sims-structural', since despite not being based on optimizing agents, it allows conducting policy experiments.

The method I propose in Chapter 2 allows comparing alternative views on the way state variables (macroeconomic variables in the paper) influence the yield curve dynamics. Moreover, it is parsimonious, avoiding curse of dimensionality issues commonly arising in traditional models. Finally, it is in a position to provide more reliable inference by using both the cross-sectional and the time series dimension of the data. In the empirical implementation of the method, I conduct in- and out-of-sample studies using a comprehensive set of US data. There I show that even a parsimonious model where the level, slope and curvature factors of the term structure are driven by, respectively, measures of inflation, monetary policy and economic activity consistently outperforms the (latent-variable) benchmark model out-of-sample, when considering the five NBER-dated recessions of the last three decades.

The third essay empirically evaluate the incentives to tacitly collude in differentiated product markets. The idea I follow is to estimate the incentive-compatibility constraint for collusion *vis-à-vis* defection for each firm in a given market within a repeated game where firms act strategically on prices. The questions I address are whether there are incentives to collude prior to a merger and whether these incentives change as a result of a merger, given the repeated interaction between a smaller number of players. I show in the paper that the intuition many economists have for the effect of mergers on the incentives to tacitly collude is actually wrong, since they do not necessarily increase with less players.

Following Friedman (1971), I consider the feasibility of sustaining a candidate collusive equilibrium using 'grim' strategies. To examine the incentives to collude using grim strategies, one needs to consider the returns achieved by each firm in the three pricing scenarios – 'Collusion', 'Nash equilibrium pricing' and 'Defection'. Fundamental ingredients for this computation are a demand model to compute profits of the stage game, an asset pricing model to estimate the discount factor of the firms, and an algorithm to compute the value functions of the firms in each of the above scenarios. In addition to exploring the incentives for collusion in one market I also examine the impact of (i) multi-market contact on firms' incentive and ability to sustain tacit coordination and (ii) a competitive fringe of smaller players who co-exist with a subset of the larger players in an industry who tacitly collude. I apply the techniques using data from the market for network servers prior to the merger between Compaq and Hewlett-Packard. I find on the incentives for firms in the server industry to tacitly collude. The results suggest that the incentives to collude in the network server market are substantial, but actively decreased following the merger between HP and Compaq.

1

How Do Incumbents React to Entry: Evidence from Differentiated Product Markets

1.1 Introduction

This paper investigates how incumbents respond to the threat of entry of a new competitor. To do so, I estimate the reactions to route entry of a low-cost carrier (LCC) in the Brazilian domestic air transport market.

The Brazilian domestic air transport market is not only interesting due to the dimensions of the country and its emergence in economic terms, resulting in increased demand for air transport, but also for providing a case study of a low-cost carrier which entered the market following its liberalization after a number of failed attempts to take on the big players in the industry before its establishment. In fact, in 2008, seven years after entering the Brazilian market with six aircraft, Gol Airlines commands some 45% of the market, with close to $3bn in net revenues in 2007 and over $3bn in market value.[1] The industry has also been closely watched by the antitrust authorities given a number of mergers, acquisitions and associations (such as code-sharing agreements) between players.

To estimate the reactions to Gol's entry, I estimate the response in prices, quantities, and a number of supply variables for a panel of routes disaggregated at the airline-route level. Besides examining the existence of preemptive responses, I investigate whether I can find evidence of entry deterrence or accommodation using information on aircraft utilization, number of flights and measures of flight schedule, accounting for the asymmetry in responses arising due to product differentiation, and controlling for time-varying unobservables at both the market and the carrier level.

[1] See http://www.voegol.com.br/ir/ for Gol's annual reports and other institutional information.

Based on the entry pattern of Gol Airlines, I define a route as being threatened by Gol whenever it is not flying the route but operating within an area of influence of that route (namely of the endpoint airports). This relates to an extensive literature on airline competition, in particular that related to the debate of airport presence vs route presence as sources of competitive advantage and market power (see, for instance, Borenstein, 1989, and Evans and Kessides, 1993 for the opposing views). In contrast with most of the papers on the airline industry, which tend to focus on what happens *after* entry (as Berry, 1990, 1992), I follow the more recent literature and concentrate on what happens prior to entry (as Goolsbee and Syverson, 2008).

By using an empirical framework to estimate preemptive actions I can also test for entry deterrence and accommodation. In particular, I focus on the plausibility of limit pricing in the spirit of Milgrom and Roberts (1982) and excess capacity *à la* Dixit (1980).

The results show evidence of preemptive behaviour by the incumbents in the form of fare cuts. The incumbents do respond to both potential and actual entry, and the former is at least as important as the latter. Following the entry of Gol, the incumbents do not sustain these responses, suggesting accommodation. There is also evidence of a repositioning or rescheduling of flights, as the share of weekday peak time flights increase at the expense of the share of weekday off-peak time flights, in what can be interpreted as an attempt to avoid head-to-head competition with the entrant.

The paper is organized as follows. Section 2 presents the Brazilian domestic air transport market. Section 3 discusses what the literature tells us about how incumbents respond to entry and to which extent we can take the implications of the models to data. The Brazilian dataset collected by DAC used in this study is discussed in Section 4, whereas the empirical strategy is presented in Section 5. Section 6 presents the results and the final section concludes.

1.2 The Brazilian Domestic Airline Market

1.2.1 Industry Overview

The process of liberalization of the Brazilian domestic air transport market initiated in the early 1990s had a crucial role in attracting newcomers to the industry. As opposed to the early entrants, which lacked the financial resources and the infrastructure to succeed in the market, and eventually went burst, Gol Airlines, whose operations started in January 2001, was part of one of the biggest Brazilian transportation groups (Grupo Aurea).[2] Moreover, Gol's internal organization tried to adapt

[2] Grupo Aurea was at that time the biggest Mercedes-Benz bus customer in Latin America and one of the biggest worldwide. (Evangelho, 2002)

to the Brazilian context some key aspects of the internal organization of Southwest Airlines, a low-cost carrier operating in the US.

By early 2001, the process of liberalization of the market, supervised by the Department of Civil Aviation (DAC) was in a stage where, in stark contrast with the thirty or so years of strict control over key variables, airlines had substantial freedom to operate, in the sense that fares and flight frequencies were almost entirely liberalized. In particular, by that time airlines were free to enter and exit markets as quick as allowed by technical regulations, so that it would take no longer than one month for a firm to have a proposed new flight authorized by the regulator. As a result, the market grew steadily over time, reaching 60mn and 83mn travellers in 2003 and 2005, respectively.

The new regulatory framework made it easier for both newcomers and incumbents to adjust their whole network in response to changing market conditions. In such a setting, entry threats may arise and might have significant – even if short-lived – impact on fares.

This very process which attracted Gol Airlines, the player I focus in this paper, saw the incumbent airlines in a fragile financial situation and having to struggle against a competitor with deep pockets but, more importantly, with a business strategy unknown to them, since Gol was the first scheduled LCC in Latin America.

Table 1 compares some characteristics of Gol to those of the main incumbents (VARIG, TAM and VASP), for year 2002. Although its yields (price normalized by passenger and kilometers, the price measure used in the industry) were some 30% lower than those of the incumbents, the load factors (fraction of seats sold) were on average at least 10% higher which, coupled with costs estimated to be some 40% lower than those of the competitors, resulted in Gol being the only profitable airline among the major ones.

1.2.2 The Entry Pattern of Gol

In contrast with a number of well-known LCCs such as Ryanair and Southwest, which operate using the point-to-point transit model, Gol follows most of the major carriers in the use of a hub-and-spoke network of routes. Whereas the former essentially means that an aircraft will fly between airports A and B without any connecting flights and the endpoint airports, the latter results in a more complex network structure, usually compared to a chariot wheel, in which all traffic moves between the (peripheral) spokes and the (central) hub. The main benefits of the hub-and-spoke over the poin-to-point model are the fewer number of routes needed to connect all airports and the easiness with which new spokes can be added to the network. On the other hand, route scheduling is more complicated for the carrier, and the model is less flexible, needing a substantial amount of time to be redesigned. Moreover, delays can affect the whole network, as often experienced by the common airline traveller.

When coupled with the characteristics of the Brazilian market, a vast country whose population is concentrated along the Atlantic coast, especially in the Southeastern region, the pattern of entry followed by Gol differs substantially from the pattern of pointwise entry of most LCCs, being more closely related to that of Wal-Mart.[3] In practice, this means that, starting from its main hub in Sao Paulo (the economic powerhouse of the country), Gol's operation would radiate from the inside out, first towards Rio de Janeiro and Belo Horizonte, respectively second and third cities of the country, also located in the Southeastern region, and would then quickly expand through three main corridors.

The first corridor goes Southwards until Porto Alegre, the second goes towards the Northeast (including cities such as Salvador, Recife and Fortaleza), whereas the third corridor goes towards Brasilia, the country's capital (and, at a later stage, from there towards the "inner-North", including Manaus, the capital of the Amazon state). As a result, the choice of routes starting from an airport Gol entered was rather straightforward: one from the hub outwards, and another from outwards toward the hub, with a number of connections in between in both cases. This feature explains why whenever it entered an airport, Gol immediately started flying "all" routes.

Another aspect of Gol's entry pattern that resembles that of Wal-Mart is that Gol did not jump to far-off locations to later fill-in the area in-between. Before flying from Sao Paulo towards Porto Alegre, the state capital located furthest to the South, for instance, Gol would first start flying to the capitals located in between, Curitiba and Florianopolis. After the routes along the main corridor were established, it was time to consider entering medium-sized cities relatively close to the corridor.

The economic justification the literature gives for this entry pattern are economies of density.[4] An economy of density is a type of economy of scale, which arises when an airline increases the frequency of flights on a given route structure/region instead of expanding the route network. Economies of density can be potentially enjoyed through channels such as management (it is easier for upper-level management to oversee a set of locations if they are closer together), marketing and advertising, but especially maintenance in the case of Gol: given that its aircraft operate for more than 12 hours on a daily basis, maintenance needs to be done overnight in a number of different locations, so having a dense route network minimizes the number of maintenance centres.

[3] See Holmes (2008) for a thorough description and detailed analysis of the entry patterns of Wal-Mart.
[4] See, for instance, Caves, Christensen, and Tretheway (1984).

1.3 How Incumbents React to Entry

1.3.1 Preemptive and Post-Entry Reactions

Most of the empirical literature focuses at the strategic behavior of incumbents after entry occurs.[5] Exceptions to this line of research are Morrison (2001) and Goolsbee and Syverson (2008). While the former investigates the impact on prices of actual, adjacent and potential route presence, the later focuses on how far back prior to entry do incumbents react, following airport presence of the newcomer in the two endpoints (airports) of a route.

Studies of the effects of (potential and actual) entry on the strategic behaviour of incumbents go back at least to the work of Bain (1956). Although there is some evidence documenting the existence of preemptive price-cutting by incumbent airlines (Windle and Dresner, 1999; Morrison 2001), there is no agreement about the underlying reasoning leading to it. Since there is limited information in the DAC dataset I use, I describe alternative theories and try to provide suggestive pieces of evidence that favour some of the theories in the sequel.

On the one hand, Dixit (1979, 1980) puts forth the idea that incumbents may invest in excess capacity in order to make entry less attractive. In our setting, this would mean that incumbents add either bigger aircraft or increase the number of flights in a given route prior to entry of the incumbent in order to make it less attractive. On the other hand, Kreps and Wilson (1982) argue that incumbents might use a limit pricing strategy, according to which pre-entry prices are set so as to discourage potential entrants (in particular, prices generating zero profits for the entrant). A particular way to react in capacity in the airline market is by rearranging the network, so that the incumbent can offer more flights leaving at specific times.

Alternative preemptive stories that depend on the micro-level content of a dataset include switching costs, as in Klemperer (1987) and long-term contracting, as in Aghion and Bolton (1987). Although all of them are extremely interesting avenues of research, they would require micro-level information about the ticket (price, fare class etc) and about both who bought the ticket and the purpose of the trip (frequent-flyer membership, business or leisure trip, company discounts etc), which unfortunately is unavailable in most cases.

[5] See, for instance, Reiss and Spiller (1989), Borenstein (1989, 1991, 1992), Berry (1990, 1992), Evans and Kessides (1993) and Peteraf and Reed (1994).

1.3.2 Product Differentiation and Asymmetric Responses

> "Product differentiation is pervasive in markets. It is at the heart of structural empiricism and it smoothes jagged behavior that causes paradoxical outcomes in several theoretical models.", Anderson (2008)

When compared to the product offered by the major airlines, the LCC product is quite differentiated, in the sense that it is a 'no-frills' product: buying a ticket from a LCC might mean using airports far from city centres, having fewer flights to choose from, arriving and departing at times not necessarily the most convenient (as associated costs such as airport fares tend to be lower at these times), sometimes having to pay extra charges for dispatched luggage, not having the convenience of in-flight service and very often having to struggle with little legroom and seats that will not recline. Typical examples of airlines operating in the low-cost segment are Ryanair (in Europe) and Southwest (in the US).

Although not all of these factors might bite for short-haul flights, they become more and more important as the flight distance increases. In other words, product differentiation increases as a function of flight distance, as it becomes increasingly difficult to cope with, for instance, little legroom, no in-flight service, and to carry only hand-luggage as flight lenght increases. An immediate consequence of this fact is that incumbents' responses to entry might also differ according to product differentiation, so that they tend to soften in routes where the LCC product is more distant (in the space of characteristics, see Lancaster, 1966) from the product offered by major airlines.

In the Brazilian market, the main sources of product differentiation (besides flight distance) are flight frequency and flight scheduling. This happens because Gol's in-flight service and general cabin comfort have always been comparable to those offered by the incumbents. Besides not having to pay for in-flight drinks or snacks, there were no extra charges to baggage handling. What is more, most Brazilian cities have only one airport; those with more than one airport do not usually have flights leaving for the same airport. For instance, in the case of Rio de Janeiro, the Santos Dumont airport, located in the city centre, operates only scheduled flights to Sao Paulo or regional and unscheduled flights to smaller cities, which are too small to be included in the DAC dataset.

It then follows that, in a market where the business travellers are estimated to command some 70% of the market (Evangelho, 2002) flight distance (which increases the contrast between full-service and low-cost carriers), flight frequency and flight schedules are the main sources of differentiation of the LCC product. I thus investigate to which extent product differentiation softens the reactions from incumbents using these variables. More precisely, I investigate whether a product more distant from the incumbents' one in the space of characteristics will result in softer reactions by the incumbents ie. I allow price responses to be asymmetric.

1.3.3 Further Particulars and Data Availability

Price information is quite transparent in the Brazilian market, since these can be learned from the companies' websites for at least the last eight years. The number of tickets sold (ie. realized demand) and the number of seats supplied by the airlines is however sensitive information – although both are known by the authorities, only ticket sales are usually made available to researchers; instead of the number of seats supplied, authorities usually make available information on average aircraft size and the fraction of seats offered in different periods (weekdays vs weekends, peak-time vs off-peak-time). Although airline quality is not exactly observable, it is quite stable over time and can be controlled for with the use of fixed-effects, as discussed in Section 1.5. Product differentiation in other dimensions (for instance, baggage allowance or extra costs for dispatching baggage, use of different airports etc) is non-existent in the Brazilian market during the sample period.

As a result, one can study price reactions in a quite detailed way, whereas the study of quantity reactions is more limited – I thus resort to the study of proxies to quantity reactions taking the forms of (i) aircraft size; (ii) (an estimate of) number of flights; (iii) market shares.

1.4 The Data

The dataset used in this study was collected and made available by the then Brazilian regulator, the Department of Civil Aviation (DAC). Information of the panel of routes comes from the report entitled "Average Yield of Monitored Airport Pairs". Data is observed at the monthly frequency and consists of airline-specific data for up to 94 origin-destination pairs from September 2001 to March 2004.

The dataset is aggregated at the city-pair level, meaning to say that a flight going from A to B is different from one going from B to A, something usual in air transport datasets and studies. As a result, we have observations over 31 months and up to four airlines operating in each city-pair. According to the DAC report, this amounts to 98% of the passenger-weighted kilometers flown within the sample period, not including small regional and/or non-scheduled airlines operating in the Brazilian market. Following the expansion of the market in the period, the dataset begins reporting Gol operating in 51 routes serving 11 cities to reach 76 routes serving 25 cities in the final period.

The variables I observe include, for every route (origin-destination pair), the average yield (which is deflated by the Brazilian's Bureau of Statistics [IBGE] CPI) and number of tickets sold; the identity of the incumbents flying a route; the timing of the entry of Gol; a dummy variable indicating a code-share agreement between the incumbents TAM and VARIG; the share of non-stop seats during peak time of the i-th incumbent on the od-th city-pair at month t; the share of non-stop seats linking central airports of the i-th incumbent on the od-th city-pair at month t; the share peak and off-peak time flights operating on weekdays and weekends for the i-th incumbent on the od-th city-pair at month t; average aircraft size; the average operating costs. The Appendix contains detailed variable descriptions.

Actual entries of the LCC are concentrated in the first half of the dataset. This will constrain the use of lagged variables for lags above four in the empirical exercise, since the first entry of Gol in the data occurs at period four, but I show evidence that this does not impact the final results.

Using the geographic coordinates of each airport, I also compute the distance between all airports in the sample. Following these calculations, I obtain that twelve of the airports have at least one airport located less than 250km away, twenty two airports have at least one airport located less than 500km away, and twenty four airports have at least one airport located less than 1,000km away (approximately 155, 310 and 620 miles, respectively). These variables will be used when investigating the threat of entry using alternative definitions of the neighbourhood (area of influence) of a route.

As opposed to the US Department of Transport (DOT) DB1A files, the dataset I use is not a 10% random sample of all domestic tickets in a given quarter. On the low side, this means that there is no detail such as in a micro-level dataset with detailed information on, say, fares, just average fares. On the other hand, the DAC dataset has monthly observations, thus supplying more detailed information in the time series aspect than its US counterpart. Information that would be useful to take a number of theoretical models to data but which the DAC dataset lacks include fare class, whether the passenger is a frequent flyer, the place of residence, the purpose of the trip.[6]

1.5 Empirical Strategy

The empirical strategy I follow can be cast as a version of an event-study, since I investigate an event of interest looking at what happened prior to it, during its occurrence and after it had occurred. In what follows I propose four main specifications to study the reactions to entry of Gol, which I then take to data. The first – or Baseline – specification only allows for responses of the incumbents to Gol's entry *after* Gol has actually entered a route.

[6] The best one can get are estimates from the DAC that the business traveller segment is responsible for 70% of the tickets issued in the country, as reported in Evangelho, 2002.

The second specification – which I refer to as the *P* specification – allows for *preemptive* responses whenever a route is under threat. More precisely, whenever the LCC is operating within a neighbourhood (or area of influence, to be defined below) of one of the endpoints of a route, but not at the route itself, this route is considered to be under threat. To estimate this effect, I consider alternative measures of neighbourhood and take all of them to data. This generalizes the approaches of Berry (1992), Windle and Dresner (1999), Morrison (2001), Goolsbee and Syverson (2008) and others since, instead of focusing on competition at the airport (and then route) level, I am considering competition prior to airport entry.

The third specification – which I refer to as the *PU* specification – builds on the P specification but differs from it by controlling for time-varying market unobservables. By doing so, it accounts for time-varying variables such as income changes in a given market, national advertising and changes in the cost structure of a given market.

Finally, I account for product differentiation as a determinant of reactions to entry in what I refer to as the *PUD* specification. This specification accounts for the fact that *product differentiation* can soften reactions to entry in a significant way, thus allowing these reactions to be asymmetric.

Baseline Specification (B)

I define $y_{ir,t}$ as be the outcome of interest, such as the logarithm of either prices or quantities for incumbent i flying route r at time t. The baseline specification is given by

$$y_{ir,t} = \gamma_{ir} + \mu_t + \sum_{\tau=0}^{K} \beta_\tau LCCpres_{r,t^*+\tau} + X_{ir,t}\alpha + \varepsilon_{ir,t}$$

where γ_{ir} and μ_t are, respectively, carrier-route and time fixed effects, t^* *is the period when Gol starts operating a route*, $LCCpres_{r,t^*+\tau}$ are dummy variables indicating that Gol is operating route r at time $t^* + \tau, \tau \geq 0$ (I let $K = 5$ in the empirical implementation), and $X_{ir,t}$ is a vector of controls which might also be included, such as cost shifters in the case of a pricing equation. The coefficients β_τ measure the impact of LCC presence on the variable of interest.

Preemptive Specification (P)

The baseline specification does not capture the effects of the threat of entry posed on the variable of interest when Gol operates within a neighbourhood of the endpoints of a route without actually flying the route. I thus define the variable $LCCthreat_{r,t^*-\tau}, \tau \geq 1$ taking value one whenever the LCC Gol is present in a neighbourhood (area of influence) of at least one of the endpoints of route r, but not route r itself at time $t^* - \tau$ (t^* is the period where actual route entry at occurs). The preemptive specification thus consists on the baseline specification augmented with the indicators measuring the threat of entry, viz.

$$y_{ir,t} = \gamma_{ir} + \mu_t + \sum_{\tau=0}^{K} \beta_\tau LCCpres_{r,t^*+\tau} + \sum_{\tau=1}^{L} \delta_\tau LCCthreat_{r,t^*-\tau} + X_{ir,t}\alpha + \varepsilon_{ir,t}$$

where the coefficients δ_τ measure the impact of the LCC threat on the variable of interest.

Although a particular case of the *LCCthreat* variable consists on the case where Gol operates at the airport but not at a given route, *the neighbourhood definition I adopt is more general than the airport one*, since it allows defining an area of influence of a route (such as a circle centered at the endpoint airports and radius d). It is worth noting that in the Brazilian dataset, the threats I observe are outside the airport, meaning to say that once Gol enters an airport, it will immediately operate all the routes originating or ending at the airport.[7]

Preemptive Specification Controlling for Time-Varying Unobservables (PU)

The fixed-effects in specification P are unable to capture time-varying market unobservables. To illustrate the importance of controlling for this effect, assume that LCC presence in a route is correlated with time-varying unobservables that are carrier- and market-specific, such as (i) income changes in a given market; (ii) nationwide advertising to stimulate LCC demand; (ii) a change in the cost structure in a given market, following the redesign of the network and resulting LCC increased presence in this market.

As the carrier-route and time fixed-effects are unable to adequately control for these unobservables, I adopt an error-component decomposition of the time fixed-effects that uses a fact specific to airline markets which allows to group routes into endpoint cities to control for city-time effects. As a result, the time fixed-effects can be decomposed into carrier-time fixed effects and city-time fixed effects, thus allowing for modelling time-varying observables at both the market and the carrier level. To perform this, note that a route r can be written as a combination of cities of origin o and destination d and that routes can be grouped into cities of origin and destination (thus markets, which appear in a much smaller number in the data). I thus replace the time- and carrier-route- fixed-effects with carrier-time and city-time fixed-effects. While the former are able to capture time-varying unobservables related to general strategic variables of the carriers, such as nationwide advertising and network decisions, the

[7] Again, I refer the reader to the section describing the institutional details and the pattern of entry of Gol; within the hub-and-spoke network it adopts (which is in stark contrast with the point-to-point nework adopted by carriers such as Ryanair and easyJet), this means flying both from and to the hub.

latter capture market-specific changes. The PU specification then reads

$$y_{iod,t} = \gamma_{iod} + \mu_{it} + v_{o,t} + v_{d,t} + \sum_{\tau=0}^{K} \beta_\tau LCCpres_{od,t^*+\tau} + \sum_{\tau=1}^{L} \delta_\tau LCCthreat_{od,t^*-\tau} + X_{iod,t}\alpha + \varepsilon_{iod,t}$$

where $y_{iod,t}$ measures the variable of interest of incumbent i flying from origin o to destination d at time t, γ_{iod}, μ_{it}, $v_{o,t}$, $v_{d,t}$ are respectively carrier-route, carrier-time, origin-time and destination time fixed-effects.

Preemptive Specification Controlling for Time-Varying Unobservables and Product Differentiation (PUD)

The specifications proposed so far do not account for product differentiation. As a result, given two routes entered (or threatened) by the entrant in a given period and origin, the incumbent reactions will differ only by the destination-time fixed effect. If, however, there are variables that shift the intensity of the reactions, such as flight distance, one can generalize the PU model by defining the PUD model as follows:

$$\begin{aligned} y_{iod,t} &= \gamma_{iod} + \mu_{it} + v_{o,t} + v_{d,t} + \sum_{\tau=0}^{K} \beta_\tau LCCpres_{od,t^*+\tau} + \sum_{\tau=1}^{L} \delta_\tau LCCthreat_{od,t^*-\tau} + \\ &\quad + Z_{iod,t} LCCpres_{od,t}\rho + X_{iod,t}\alpha + \varepsilon_{iod,t} \end{aligned}$$

where, $Z_{iod,t}$ is a vector of shifters of the intensity of reactions and ρ is a parameter to be estimated.

In the case of price reactions to entry, one would expect at least some of the δ and β coefficients to be negative and statistically significant, indicating fare cuts given the threat of actual entry of Gol. If product differentiation issues play a role in price responses, the ρ coefficients would be expected to have a positive sign, softening these reactions for routes where the rivalry between the incumbents and the entry is less intense.

Discussion: Identification, Fixed-Effects

Throughout my analysis I assume that network decisions made by the LCC Gol (in particular, entry decisions) are exogenous. More precisely, I assume that its network decisions are made in advance (ie. are predetermined) of pricing decisions and that the fixed-effects can control for (time-varying) unobservables, so that my least squares estimates are consistently estimated. In what follows I justify this assumption economically, leaving a thorough set of robustness tests (which give additional support to the exogeneity assumption) to the Appendix.

As discussed above (Section 1.2.2), network-related decisions have to be made well in advance of reactions for a number of reasons.[8] First, in the case of a start-up, as was Gol in 2001, one needs a detailed business plan before starting operations, not only to present to potential investors, but also serve as guidance when establishing branches and maintenance centres at airports or hiring and training staff.

Second, one needs to plan (and order) well in advance a number of aircraft, choose their specifications and delivery times; typically, these decisions are made years in advance and impose extra restrictions on the company regarding network adjustment, since specific types of aircraft are designed for specific route lenghts.[9]

Third, the redesign of hub-and-spoke networks as the one of Gol also involves dealing with passengers who bought tickets in advance, in case of routes to be exited, and selling tickets for a new route before entry.

Fourth, while the response variables (be it a price or a quantity measure) are measured at the route level for a given incumbent, $LCCthreat$ and $LCCpres$ are the consequence of the complex network design of the entrant, so rejecting the exogeneity assumption in this specific case would amount to saying that the responses of a single route would lead to a network redefinition by the entrant.

Fifth, exogeneity is consistent with Gol's entry pattern described in Section 1.2.2 – the underlying behavioural assumption is that the "radial" entry pattern of Gol starting from Sao Paulo, is not influenced by the reactions to entry. In other words, Gol has to enter the key markets (starting from Sao Paulo) if it is to succeed.

Another argument that needs to be addressed is the fact that route profitability is an unobserved variable (at least to the econometrician), implying that least squares renders inconsistent estimates. To account for that, previous studies since at least Windle and Dresner (1999) and Morrison (2001) have invested in careful modelling of fixed-effects. More recently, Gooslbee and Syverson (2005) put forth a model similar to my P (Preemptive) specification, whereas their 2008 paper is a particular case of my PU specification, with route fixed-effects γ_{iod} and incumbent-time fixed effects, μ_{it}. The way I specify my general model (PUD, but also PU) generalizes these earlier contributions by controlling also for time-varying market effects (profitability, in particular) using time-origin and time-destination fixed-effects, $v_{o,t}$ and $v_{d,t}$, respectively. All in all, my final specification thus captures time-varying market unobservables and time-varying carrier unobservables (unobserved quality, for instance). If the route

[8]As stated by Busse (2002, p. 304) in her study of airline price wars, "(...) airlines set their schedules well in advance of the actual flight; most airlines begin selling tickets six months before a flight." Busse also stresses the existence of airport slot restrictions, timing constraints imposed by by hub-and-spoke networks, and the necessity to physically coordinate with other (partner) airlines.

[9]When it comes to planning its fleet size, Gol's first annual report had a planning horizon of five years, while in its 2007 one it goes until 2014. Moreover, it started operating with a fleet of 6 aircraft in January 2001, which it increased to 10 by the end of 2001, 19 by the end of 2002, 22 by the end of 2003, 27 by the end of 2004 and 41 by the end of 2005 - see Souza (2006).

fixed-effects γ_{iod} capture market heterogeneity constant over time (such as the fact that a given route connects two state capitals), the carrier-time fixed-efects μ_{it} capture, for instance, varying financial conditions of an airline (bankruptcy, in particular), where the time-varying origin and destination fixed-effects capture changes in economic conditions (such as economic growth in the countryside due to the commodities boom), seasonality patterns (or tourism effects during holidays), and profitability.

One important variable not observed but which can be controlled for by the above fixed-effects is airline quality (one can think about leg room, or comfort in general to fix ideas). If quality is time-varying, but homogeneous across routes of a given carrier, then it is controlled for by μ_{it}. In practice, quality tends to be quite stable over time for a given route, since fleets tend to be standardized to minimize costs and ordered well in advance (at least 2-3 years). What might vary is the difference in quality between an incumbent and the LCC, and this will be accounted for by using alternative variables in the interaction term $Z_{iod,t}LCCpres_{od,t}$. In particular, I use flight distance as a control, since the difference in quality tends to increase in longer flights.

1.6 Estimation Results and Analysis

This section reports results on price and quantity reactions of the incumbents given the threat of entry and the actual presence of the LCC Gol. I start by estimating price reactions and the associated responses in demand (ticket sales), controlling for costs and alternative measures of product differentiation. Next I focus on different measures of quantity responses (responses in the supply of seats by each incumbent at every route): first, I investigate excess capacity by estimating the responses in aircraft size and (a lower bound on) the number of flights; second, I investigate product repositioning by estimating the responses in the share of seats supplied in weekdays/weekends and peak- and off-peak-time flights.

1.6.1 Incumbents Do React Preemptively Using Prices

Baseline Specification (B)

The estimates from the baseline specification are hard to justify in economic terms: they show a 14% fare reduction at the period Gol starts flying a route with no response in the number of tickets sold for five months, as reported in Table 1.3.[10]

[10] This results implies that demand reacts very inelastically to fare reductions, a fact at odds with findings for the Brazilian market, where average own- and cross-price elasticities are in the range 2-4 (in absolute values) and 0.4-1.8, respectively, see Oliveira (2007).

Preemptive Specification (P)

When investigating preemptive effects, I consider a number of measures of neighbourhood or area of influence. First, I assume a route to be threatened whenever Gol entered the geographic region in either of the endpoint airports of a route, but not the route itself. Brazil is divided into five geographic regions (South, Southeast, Northeast, North and Midwest), so this is a very broad concept of neighbourhood. I also consider an alternative definition of regions by defining the coastal area into three regions and the inner part of the country into two regions (inner-North and inner-South), but the results were identical to that of geographic regions.[11] I then focused on more localized measures of area of influence: after obtaining the geographic coordinates of each airport and computing the distance between them, I defined an area of influence (or neighbourhood) to be a circle centered at a given airport and with radius d.[12] Following this reasoning, whenever Gol operates within an area of influence, but not at a given route, this route is considered to be under threat. To this end, I define radii of 500km and 1,000km . Overall, the results for the $d_{1,000}$ criterion do not differ significantly from that of region, but those for d_{500} tend to show stronger yet sometimes more unstable reactions to entry; this can be partly attributed to the fact that by considering a smaller neighbourhood one is effectively using a smaller sample, but the increasing reactions to an ever approaching threat by a competitor also makes economic sense.[13] So, for every route, we investigate whether there were reactions to the threat of entry going back up to four months using four alternative measures of area of influence.

Although one might argue that four months is not a long enough window, I choose to do so for a number of reasons. First, I am constrained by the data, since the first period an entry occurs in the dataset is period four; by increasing the number of lags, I am at risk of dropping important information and reducing the sample, a critical issue especially for the d_{500} area of influence.[14] Second, given the historically high Brazilian (real) interest rates, it would be surprising to see significant fare reductions many periods before entry, since an incumbent following this strategy would forego a substantial amount of revenues.[15] Third, the time it takes between Gol requiring to operate a route

[11] When the area of influence is defined as the geographic region, there are 5 airports in the Southern region, 11 in the southeast, 4 in the Northeast, 3 in the North and 2 in the Midwest. When considering the alternative criterion of zones of influence, I divide the coastal cities into three regions (with 5 airports in the Southern region, 3 in the Southeast and 5 in the Northeast) and divided the inner part of the country into 2 regions, "inner-North", with 7 airports, and "inner-South", with 5 airports. The results are robust to these definitions.

[12] The use of areas of influence or neighbourhoods is by no means new – see for instance Davis (2006), who studies demand for movie theathres in the US accounting for their location.

[13] Of the 12, 22 and 24 airports which have neighbourhoods within, respectively, 250km, 500km and 1,000km, only 5, 11 and 12 are threatened by Gol.

[14] The results of a larger number of lags for the $d_{1,000}$ and region neighbourhood definitions shows that, by and large, fare cuts and quantity increases occur the same way as for the specifications whose results we report.

[15] Alternatively, in the context of a repeated game where players have a discount $\delta := \frac{1}{1+r}$, a high (real) interest rate implies a low discount factor, so players give less importance to future payoffs ie. are impatient. As a result, foregoing profits today to avoid a future loss of profits is less likely to happen.

and the regulator authorizing it is roughly one month, as discussed above. Finally, the state of financial fragility of the incumbents makes it hard for them to sustain aggressive fare cuts for long periods.

When allowing for preemption one obtains a fare reduction of about 21% at the month *before* Gol starts flying a route, as reported in Table 1.4. At the period entry occurs, the incumbents reduce fares again by 16-20%, which is followed by another reduction of about 6% four periods after entry, suggesting that demand is indeed price-sensitive.

Together with the fare reduction prior to actual entry, there is an associated increase in ticket sales ranging from 31-52%, depending on the area of influence criterion considered. There is no significant effect at the month Gol enters a route, but further (marginally significant) increases occur two and five periods after entry.

Overall, the results for both price and ticket sales are robust to the choice of area of influence. Importantly, the preemption-related coefficients of the price equations are not only significant, but are also larger than those measuring fare reduction at entry, and the reaction in demand responds accordingly.

Preemptive Specification Controlling for Time-Varying Unobservables (PU)

The PU estimates suggest a significant fare reduction at the month prior to entry (24-36%), followed by more moderate reactions at the period of entry (18-12%) and four and five periods after the actual entry of Gol (3-5% and 6-9%, respectively), as reported in Table 1.5. As in the case of the Preemptive model, the effect of preemption on fares dominates that of post-entry reaction.

The response in ticket sales occurs only for the models with a wider area of influence ($d_{1,000}$ and region), being insignificant whenever d_{500}. Although no model identifies reactions in quantity at the period Gol actually starts operating a route, all of them measure increases two and five periods after actual entry (25-28% and 15-21%, respectively). As opposed to what happens with price, the post-entry responses in demand seem to dominate the preemptive effects, suggesting that they are somewhat sluggish.

By controlling for time-varying market unobservables, the results for the PU specification suggest that the P specification tends to underestimate preemptive reactions in prices and overestimate reactions to actual entry at time t^*. On the other hand, when it comes to quantity reactions the Preemptive specification tends to overestimate the preemptive responses and to underestimates the post-entry reactions.

1.6.2 Product Differentiation Might Soften Price Responses

Preemptive Specification Controlling for Time-Varying Unobservables and Product Differentiation (PUD)

I now report the estimates for the PUD specification. Besides showing how product differentiation can soften both price and quantity responses, the results below are more in line with economic intuition.

I account for product differentiation using two alternatives. First, I interact LCC presence with flight distance, since the increased comfort offered by a legacy carrier as compared to a LCC tends to have a larger impact the longer the flight takes. Second, I also construct product differentiation indices using the distance-metric variables of Slade (2004) and Pinkse, Slade and Brett (2002). These variables measure the distance between products in the space of characteristics, thus providing an index of product differentiation. In what follows, I use the distance-metric between the characteristic x of two products i and j defined by

$$wx_{ij} = dm(x_i, x_j) = -\frac{3}{2}\left[\frac{1}{1+2|x_i - x_j|} - 1\right]$$

where the characteristics are $nssp_{iod,t}$ and $nsscc_{iod,t}$ which are, respectively, the share of non-stop seats during peak time of the i-th incumbent on the od-th city-pair at month t and the share of non-stop seats linking central airports of the i-th incumbent on the od-th city-pair at month t, both of which are proxies for convenience and level of service of airline i at a route.

The estimates of distance only and distance plus distance-metric variables are reported in Tables 1.6 and 1.7, respectively. In the former, there is mixed evidence about the significance of distance across alternative measures of neighbourhood for prices, whereas in the latter the distance-metric variables are always significant. Despite the concerns regarding the endogeneity of the distance-metric variables when using shares as characteristics, the results with and without distance-metric proxies send the same basic message: incumbents respond strongly in prices, both before and at the period Gol starts operating a route. The post-entry effects are concentrated two months after Gol's entry occurred. The main difference induced by the inclusion of the distance-metric variables is the stronger price response at the period of entry.

The estimates for the PUD specification are relatively robust to the choice of the area of influence: as reported in Table 1.6, incumbents tend to preempt by reducing fares by 25-35% at the month before Gol starts flying a route. Then, given the actual entry of the LCC, incumbents reduce fares again by 20-25%, but tend not to react in prices in the periods following entry. Product differentiation tends to soften these price reactions by about 8% per 1,000km flown.

The impact on quantities also tends to precede LCC entry: for the wider areas of influence ($d_{1,000}$ and region), the increase of the number of tickets sold at the period immediately before actual entry of the LCC is in the range 14-22%, and this increase is going to be at least 30% once Gol starts flying a route.[16] Except for the second month after entry, when quantities increase by about 23%, the quantity responses are not significant up to period $t^* + 5$. As in the case of prices, product differentiation (or, more precisely, the interaction between LCC presence and distance) softens the response in quantities by 23-32%, which is slightly less than the quantity increase observed at the period t^* of entry.

When comparing the PUD and PU estimates, the latter seems to underestimate price reactions at the period Gol starts flying a route and to overestimate price reactions after entry (price reactions at period $t^* + 5$ are significant for the PU specification). Moreover, the PU specification does not detect any quantity reactions at the period of entry (especially worrying given the price reactions at periods $t^* - 1$ and t^* itself) and again overestimate the reactions at period $t^* + 5$. When comparing the alternative criteria of area of influence, the PU specification results suggests that price reactions at the period of entry are stronger the larger the area of influence, exactly the opposite result obtained for the PUD specification.

The findings above suggest that there occurs preemption in the form of fare reductions at the period immediately before entry of the LCC in a route. This is followed by further fare reductions at the period of entry, then by accommodation after entry has occurred. Regardless of the measure of area of influence considered, the preemptive reductions exceed those occurring at entry by at least 20%. Quantity responds to the preemptive fare reductions at the month before entry and again two months after entry has occurred. Product differentiation will, however, soften price responses, suggesting that most of the competitive action happens at shorter routes where the schedules of the entrant is similar to those of the incumbents. Finally, price and quantity responses tend to be stronger for the smaller areas of influence ie. the closer Gol gets, the stronger the response of the established players.

1.6.3 Price Responses Are Not Cost-Driven

So far I have abstracted from costs when investigating price responses, and Table 1.8 reports that costs do not drive price responses.

The cost information available is, for a given route and incumbent, the average operating costs excluding (financial variables such as) depreciation, amortization, leasing and insurance, thus being

[16] The quantity estimates for the smaller areas of influence look less plausible. In particular, the results for the d_{250} specification, according to which quantities decrease by 33% two months before entry and then increase by 79% one month before entry might be caused by the small number of airports/routes involved.

closely related to route an airport costs. The results should also reflect the good job done by the fixed-effects in controlling for unobservables.

1.6.4 Excess Capacity?

I do not have information on the amount of seats supplied by each incumbent in a given route, so I resort to indirect information to try to learn about the existence of preemption in the form of excess capacity. Basically, incumbents can generate excess capacity for a route either by increasing aircraft size (replacing a 95-seater with, say, a 120- or 150-seater) or by adjusting the network, which consists of creating (or rescheduling) flights.

I start by using information on the average aircraft size used by incumbent i at route *od*. Overall, there is an increasing trend in aircraft size in the dataset regardless of Gol's presence, but using the framework of the PUD specification I investigate whether there is a response in terms of aircraft utilization to the threat of entry of Gol. Following this reasoning, an incumbent feeling threatened by the LCC could, even in the short run, use larger aircraft in a route in an attempt to dissuade Gol from operating this route. The results in Table 1.9 show no evidence of preemptive behaviour of that sort. Although there are responses at the period when Gol starts flying a route, they are marginal – the 4-5% coefficients correspond to roughly six seats in the median 118-seater in the data and ten seats in the case of the biggest aircraft observed in the sample. What incumbents really tend to look at when determining aircraft size for a route are costs and flight distance. I thus conclude that there are no substantial adjustments in aircraft size given the potential entry of Gol.

I then have a second look at excess capacity by deriving an estimate of the number of flights in a route using data on aircraft size and the number of tickets sold by each incumbent at a given period and route. By dividing the number of tickets by the aircraft size, I get a lower bound on the number of flights (LBF) each incumbent is operating on a route at a given period (this is the number of fully booked flights each incumbents would need to satisfy its demand).

The results from a regression using the natural logarithm of LBF as the dependent variable are reported in Table 1.9. The results show a 16-26% increase at the period Gol starts flying a route plus another 24-29% increase two months afterwards, with no significant evidence for action before Gol enters the route. I then go back to the price regressions (Table 1.6) and note that there were fare cuts at periods $t^* - 1$ (between 13% and 37%, depending on the measure of neighbourhood used) and t^* (between 14% and 19%), so I interpret the changes in LBF as being fuelled by these price changes.

1.6.5 Network Adjustment and Product Repositioning

I now use the information on the shares of seats supplied to investigate a potential repositioning of the incumbent carriers given the threat of Gol's entry. Although I do not observe the quantity of seats supplied, I observe shares of seats offered at (1) peak time during weekdays; (2) off-peak time during weekdays; (3) peak time during weekends; and (4) off-peak time during weekends and the typical route sees a clearly decreasing pattern for (1)-(4).

The results reported in Tables 1.10 and 1.11 show no evidence of any meaningful preemptive adjustment in any of the shares, and not much action at all in what regards the weekend shares, either pre-, at or post-entry. There is, however, some evidence of an increase in the share of peak time flights during weekdays and a decrease in the share of off-peak flights also on weekdays, both of which tend to be followed by a further increase and decrease, respectively, two months after the entry of Gol. Given the tendency of the LCCs in general, and Gol in particular, to operate in off-peak times, this is suggestive of accommodation by the incumbents, as they are avoiding head-to-head competition with the entrant by focusing on the supply of flights more conveniently scheduled (from the traveller's viewpoint), and which are more highly priced.

1.7 Conclusion

This paper studied the reactions by incumbent airlines to the threat and actual entry of the low-cost carrier Gol in the Brazilian domestic air transport market.

The results show evidence of preemptive behaviour by the incumbents in the form of fare cuts. Although the incumbents do respond to the threat of entry ie. to the fact that Gol is operating within the area of influence of a route, and to actual entry, they do not sustain these responses post-entry, in what I interpret as accommodation by the incumbents. Importantly, the responses to potential entry are at least as important as the ones to actual entry.

There is also evidence of repositioning in the supply of seats, with an increase of peak time flights on weekdays and a decrease of off-peak time flights also on weekdays, in what suggests an attempt to avoid head-to-head competition with the entrant.

The results show the importance of controlling for time-varying market unobservables, that costs are not the driving force behind price reactions and that product differentiation tends to soften price responses, suggesting that most of the competitive action tends to occur in short-haul routes.

On a more general level, the results of the paper suggest that reactions to entry are not only preemptive with respect to entry at the route itself, but also with respect to entry at the airport, a traditional predictor of route entry in the literature.

TABLE 1.1. Operational Information of the Main Airlines in the Brazilian Domestic Market – 2002

Item	Unit	Gol	TAM	VARIG	VASP
Revenue per Passenger-Kilometers (RPK)	Passenger×km (bn)	3.22	9.34	10.48	8.89
RPK Market Share	Fraction	0.12	0.35	0.39	0.13
Traffic per Employee	Passenger×km (mn)	1.56	1.23	0.75	0.70
Load Factor	Fraction	0.68	0.53	0.59	0.55
Unit Cost	BRL/Passenger×km	0.20	0.33	0.33	0.31
Yield	BRL/Passenger×km	0.21	0.29	0.31	0.27
Operating Margin	Fraction	0.06	−0.12	−0.05	−0.16

Note: This table compares operational characteristics of the LCC Gol to those of the incumbent airlines in the Brazilian market. BRL is the Brazilian currency, the Real. Passenger indicates the number of revenue-passengers traveled whereas RPK is defined as revenue-passenger times kilometers. The market share is defined as the ratio of the firm's RPK over the RPK of the industry. Operating margin is defined as the ratio of operating profits (or losses) over total revenues.

TABLE 1.2. Evolution of Gol's Actual Presence in the Sample

		Sample	
	Initial Period	Mid-period	Final Period
Gol's Presence	September 2001	December 2002	March 2004
Routes Served	51	70	76
% Sample Routes	54	74	81
Cities Served	11	22	25
% Sample Cities	38	76	90

1.A Appendix: Variable Description

This Appendix describes the variables used in the study.

- $yield_{iod,t}$: is the natural logarithm of average yield (revenue per passenger-kilometers) of the i-th incumbent on the od-th route at month t. A route is defined as a directional (one-way) city-pair. Yields are expressed in local currency (BRL – Brazilian Real) and are related to all tickets sold in a given period for a given airline on a given route. Information available in DAC's (non-published) Average Yield of Monitored Airport-Pairs Report. Only major incumbent's average yields are considered. Average yields are not disaggregated by fare class. Yields were deflated by a consumer price index, IPCA (source: IBGE).
- $LCCpres_{od,t^*+\tau}$: a dummy variable that controls for the route presence of the entrant Gol Airlines at time $t^* + \tau$. t^* is the time period in which it actually establishes presence (flights) on the route. Information available in the HOTRAN reports.

TABLE 1.3. Baseline Specification Results: Price and Demand Responses

	Price	Ticket Sales
Variable	**Time+Route FE**	**Time+Route FE**
LCCpres$_{t^*}$	-0.137**	-0.064
	(-3.31)	(-0.73)
LCCpres$_{t^*+1}$	-0.016	0.065
	(-0.89)	(0.96)
LCCpres$_{t^*+2}$	0.029	0.065
	(0.73)	(0.91)
LCCpres$_{t^*+3}$	-0.011	-0.127
	(-0.24)	(-1.26)
LCCpres$_{t^*+4}$	-0.032**	0.005
	(-2.27)	(0.08)
LCCpres$_{t^*+5}$	-0.014	0.047*
	(-0.55)	(0.79)
R-squared	0.834	0.923

Note: This table reports estimates of the Baseline specification for (logarithms of) prices and demand (ticket sales) as the dependent variables. LCCpres denotes the actual entry of the LCC on a route. Standard errors are clustered by route to control for intertemporal correlation, and the corresponding t-statistics are reported within parentheses. Significance at the 10% and 5% significance levels are denoted by * and ** respectively. Fixed effects not reported.

- $LCCthreat_{od,t^*-\tau}$: a dummy variable that controls for the threat of route presence of the entrant Gol Airlines. A route is threatened if (1) it operates in an airport in the same geographic region as either airport o or airport d (this leads to the "region" concept); (2) if it operates in the same zone of influence of either airport o or airport d (this leads to the same results as the region concept - results not reported); (3) it operates whithin 250/500/1,000km as either airport o or airport d (this leads to the variables d_{250}, d_{500} and $d_{1,000}$).

- $nssp_{iod,t}$: the share of non-stop seats during peak time of the i-th incumbent on the od-th city-pair at month t (= number of non-stop and peak-time seats of airline i over number of total non-stop and peak-time seats in the market). It is a proxy for convenience and level of service of airline i at a route. For this calculation, "peak time" was defined considering all flights with departure within 5am to 10am (morning peak) and 4.30pm to 10pm (evening peak) on weekdays, and those with departure from 7pm to 10pm on Sundays. DAC's HOTRAN Report provides the information of flight number / weekdays / departure times, which made possible the segregation into "peak" and "off-peak" periods.

- $nsscc_{iod,t}$: the share of non-stop seats linking central airports of the i-th incumbent on the od-th city-pair at month t (= number of non stop seats of airline i linking airports close to the city center over number of total non stop seats linking airports close to the city center in the market). Also a proxy for convenience and level of service of airline i. Airports close to the city center are: São Paulo's Congonhas (CGH), Rio de Janeiro's Santos Dumont (SDU) and Belo Horizonte's Pampulha

TABLE 1.4. Preemptive Specification Results: Price and Demand Responses

	Price			Ticket Sales		
Variable	d_{500}	$d_{1,000}$	**Region**	d_{500}	$d_{1,000}$	**Region**
LCCthreat$_{t^*-4}$	0.036	0.031	0.085**	-0.002	-0.046	-0.018
	(1.41)	(1.27)	(3.50)	(-0.03)	(-0.75)	(-0.29)
LCCthreat$_{t^*-3}$	-0.007	0.001	-0.005	0.066	0.065	0.094
	(-0.26)	(0.02)	(-0.22)	(0.75)	(0.82)	(1.20)
LCCthreat$_{t^*-2}$	0.006	0.024	0.023	-0.189*	-0.140	-0.128
	(0.19)	(0.89)	(0.84)	(-1.95)	(1.47)	(1.33)
LCCthreat$_{t^*-1}$	-0.210**	-0.207**	-0.210**	0.400**	0.340**	0.315*
	(-4.69)	(-4.80)	(-4.04)	(4.12)	(3.45)	(2.21)
LCCpres$_{t^*}$	-0.167**	-0.189**	-0.185**	-0.011	0.045	0.060
	(-4.28)	(-5.33)	(-4.30)	(-0.11)	(0.57)	(0.64)
LCCpres$_{t^*+1}$	-0.018	-0.014	-0.015	0.059	0.053	0.051
	(-0.97)	(-0.74)	(-0.76)	(0.85)	(0.81)	(0.76)
LCCpres$_{t^*+2}$	-0.013	-0.001	-0.007	0.105*	0.109*	0.119*
	(-0.47)	(-0.03)	(-0.31)	(1.70)	(1.85)	(1.84)
LCCpres$_{t^*+3}$	-0.045	-0.038	-0.036	-0.176	-0.163	-0.156
	(-0.98)	(-0.92)	(-0.91)	(-1.60)	(-1.64)	(-1.60)
LCCpres$_{t^*+4}$	-0.060**	-0.068**	-0.060*	0.039	0.073	0.065
	(-2.77)	(-3.38)	(-3.19)	(0.52)	(1.00)	(0.86)
LCCpres$_{t^*+5}$	-0.060*	-0.048	-0.073**	0.105*	0.057*	0.069
	(-1.71)	(-1.51)	(-2.76)	(1.94)	(0.94)	(1.11)
R-squared	0.835	0.835	0.834	0.923	0.923	0.923

Note: This table compares alternative versions of the Preemptive specification for (logarithms of) prices and demand (ticket sales) as the dependent variables. LCCthreat denotes the fact that Gol is operating within a neighbourhood (measured by distance within 500 and 1,000km from a given airport) but not on the airport itself. LCCpres denotes the actual entry of the LCC on a route. Standard errors are clustered by route to control for intertemporal correlation, and the corresponding t-statistics are reported within parentheses. Significance at the 10% and 5% significance levels are denoted by * and ** respectively. Fixed effects not reported.

(PLU); these cities have one downtown airport and one international airport. Information available in the HOTRAN reports.

- Code Share$_{iod,t}$: a dummy variable indicating a code share agreement between Tam and Varig. cdshare = 1 from March 2003 to the end of the sample and only for routes in which both Tam and Varig had flights. Route-specific data on flight operations was collected from DAC's HOTRAN, "Horário de Transporte", a data system that generates reports containing information of all scheduled flights within the country (non-published information), disaggregated by airline/flight code. Therefore, data on flight frequency and number of available seats were accessible on a monthly basis with information from HOTRAN being collected for the mid-point day of each month.
- The shares of peak time and off-peak time flights on weekdays and weekends of the i-th incumbent on the od-th city-pair at month t.
- Average aircraft size of the i-th incumbent on the od-th city-pair at month t.

TABLE 1.5. PU Specification Results: Price and Demand Responses

	Price			Ticket Sales		
Variable	d_{500}	$d_{1,000}$	**Region**	d_{500}	$d_{1,000}$	**Region**
LCCthreat$_{t^*-4}$	0.035**	0.035**	0.038**	-0.085	-0.021	-0.001
	(2.72)	(2.82)	(3.24)	(-1.14)	(-0.29)	(0.02)
LCCthreat$_{t^*-3}$	-0.011	-0.005	-0.007	0.047	0.041	0.076
	(-0.40)	(-0.21)	(-0.28)	(0.57)	(0.56)	(1.02)
LCCthreat$_{t^*-2}$	0.029**	0.031**	0.029**	0.064	0.038	0.042
	(2.40)	(2.88)	(2.87)	(-1.06)	(0.65)	(0.72)
LCCthreat$_{t^*-1}$	-0.328**	-0.248**	-0.244**	0.206	0.250*	0.238*
	(-6.19)	(-5.33)	(-4.73)	(1.31)	(1.96)	(1.65)
LCCpres$_{t^*}$	-0.100**	-0.109**	-0.120**	-0.026	0.028	0.068
	(-3.26)	(-3.14)	(-3.22)	(-0.28)	(0.32)	(0.71)
LCCpres$_{t^*+1}$	-0.030	-0.009	-0.010	0.084	0.063	0.062
	(-1.30)	(-0.42)	(-0.47)	(1.50)	(1.23)	(0.71)
LCCpres$_{t^*+2}$	-0.010	-0.013	-0.010	0.284**	0.264**	0.246**
	(-0.55)	(-0.58)	(-0.42)	(2.22)	(2.12)	(2.07)
LCCpres$_{t^*+3}$	-0.012	-0.013	-0.013	-0.078	-0.076	-0.073
	(-0.53)	(-0.50)	(-0.51)	(-1.09)	(-1.09)	(-1.10)
LCCpres$_{t^*+4}$	-0.030**	-0.049*	-0.048*	0.008	0.059	0.069
	(-1.99)	(-1.86)	(-1.82)	(0.14)	(0.92)	(1.17)
LCCpres$_{t^*+5}$	-0.094**	-0.067**	-0.065**	0.211**	0.163*	0.147*
	(-3.10)	(-2.04)	(-2.01)	(2.31)	(1.81)	(1.73)
R-squared	0.933	0.933	0.934	0.950	0.950	0.950

Note: This table compares alternative versions of the PU specification for (logarithms of) prices and demand (ticket sales) as the dependent variables. LCCthreat denotes the fact that Gol is operating within a neighbourhood (measured by distance within 500 and 1,000km from a given airport) but not on the airport itself. LCCpres denotes the actual entry of the LCC on a route. Standard errors are clustered by route-carrier to control for intertemporal correlation, and the corresponding t-statistics are reported within parentheses. Significance at the 10% and 5% significance levels are denoted by * and ** respectively. Fixed effects not reported.

- $\text{Cost}_{iod,t}$ the average operating costs excluding depreciation, amortization, leasing and insurance for the i-th incumbent on the od-th city-pair at month t.
- Distance_{od}: the distance (in 1,000km) between origin and destination, computed using the geographic coordinates of the airports o and d.

TABLE 1.6. PUD Specification Results: Price and Demand Responses – Distance Variable

	Price			Ticket Sales		
Variable	d_{500}	$d_{1,000}$	**Region**	d_{500}	$d_{1,000}$	**Region**
$\mathbf{LCCthreat}_{t^*-4}$	0.036**	0.035**	0.036	-0.087	-0.024	0.045
	(2.77)	(2.86)	(0.79)	(-1.16)	(-0.35)	(0.83)
$\mathbf{LCCthreat}_{t^*-3}$	-0.012	-0.006	-0.094	0.054	0.045	-0.039
	(-0.45)	(-0.24)	(2.47)	(0.64)	(0.61)	(-0.44)
$\mathbf{LCCthreat}_{t^*-2}$	0.028**	0.030**	0.035	0.066	0.040	0.026
	(2.35)	(2.81)	(0.92)	(1.09)	(0.68)	(0.24)
$\mathbf{LCCthreat}_{t^*-1}$	-0.337**	-0.242**	-0.128**	0.238	0.216*	-0.063
	(-6.65)	(-4.84)	(-2.15)	(1.56)	(1.67)	(-0.63)
$\mathbf{LCCpres}_{t^*}$	-0.186**	-0.159**	-0.141**	0.310**	0.305**	0.211**
	(-3.40)	(-3.21)	(-2.46)	(2.80)	(2.68)	(2.31)
$\mathbf{LCCpres}_{t^*+1}$	-0.025	-0.006	-0.021	0.065	0.046	0.059
	(-1.13)	(-0.27)	(-0.89)	(1.28)	(0.88)	(1.22)
$\mathbf{LCCpres}_{t^*+2}$	-0.003	-0.006	-0.021	0.233*	0.227*	0.265**
	(-0.16)	(-0.26)	(-0.81)	(1.96)	(1.94)	(2.12)
$\mathbf{LCCpres}_{t^*+3}$	-0.013	-0.013	-0.005	-0.076	-0.075	-0.074
	(-0.56)	(-0.50)	(-0.13)	(-1.08)	(-1.07)	(-0.87)
$\mathbf{LCCpres}_{t^*+4}$	-0.028*	-0.046*	-0.018	0.002	0.042	-0.035
	(-1.95)	(-1.80)	(-0.76)	(0.03)	(0.70)	(-0.58)
$\mathbf{LCCpres}_{t^*+5}$	-0.058*	-0.046	-0.063	0.069	0.042	0.065
	(-1.86)	(-1.22)	(-1.78)	(0.73)	(0.48)	(0.70)
LCCpres×Distance	0.083**	0.051	0.052	-0.324**	-0.284**	-0.313**
	(2.31)	(1.35)	(1.35)	(-4.52)	(-3.73)	(-4.14)
Code-share	-0.070	-0.070	-0.070*	0.166	0.166	0.166
	(-1.60)	(-1.61)	(-1.61)	(1.46)	(1.46)	(1.46)
R-squared	0.935	0.933	0.933	0.950	0.950	0.950

Note: This table compares alternative versions of the PUD specification for (logarithms of) prices and demand (ticket sales) as the dependent variables. LCCthreat denotes the fact that Gol is operating within a neighbourhood (measured by distance within 500 and 1,000km from a given airport) but not on the airport itself. LCCpres denotes the actual entry of the LCC on a route. Standard errors are clustered by route-carrier to control for intertemporal correlation, and the corresponding t-statistics are reported within parentheses. Significance at the 10% and 5% significance levels are denoted by * and ** respectively. Fixed effects not reported.

1.B Appendix: Endogeneity Issues

1.B.1 Identification

So far, I have estimated models such as the PUD specification,

$$\begin{aligned} y_{iod,t} &= \gamma_{iod} + \mu_{it} + v_{o,t} + v_{d,t} + \sum_{\tau=0}^{K} \beta_\tau LCCpres_{od,t^*+\tau} + \sum_{\tau=1}^{L} \delta_\tau LCCthreat_{od,t^*-\tau} + \\ & +Z_{iod,t} LCCpres_{od,t} \rho + X_{iod,t}\alpha + \varepsilon_{iod,t} \end{aligned}$$

assuming that both the threat and the entry of Gol are exogenous, as in much of the literature, see eg. Goolsbee and Syverson (2008). If, however, this reasoning is not valid, the OLS method can generate

TABLE 1.7. PUD Specification Results: Price and Demand Responses – Distance-Metric Variables

	Price			Ticket Sales		
Variable	$\mathbf{d_{500}}$	$\mathbf{d_{1,000}}$	**Region**	$\mathbf{d_{500}}$	$\mathbf{d_{1,000}}$	**Region**
LCCthreat$_{t^*-4}$	0.037**	0.037**	0.040**	-0.090	-0.027	-0.009
	(2.88)	(2.91)	(3.40)	(-1.20)	(-0.39)	(-0.14)
LCCthreat$_{t^*-3}$	-0.013	-0.006	-0.007	0.054	0.045	0.078
	(-0.49)	(-0.26)	(-0.31)	(0.65)	(0.61)	(1.05)
LCCthreat$_{t^*-2}$	0.028**	0.030**	0.029**	0.066	0.040	0.046
	(2.32)	(2.73)	(2.79)	(-1.11)	(-0.70)	(-0.80)
LCCthreat$_{t^*-1}$	-0.350**	-0.257**	-0.262**	0.249	0.222*	0.143*
	(-7.39)	(-5.22)	(-3.92)	(1.59)	(1.67)	(0.80)
LCCpres$_{t^*}$	-0.242**	-0.214**	-0.191**	0.342**	0.336**	0.300**
	(-4.74)	(-4.63)	(-4.00)	(2.90)	(2.79)	(2.58)
LCCpres$_{t^*+1}$	-0.024	-0.005	-0.008	0.065	0.045	0.047
	(-1.10)	(-0.02)	(-0.36)	(1.27)	(0.86)	(0.92)
LCCpres$_{t^*+2}$	-0.003	-0.012	-0.012	0.234**	0.229*	0.231**
	(-0.18)	(-0.54)	(-0.55)	(2.00)	(1.96)	(2.01)
LCCpres$_{t^*+3}$	-0.016	-0.016	-0.017	-0.073	-0.072	-0.071
	(-0.76)	(-0.66)	(-0.68)	(-1.06)	(-1.05)	(-1.04)
LCCpres$_{t^*+4}$	-0.024	-0.044*	-0.046*	0.001	0.042	0.037
	(-1.64)	(-1.67)	(-1.75)	(0.01)	(0.69)	(0.58)
LCCpres$_{t^*+5}$	-0.048	-0.035	-0.048	0.067	0.040	0.054
	(-1.61)	(-0.96)	(-1.27)	(0.71)	(0.45)	(0.60)
LCCpres×Distance	0.085**	0.051	0.013	-0.320**	-0.280**	-0.234**
	(2.11)	(1.20)	(0.26)	(-4.32)	(-3.52)	(-2.42)
LCCpres×wnssp	0.042**	0.039**	0.040**	-0.056	-0.053	-0.054
	(2.18)	(2.00)	(2.05)	(-0.65)	(-0.62)	(-0.62)
LCCpres×wnscc	0.100**	0.102**	0.103**	-0.007	-0.011	-0.011
	(4.82)	(4.89)	(4.92)	(-0.13)	(-0.20)	(-0.21)
Code-share	-0.089*	-0.090*	-0.090*	0.164	0.166	0.166
	(-1.82)	(-1.84)	(-1.84)	(1.43)	(1.44)	(1.45)
R-squared	0.935	0.934	0.934	0.950	0.950	0.950

Note: This table compares alternative versions of the PUD specification for (logarithms of) prices and demand (ticket sales) as the dependent variables. LCCthreat denotes the fact that Gol is operating within a neighbourhood (measured by distance within 500 and 1,000km from a given airport) but not on the airport itself. LCCpres denotes the actual entry of the LCC on a route. Standard errors are clustered by route-carrier to control for intertemporal correlation, and the corresponding t-statistics are reported within parentheses. Significance at the 10% and 5% significance levels are denoted by * and ** respectively. Fixed effects not reported.

inconsistent estimates. To address such a problem, I also use instrumental variables to estimate the reactions to entry using the two-step GMM estimator with standard errors clustered at the route-level.

The first task is to devise an identification strategy under endogeneity using data at a quite disaggregated level (regional Brazilian data, municipalities in particular) which is observed at a high enough frequency during the sample period (September 2001-March 2004) to be highly correlated with the 0-1 decision of the LCC to enter a route, but not with the error term. To this end I resort to the data from the Brazilian Statistics Bureau (Instituto Brasileiro de Geografia e Estatística - IBGE) and the

TABLE 1.8. PUD Specification Results: Price Responses – Cost Controls

	Price		
Variable	d_{500}	$d_{1,000}$	**Region**
LCCthreat$_{t^*-4}$	0.037**	0.036**	0.036
	(2.82)	(2.89)	(0.80)
LCCthreat$_{t^*-3}$	-0.012	-0.006	-0.096**
	(-0.44)	(-0.23)	(2.58)
LCCthreat$_{t^*-2}$	0.028**	0.030**	0.032
	(2.33)	(2.78)	(0.85)
LCCthreat$_{t^*-1}$	-0.338**	-0.242**	-0.126**
	(-6.65)	(-4.82)	(-2.12)
LCCpres$_{t^*}$	-0.186**	-0.159**	-0.141**
	(-3.41)	(-3.21)	(-2.46)
LCCpres$_{t^*+1}$	-0.024	-0.006	-0.020
	(-1.09)	(-0.25)	(-0.86)
LCCpres$_{t^*+2}$	-0.003	-0.006	-0.021
	(-0.18)	(-0.25)	(-0.81)
LCCpres$_{t^*+3}$	-0.014	-0.014	-0.006
	(-0.60)	(-0.54)	(-0.15)
LCCpres$_{t^*+4}$	-0.028*	-0.046*	-0.017
	(-1.94)	(-1.79)	(-0.75)
LCCpres$_{t^*+5}$	-0.059*	-0.047	-0.064*
	(-1.89)	(-1.25)	(-1.81)
LCCpres×Distance	0.083**	0.051	0.053
	(2.32)	(1.36)	(1.35)
Code-share	-0.069	-0.069	-0.069
	(-1.54)	(-1.55)	(-1.55)
Cost Controls	-0.004	-0.003	-0.003
	(-1.17)	(-1.12)	(-1.14)
R-squared	0.933	0.933	0.933

Note: This table compares alternative versions of the PUD specification for (logarithms of) price as the dependent variable. LCCthreat denotes the fact that Gol is operating within a neighbourhood (measured by distance within 500 and 1,000km from a given airport) but not on the airport itself. LCCpres denotes the actual entry of the LCC on a route. Standard errors are clustered by route-carrier to control for intertemporal correlation, and the corresponding t-statistics are reported within parentheses. Significance at the 10% and 5% significance levels are denoted by * and ** respectively. Fixed effects not reported.

Secretariat of the Treasury of the Finance Ministry (Secretaria do Tesouro Nacional, Ministério da Fazenda - STN). The related literature has resorted to three main classes of instruments:

1. exogenous demand and cost shifters;

2. demand and/or cost characteristics of rivals on a given route or of the same firm on other routes; and

3. lagged or transformed variables.

TABLE 1.9. Incumbents' Response in Aircraft Utilization and Number of Flights

	Average Aircraft Size			Number of Flights		
Variable	**d_{500}**	**$d_{1,000}$**	**Region**	**d_{500}**	**$d_{1,000}$**	**Region**
LCCthreat$_{t^*-4}$	-0.002	0.008	0.041**	-0.070	-0.019	-0.012
	(-0.55)	(1.25)	(2.62)	(-0.92)	(-0.27)	(-0.21)
LCCthreat$_{t^*-3}$	0.002	0.005	-0.054*	0.056	0.042	0.071
	(0.30)	(0.87)	(-1.99)	(0.65)	(0.55)	(0.68)
LCCthreat$_{t^*-2}$	0.007**	0.004	0.055*	0.056	0.031	-0.92
	(2.38)	(1.32)	(1.82)	(0.91)	(0.54)	(0.82)
LCCthreat$_{t^*-1}$	-0.024**	-0.018*	-0.019	0.230	0.224*	-0.013*
	(-2.20)	(-1.71)	(-1.13)	(1.50)	(1.74)	(-0.14)
LCCpres$_{t^*}$	0.043**	0.047**	0.052**	0.259**	0.256**	0.160*
	(2.41)	(2.75)	(3.43)	(2.34)	(2.23)	(1.74)
LCCpres$_{t^*+1}$	0.004	0.004	0.004	0.071	0.053	0.064
	(0.72)	(0.92)	(0.72)	(1.36)	(1.01)	(1.28)
LCCpres$_{t^*+2}$	-0.016*	-0.018*	-0.023**	0.257**	0.249**	0.287**
	(-1.71)	(-1.90)	(-2.42)	(2.12)	(2.07)	(2.27)
LCCpres$_{t^*+3}$	-0.010*	-0.009**	-0.006	-0.088	-0.088	-0.089
	(-1.92)	(-1.84)	(-0.97)	(-1.21)	(-1.22)	(-1.00)
LCCpres$_{t^*+4}$	-0.003	-0.002	-0.002	0.007	0.050	-0.026
	(-0.31)	(-0.23)	(-0.21)	(0.12)	(0.81)	(-0.44)
LCCpres$_{t^*+5}$	-0.038**	-0.040**	-0.041**	0.075	0.049	0.076
	(-2.55)	(-2.63)	(-2.86)	(0.75)	(0.52)	(0.79)
LCCpres×Distance	-0.042**	-0.043**	-0.045**	-0.276**	-0.237**	-0.260**
	(-2.50)	(-2.57)	(-2.76)	(-3.73)	(-2.93)	(-3.27)
Code-share	0.076**	0.076**	0.076**	0.119	0.119	0.118
	(2.45)	(2.45)	(2.45)	(1.17)	(1.17)	(1.17)
Costs Controls	-0.076**	-0.076**	-0.076**			
	(-15.44)	(-15.44)	(-15.39)			
R-squared	0.931	0.931	0.931	0.943	0.943	0.943

Note: This table compares alternative versions of the PUD specification for (logarithms of) aircraft size and the (lower bound on the) number of flights as the dependent variables. LCCthreat denotes the fact that Gol is operating within a neighbourhood (measured by distance within 500 and 1,000km from a given airport) but not on the airport itself. LCCpres denotes the actual entry of the LCC on a route. Standard errors are clustered by route-carrier to control for intertemporal correlation, and the corresponding t-statistics are reported within parentheses. Significance at the 10% and 5% significance levels are denoted by * and ** respectively. Fixed effects not reported.

Berry, Carnall and Spiller (1996) and Borenstein (1989) have used a combination of approaches (1) and (2), notably population at the endpoint cities, whereas a number of papers have used approach (3) – Evans, Froeb and Werden (1995) and Marin (1995) have used lagged variables (which I difficult to use in our short panel) and Evans and Kessides (1993) have used the ranking of the firms according to their presence at a given route.

The instruments I use are measures of market (origin and destination) population and tax revenues, both in levels and in growth rates. While population provides a rough estimate of the potential market of a given route, tax revenues are a proxy for the economic activity of a market. Importantly, these

TABLE 1.10. Incumbents' Response in the Share of Seats Supplied – Weekdays

	Peak Time on Weekdays			Off-peak Time on Weekdays		
Variable	d_{500}	$d_{1,000}$	**Region**	d_{500}	$d_{1,000}$	**Region**
LCCthreat$_{t^*-4}$	-0.007	-0.014**	0.047	0.005	0.012	-0.037
	(-1.32)	(-2.15)	(1.52)	(0.77)	(1.65)	(-1.31)
LCCthreat$_{t^*-3}$	0.011	0.010	-0.020	-0.014	-0.013	-0.005
	(1.03)	(1.10)	(-0.45)	(-1.39)	(-1.55)	(-0.12)
LCCthreat$_{t^*-2}$	0.009**	0.011**	0.028	-0.010**	-0.012**	-0.009
	(2.01)	(2.68)	(0.68)	(-2.21)	(-2.86)	(-0.25)
LCCthreat$_{t^*-1}$	0.000	0.006	0.014	-0.004	-0.013	-0.030
	(0.01)	(0.32)	(0.55)	(-0.19)	(-0.63)	(-1.20)
LCCpres$_{t^*}$	0.056*	0.056*	0.071**	-0.058*	-0.058*	-0.075**
	(1.87)	(1.95)	(2.63)	(-1.65)	(-1.76)	(-2.38)
LCCpres$_{t^*+1}$	0.015	0.014	0.009	-0.016	-0.015	-0.009
	(1.27)	(1.22)	(1.04)	(-1.48)	(-1.38)	(-1.09)
LCCpres$_{t^*+2}$	0.040**	0.041**	0.026*	-0.050**	-0.050**	-0.035**
	(2.13)	(2.18)	(1.71)	(-2.50)	(-2.51)	(-2.17)
LCCpres$_{t^*+3}$	0.005	0.003	0.015*	0.000	0.002	-0.011
	(0.77)	(0.50)	(1.89)	(0.02)	(0.29)	(-1.48)
LCCpres$_{t^*+4}$	0.014	0.016**	0.015*	-0.017*	-0.023**	-0.020**
	(1.44)	(2.02)	(1.99)	(-1.71)	(-2.82)	(-2.51)
LCCpres$_{t^*+5}$	0.003	0.003	0.000	-0.015	-0.014	-0.013
	(0.13)	(0.13)	(0.02)	(-0.60)	(-0.57)	(-0.56)
LCCpres×Distance	-0.046	-0.044	-0.043	0.042	0.038	0.037
	(-1.62)	(-1.56)	(-1.52)	(1.32)	(1.22)	(1.15)
Code-share	-0.003	-0.003	-0.003	-0.008	-0.008	-0.009
	(-0.17)	(-0.17)	(-0.16)	(-0.42)	(-0.42)	(-0.43)
R-squared	0.843	0.843	0.844	0.833	0.833	0.833

Note: This table compares alternative versions of the PUD specification for the share of seats at weekday peak and off-peak times as the dependent variables. LCCthreat denotes the fact that Gol is operating within a neighbourhood (measured by distance within 500 and 1,000km from a given airport) but not on the airport itself. LCCpres denotes the actual entry of the LCC on a route. Standard errors are clustered by route-carrier to control for intertemporal correlation, and the corresponding t-statistics are reported within parentheses. Significance at the 10% and 5% significance levels are denoted by * and ** respectively. Fixed effects not reported.

variables are correlated with the attractiveness of a market and should be exogenous, as desired from a candidate instrument.

1.B.2 Data Sources and Construction of Instruments

To the best of my knowledge, no historical series at the municipal level is recorded at the monthly frequency, and most of them are recorded in intervals of five or ten years, often in connection with the Brazilian Census Survey. Most series are recorded only up to year 2000, including:

1. Among those recorded at ten-year intervals:

TABLE 1.11. Incumbents' Response in the Share of Seats Supplied – Weekends

	Peak Time on Weekends			Off-peak Time on Weekends		
Variable	**d_{500}**	**$d_{1,000}$**	**Region**	**d_{500}**	**$d_{1,000}$**	**Region**
$LCCthreat_{t^*-4}$	0.004	0.004	-0.006	0.004	0.004	-0.042
	(1.13)	(1.26)	(-0.85)	(1.13)	(1.26)	(-1.42)
$LCCthreat_{t^*-3}$	0.003	0.002	0.020**	0.003	0.002	0.015
	(0.81)	(0.72)	(2.00)	(0.81)	(0.72)	(0.34)
$LCCthreat_{t^*-2}$	0.000	0.000	-0.017**	0.000	0.000	-0.026
	(0.12)	(0.00)	(-2.16)	(0.12)	(0.00)	(-0.65)
$LCCthreat_{t^*-1}$	0.001	0.004	0.015**	0.001	0.004	-0.015
	(0.19)	(0.72)	(2.12)	(0.19)	(0.72)	(-0.56)
$LCCpres_{t^*}$	-0.003	-0.002	0.000	-0.003	-0.002	-0.075**
	(-0.30)	(-0.25)	(-0.01)	(-0.30)	(-0.25)	(-2.60)
$LCCpres_{t^*+1}$	0.000	0.000	-0.001	0.000	0.000	-0.010
	(0.02)	(-0.04)	(-0.54)	(0.02)	(-0.04)	(-1.12)
$LCCpres_{t^*+2}$	0.007	0.006	0.006	0.007	0.006	-0.029*
	(1.19)	(1.04)	(1.05)	(1.19)	(1.04)	(-1.80)
$LCCpres_{t^*+3}$	-0.003	-0.004	-0.002	-0.003	-0.004	-0.013*
	(-1.01)	(-1.11)	(-0.70)	(-1.01)	(-1.11)	(-1.72)
$LCCpres_{t^*+4}$	0.001	0.003	0.002	0.001	0.003	-0.018**
	(0.45)	(1.03)	(0.78)	(0.45)	(1.03)	(-2.10)
$LCCpres_{t^*+5}$	0.014**	0.013**	0.015**	0.014**	0.013**	0.002
	(2.20)	(2.15)	(2.39)	(2.20)	(2.15)	(0.07)
LCCpres×Distance	0.008	0.009	0.010	0.008	0.009	0.005
	(1.00)	(1.16)	(1.29)	(1.00)	(1.16)	(1.48)
Code-share	0.008	0.008	0.008	0.008	0.008	-0.001
	(0.69)	(0.70)	(0.70)	(0.69)	(0.70)	(-0.03)
R-squared	0.825	0.826	0.826	0.825	0.825	0.844

Note: This table compares alternative versions of the PUD specification for the share of seats at weekend peak and off-peak times as the dependent variables. LCCthreat denotes the fact that Gol is operating within a neighbourhood (measured by distance within 500 and 1,000km from a given airport) but not on the airport itself. LCCpres denotes the actual entry of the LCC on a route. Standard errors are clustered by route-carrier to control for intertemporal correlation, and the corresponding t-statistics are reported within parentheses. Significance at the 10% and 5% significance levels are denoted by * and ** respectively. Fixed effects not reported.

(a) All but one of population-related measures are recorded in ten-year intervals, the last observation being from year 2000;

(b) Measures of capital related to housing (3 series);

2. Among those recorded at five year intervals:

   (a) Gross product of municipalities and all their components (including industrial production, 15 series in total);

   (b) Employment, wage-related expenses, number of commercial establishments (15 series);

   (c) Income-related measures;

3. Among those recorded annually:

   (a) Human capital;

   (b) Credit and banking-related series (including bank branches, savings, deposits).

Given this restriction, I am left with (i) an estimate of population at the municipality level computed by IBGE, and (ii) a set of public finance series, which are recorded (or estimated) annually and whose last observation is recorded after the end of the route dataset. Following a preliminary analysis accounting for missing values and collinearity among the series, I then selected two series:

1. The population estimated by the IBGE at the municipal level; and

2. The municipal revenues from the value-added tax on services, which is charged on every transaction generated by services, such as private doctors and dentists, but also cleaning and accounting services.

Both variables should be related to the size of the economy of the municipality, thus indicating how attractive the market (route) is. As the series are sampled annually, I interpolate them using a cubic spline to generate monthly data. I then use the interpolated series both in levels and growth rates, since entry could be correlated to both the level and the growth of the population (or tax revenues) of a given market. Finally, *the instruments for entry of a given route are the population (tax revenue) instrument is calculated as the product of the populations (tax revenues) of its two endpoint cities, and the growth rate in population (tax revenue) is calculated from the series in levels.* A similar construction of the population instrument has been previously used in Berry (1992) and Berry, Carnall and Spiller (1996), when studying entry and demand in the airline industry, respectively.

1.B.3 Estimation Strategy and Results

Together with the estimation of the IV regressions, I also conducted diagnostic tests to assess the reliability of the instruments and the corresponding IV estimates. First, I regress the potentially endogenous variable, $LCCpres$, on the instruments – the aim is to check whether the instruments are relevant and valid using the first-stage F-statistic of joint significance of the instruments (the rule of thumb in Staiger and Stock (1997) is that the F-statistic should exceed 10).As commonly happens with clustered standard errors, the covariance matrix of the orthogonality conditions is rank-deficient, so I resort to the Frisch-Waugh-Lovell (FWL, see Frisch and Waugh, 1933 and Lovell, 1963) adapted to instrumental variables and perform the estimation.

For each reaction equation I estimate, the validity of our instruments is assessed using the Hansen J-statistic for overidentification. Provided that the instruments satisfy the relevance and validity conditions, IV estimates can be consistent but inefficient. Therefore, if regressors are exogenous, consistent and more efficient, estimates produced by OLS are preferred.[17]

I start by considering all instruments; in case they are rejected by the J-test, I use subsets of the instruments and examine the adequacy of the specifications checking both the C- and J-tests if applicable (in some cases the models become exactly identified).

For Baseline, Preemptive and PU specifications, estimated using all instruments, (i) the Hansen-Sargan J-statistics have high p-values, suggesting that the null hypothesis is not rejected and the instrument excludability requirement is satisfied ie. the over-identification tests do not reject the null hypothesis that the instruments are independent of the second-stage disturbance terms at the usual significance levels; (ii) the C-tests do not reject the null of exogeneity of the potentially endogenous variable. All in all, the coefficient estimates of the GMM and least squares estimate, as well as their significances, are strikingly similar.

For the variants of the PUD specification, which I estimate using the instruments in levels, the models become exactly identified and I cannot reject the null of exogeneity of the potentially endogenous variables. Again, GMM and least squares estimates, as well as their significances, are quite similar.

The quantity specifications (aircraft size, number of flights and shares of seats supplied) are estimated using the instruments in growth rates and, again, one cannot reject the null of exogeneity – with the benefit of hindsight, this can be attributed, at least in part, to the good job done by the fixed-effects in accounting for unobservables. Table 1.12 summarizes the results of the J- and C-tests for all specifications.

Finally, I also report two sets of results for the sake of illustration. Table 1.13 reports GMM estimates of price reactions using the population and tax revenue instruments for the PUD specification with both distance and cost controls. The magnitudes of the IV estimates are very close to the least squares ones, as is the pattern of significant coefficients, with the exception of $LCCpres_{t^*+1}$ and the *Code-share* estimates. The same holds for the results in Table 1.14, which reports GMM estimates of aircraft utilization using the growth rates of population and tax revenue as instruments, again with both distance and cost controls.

[17] I test for exogeneity using a GMM distance, or C-test. Under conditional homoskedasticity, this is numerically equivalent to a Durbin-Wu-Hausman test statistic.

TABLE 1.12. Tests of Exogeneity and Overidentifying Restrictions

Specification	**Exogeneity** H_0: **Exogeneity**	**Overid. Restrictions** H_0: **Moments not rejected**	**Instruments**	**Estimation Method**
Table 1.3				
Baseline	Not rejected	Not rejected	All	LS
Table 1.4				
Preemptive	Not rejected	Not rejected	All	LS
Table 1.5				
PU	Not rejected	Not rejected	All	LS
Table 1.6				
PUD+Distance	Not rejected	*NA*	Levels	LS
Table 1.7				
PUD+Distance+Dist.-metric	Not rejected	*NA*	Levels	LS
Table 1.8				
PUD+Distance+Controls	Not rejected	*NA*	Levels	LS
Table 1.9				
Aircraft Size	Not rejected	*NA*	Growth rates	LS
Number of Flights	Not rejected	*NA*	Growth rates	LS
Table 1.10				
Share – peak, w'days	Not rejected	*NA*	Growth rates	LS
Share – off-peak, w'days	Not rejected	*NA*	Growth rates	LS
Table 1.11				
Share – peak, w'ends	Not rejected	*NA*	Growth rates	LS
Share – off-peak, w'ends	Not rejected	*NA*	Growth rates	LS

Note: This table summarizes the results of tests of exogeneity (C-tests) and overidentifying restrictions (J-tests) for the price specifications reported in Tables 1.3-1.8, the supply variables of Table 1.9 and the shares in Tables 1.10-1.11 using the standard significance levels of 10%, 5% and 1%. Whenever a model is exactly identified, I report *NA* for the J-test. The instruments used are, for a given route, (i) the product of the populations of the endpoint cities; (ii) the growth rate of (i); (iii) the product of the (service) value-added tax revenues of the endpoint cities and (iv) the growth rate of (iii). Given the outcomes of the exogeneity and overidentifying restrictions tests, the last column reports whether the least squares (LS) or the instrumental variable (IV) estimation method should be used.

TABLE 1.13. PUD Specification Results: IV Estimation of Price Responses – Distance Variable and Cost Controls

Variable	Price		
	d_{500}	$d_{1,000}$	Region
LCCthreat$_{t^*-4}$	0.043**	0.038**	0.033
	(2.69)	(3.25)	(0.78)
LCCthreat$_{t^*-3}$	-0.006	0.003	0.106**
	(-0.24)	(-0.12)	(3.12)
LCCthreat$_{t^*-2}$	0.027**	0.030**	0.025
	(2.41)	(2.93)	(0.73)
LCCthreat$_{t^*-1}$	-0.344**	-0.237**	-0.115**
	(-7.90)	(-5.08)	(-2.07)
LCCpres$_{t^*}$	-0.219**	-0.150**	-0.132**
	(-4.31)	(-3.26)	(-2.46)
LCCpres$_{t^*+1}$	-0.043**	-0.024	-0.039**
	(-2.71)	(-1.40)	(-2.34)
LCCpres$_{t^*+2}$	0.004	-0.003	-0.020
	(0.24)	(-0.12)	(-0.81)
LCCpres$_{t^*+3}$	-0.012	-0.010	-0.001
	(-0.60)	(-0.44)	(-0.02)
LCCpres$_{t^*+4}$	-0.031**	-0.048**	-0.022
	(-2.40)	(-1.99)	(-1.01)
LCCpres$_{t^*+5}$	-0.048*	-0.040	-0.055*
	(-1.66)	(-1.15)	(-1.69)
LCCpres×Distance	0.068**	0.044	0.047
	(1.96)	(1.28)	(1.29)
Code-share	-0.072*	-0.078*	-0.078*
	(-1.74)	(-1.90)	(-1.90)
Cost Controls	0.081**	-0.002	-0.002
	(4.73)	(-0.82)	(-0.83)
J-stat p-value	NA	NA	NA
C-stat p-value	0.42	0.42	0.41

Note: This table compares alternative versions of the PUD specification for (logarithms of) price as the dependent variable. LCCthreat denotes the fact that Gol is operating within a neighbourhood (measured by distance within 500 and 1,000km from a given airport) but not on the airport itself. LCCpres denotes the actual entry of the LCC on a route. Whenever a model is exactly identified, I report *NA* for the J-test. The instruments used are, for each given route, the product of the populations of the endpoint cities and the product of the (service) value-added tax revenues of the endpoint cities. Standard errors are clustered by route-carrier to control for intertemporal correlation, and the corresponding t-statistics are reported within parentheses. Significance at the 10% and 5% significance levels are denoted by * and ** respectively. Fixed effects not reported.

TABLE 1.14. IV Estimation of Responses in Aircraft Utilization

	Average Aircraft Size		
Variable	**d_{500}**	**$d_{1,000}$**	**Region**
LCCthreat$_{t^*-4}$	-0.002	0.008	0.040**
	(-0.51)	(1.27)	(2.75)
LCCthreat$_{t^*-3}$	0.003	0.006	-0.052**
	(0.51)	(1.14)	(-2.12)
LCCthreat$_{t^*-2}$	0.008**	0.006**	0.054**
	(3.18)	(2.07)	(2.03)
LCCthreat$_{t^*-1}$	-0.028**	-0.023**	-0.022
	(-2.72)	(-2.37)	(-1.47)
LCCpres$_{t^*}$	0.040**	0.044**	0.049**
	(2.42)	(2.79)	(3.44)
LCCpres$_{t^*+1}$	0.003	0.003	0.003
	(0.55)	(0.77)	(0.67)
LCCpres$_{t^*+2}$	-0.015*	-0.017*	-0.022**
	(-1.73)	(-1.85)	(-2.53)
LCCpres$_{t^*+3}$	-0.012**	-0.012**	-0.007
	(-2.47)	(-2.47)	(-1.36)
LCCpres$_{t^*+4}$	-0.006	-0.006	-0.005
	(-0.67)	(-0.73)	(-0.51)
LCCpres$_{t^*+5}$	-0.034**	-0.034**	-0.038**
	(-2.49)	(-2.50)	(-2.87)
LCCpres×Distance	-0.039**	-0.041**	-0.043**
	(-2.51)	(-2.65)	(-2.84)
Code-share	0.081**	0.081**	0.081**
	(2.84)	(2.84)	(2.87)
Costs Controls	-0.076**	-0.076**	-0.076**
	(-17.02)	(-16.86)	(-17.43)
J-stat p-value	NA	NA	NA
C-stat p-value	0.17	0.14	0.20

Note: This table compares alternative versions of the PUD specification for (logarithms of) aircraft size and the (lower bound on the) number of flights as the dependent variables. LCCthreat denotes the fact that Gol is operating within a neighbourhood (measured by distance within 500 and 1,000km from a given airport) but not on the airport itself. LCCpres denotes the actual entry of the LCC on a route. Whenever a model is exactly identified, I report *NA* for the J-test. The instruments used are, for each given route, growth rates of the product of the populations of the endpoint cities and the product of the (service) value-added tax revenues of the endpoint cities. Standard errors are clustered by route-carrier to control for intertemporal correlation, and the corresponding t-statistics are reported within parentheses. Significance at the 10% and 5% significance levels are denoted by * and ** respectively. Fixed effects not reported.

References

[1] Aghion, Ph. and P. Bolton (1987) "Contracts as a Barrier to Entry." *American Economic Review*, 77(3), 388-401.

[2] Anderson, S. P. (2008) Product Differentiation, *The New Palgrave Dictionary of Economics*, Eds. Steven N. Durlauf and Lawrence E. Blume, Palgrave Macmillan.

[3] Bain, J. S. (1956) Barriers to New Competition. Cambridge, MA: Harvard University Press, 1956.

[4] Berry, S. T. (1992) "Estimation of a Model of Entry in the Airline Industry." *Econometrica*, 60(4), 889-917.

[5] Berry, S., Carnall, M., Spiller, P. (2006) Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity. in Lee, D. (editor) Advances in Airline Economics. Vol. 1 - Competition Policy and Antritrust. Elsevier. Cambridge, USA, 183-213.

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

[7] Borenstein, S. (1991) "The Dominant-Firm Advantage in Multiproduct Industries: Evidence from the U. S. Airlines." *Quarterly Journal of Economics*, 106(4), 1237-1266.

[8] Borenstein, S. (1992) "The Evolution of U.S. Airline Competition." *Journal of Economic Perspectives*, 6(2), 45-73.

[9] Borenstein, S. (1996) "Repeat-Buyer Programs in Network Industries." in Werner Sichel ed., Networks, Infrastructure, and The New Task for Regulation, University of Michigan Press.

[10] Borenstein, S. and N. L. Rose (1994) "Competition and Price Dispersion in the U.S. Airline Industry." *Journal of Political Economy*, 102(4), 653-683.

[11] Brueckner, J. N. Dyer and P. T. Spiller (1992) "Fare Determination in Airline Hub and Spoke Networks," *RAND Journal of Economics*, 23, 309-333.

[12] Busse, M. (2002), "Firm Financial Condition and Airline Price Wars", *RAND Journal of Economics*, 33 (2), 298-318.

[13] Cairns, R. D. and J. W. Galbraith (1990). "Artificial Compatibility, Barriers to Entry and Frequent Flyer Programs," *Canadian Journal of Economics*, 23(4), 807-816.

[14] Caves, D. W., L R. Christensen, M. W. Tretheway (1984) "Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ," *RAND Journal of Economics*, Vol. 15, No. 4. 471-489.

[15] DAC (2002) "Average Yield of Monitored Airport Pairs".

[16] DAC (2002) "Anuario Estatistico do Transporte Aereo".

[17] DAC (2003) "Average Yield of Monitored Airport Pairs".

[18] DAC (2003) "Anuario Estatistico do Transporte Aereo".

[19] DAC (2004) "Average Yield of Monitored Airport Pairs".

[20] DAC (2004) "Anuario Estatistico do Transporte Aereo".

[21] Davis, P. (2006) "Spatial competition in retail markets: movie theaters", *RAND Journal of Economics* 37 (4) , 964–982.

[22] Dixit, A. (1980) The Role of Investment in Entry-Deterrence. *Economic Journal* 90, 95-106.

[23] Donnelly, S., James, A., and Binnion, C. (2004) Bmi's Response to the Changing European Airline Marketplace. *Journal of Revenue and Pricing Management* 3, 10-17.

[24] Dresner, M., Lin, J., & Windle, R. (1996) The Impact of Low-Cost Carriers on Airport and Route Competition. *Journal of Transport Economics and Policy* 30, 309-328.

[25] Evangelho, F. (2002) A Flexibilizacao da Aviacao Comercial Brasileira. Cadernos EBAPE, Getulio Vargas Foundation, Rio de Janeiro.

[26] Evans, W., Froeb, L. and Werden, G. (1993) Endogeneity in the Concentration-Price Relationship: Causes, Consequences, and Cures. Journal of Industrial Economics, 41, 431-438.

[27] Evans, W. & Kessides, I. (1993) Localized Market Power in the U.S. Airline Industry. *Review of Economics and Statistics* 75, 66-75.

[28] Farrell, J. and P. Klemperer (2004) "Coordination and Lock-In: Competition with Switching Costs and Network Effects." forthcoming in Michael Armstrong and Robert Porter, eds., *Handbook of Industrial Organization*, Vol. 3, Amsterdam: North-Holland.

[29] Frisch, R., and F. V. Waugh (1933) Partial time regressions as compared with individual trends. *Econometrica* 1 (4): 387–401.

[30] Goolsbee, A. and Syverson, C. (2008) How Do Incumbents Respond to the Threat of Entry? Evidence from Major Airlines. forthcoming *Quarterly Journal of Economics.*

[31] Hendricks, K., M. Piccione, and G. Tan (1997) "Entry and Exit in Hub-Spoke Networks." *RAND Journal of Economics*, 28(2), 291-303.

[32] Holmes, T. (2008) "The Diffusion of Wal-Mart and Economies of Density," NBER Working Paper #13783, February 2008.

[33] Hurdle, G. J., R. L. Johnson, A. S. Joskow, G. J. Werden, M. A. Williams (1989) "Concentration, Potential Entry, and Performance in the Airline Industry." *Journal of Industrial Economics*, 38(2), 119-139.

[34] Klemperer, Paul. "Entry Deterrence in Markets with Consumer Switching Costs." The *Economic Journal*, 97(Supplement: Conference Papers), 1987, 99-117.

[35] Kreps, D. and Wilson, R. (1982) Reputation and Imperfect Information. *Journal of Economic Theory* 27, 253-279.

[36] Lancaster, K. J. (1966) "A New Approach to Consumer Theory". *Journal of Political Economy*, 74, 132-157.

[37] Lovell, M. (1963) Seasonal adjustment of economic time series. *Journal of the American Statistical Association* 58: 993–1010.

[38] Marin, P. (1995) Competition in European Aviation: Pricing Policy and Market Structure. Journal of Industrial Economics, 43, 141-159.

[39] Mayer, C. and T. Sinai (2003) "Network Effects, Congestion Externalities, and Air Traffic Delays: Or Why All Delays Are Not Evil." *American Economic Review*, 93(4), 1194-1215.

[40] Milgrom, P. and Roberts, J. (1982) "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis," *Econometrica*, Econometric Society, vol. 50(2), pages 443-59, March.

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

[42] Morrison, S. and C. Winston (1989) "Enhancing the Performance of the Deregulated Air Transportation System." *Brookings Papers on Economic Activity*, Microeconomics, 1, 61-112.

[43] Oliveira, A. V. M. (2007) "Acesso a Recursos Essenciais e Poder de Mercado: Estudo de Caso da Concessão de Slots em Aeroportos no Brasil", NECTAR–ITA monograph.

[44] Peteraf, M. A. and R. Reed (1994) "Pricing and Performance in Monopoly Airline Markets." *Journal of Law and Economics*, 37(1), 193-213.

[45] Pinkse, J. and M. Slade (2004) Mergers, Brand Competition, and the Price of a Pint. *European Economic Review* 48, 617-643.

[46] Pinkse, J., Slade, M., & Brett, C. (2002) Spatial Price Competition: A Semiparametric Approach. *Econometrica* 70, 1111-1155.

[47] Reiss, P. C. and P. T. Spiller (1989) "Competition and Entry in Small Airline Markets." *Journal of Law and Economics*, 32, S179-S202.

[48] Richard, O. (2003) Flight frequency and mergers in airline markets. *International Journal of Industrial Organization* 21, 907-922.

[49] Slade, M. (2004) Market Power and Joint Dominance in UK Brewing. *Journal of Industrial Economics* 52, 133-163.

[50] Souza, C. A. (2006), Senso de Oportunidade da Empresa Gol Linhas Aéreas no Mercado de Aviação Civil do Brasil. Monograph in Economics, Federal University of Santa Catarina, Brazil.

[51] Staiger, D., and J. H. Stock (1997) Instrumental variables regression with weak instruments. *Econometrica* 65(3): 557–86.

[52] Thomas, L. A. (1999) Incumbent Firms' Response to Entry: Price, Advertising, and New Product Introduction. *International Journal of Industrial Organization* 17, 527-555.

[53] Whinston, M. and Collins, S. (1992) Entry and Competitive Structure in Deregulated Airline Markets: An Event Study Analysis of People Express. *RAND Journal of Economics* 23, 445-462.

[54] Windle, R. & Dresner, M. (1999) Competitive Responses to Low Cost Carrier Entry. *Transportation Research - Part E* 35, 59-75.

[55] Yamawaki, H. (2002) Price Reactions to New Competition: A Study of US Luxury Car Market, 1986-1997. *International Journal of Industrial Organization* 20, 19-39.

2

Term Structure Modelling with Observable State Variables

2.1 Introduction

The interaction between the term structure of interest rates and macroeconomic variables has been extensively explored in a number of papers in the last decade or so, thanks in part to the fact that 'yields-only' models based on no-arbitrage were found to do well in fitting the cross-section of yields at a particular point in time (de Jong, 2000; Dai and Singleton, 2000), but poorly in describing the dynamics of the yield curve (Duffee, 2002; Brousseau, 2002).[1] This paper proposes and implements a parsimonious three-factor model of the term structure whose dynamics is driven uniquely by observable state variables, as opposed to latent variables.[2] It builds upon a three-factor model describing the term structure behaviour first proposed in Nelson and Siegel (1987) and recetly reinterpreted by Diebold and Li (2006, DL) as a dynamic latent factor model.

The literature can be divided into three main streams, according to their methods and purposes. The first strand of the literature relies on the optimizing behaviour of economic agents, which can thus be cast whithin the dynamic stochastic general equilibrium (DSGE) framework.[3] Currently, this

[1]Contributions to the literature in the last decade include Fuhrer and Moore (1995), Rudebusch (1995, 2002), Evans and Marshall (1996), Fuhrer (1996), Dewachter and Lyrio (2006), Hördahl, Tristani and Vestin (2006), Wu (2002), Piazzesi (2005), Ang and Piazzesi (2003), and Bikbov and Chernov (2005).

[2]In the text I usually refer to state variables for the sake of generality, but in the literature the set of variables that have been mostly used are macroeconomic variables.

[3]Early contributions to the literature of equilibrium pricing include Cox, Ingersoll and Ross (1985), Campbell (1986) and Dunn and Singleton (1986). The recursive preferences of Epstein and Zin (1989) and Weil (1989) have been used extensively , as in Campbell (1993, 1996, 1999), Duffie, Schroder and Skiadas (1997) and more recently, Schneider and Piazzesi (2006), while Wachter (2006) adapts a consumption habit utility specification *à la* Campbell and Cochrane (1999).

approach still needs unreasonable assumptions about risk aversion and the elasticity of intertemporal substitution (EIS) to deliver satisfactory results.[4]

A second strand of the literature adopts only the very basic structure of the DSGE approach, namely the absence of arbitrage opportunities, to study the term structure of interest rates, usually imposing no-arbitrage when estimating a Vector Autoregression of yields.[5] [6] The main (empirical) lesson from this stream of literature is that one needs a combination of (observed) state variables plus an unobserved factor to explain the dynamcs of the term structure of interest rates.

The route I follow in this paper belongs to a third strand of the literature, going back to Nelson and Siegel (1987) and Diebold and Li (2006) in that it decomposes the yield curve into three factors, namely level, slope and curvature and relates them to observable state variables – macroeconomic variables in particular – to forecast the term structure of interest rates. The intuition of the modelling strategy I adopt goes as follows. If the term structure moves as a result of changes in the economic fundamentals – here represented by a set of state variables – the term structure factors (and, by consequence, the term structure dynamics) should be somehow linked to these state variables. In this paper, I make this link explicit, so that the movements of the term structure are completely exerted by the underlying state variables.

The approach I propose contributes to the literature from both the theoretical and the empirical viewpoints. First, the replacement of latent factors with observable state variables as the only drivers of the term structure factors allows comparing alternative views on the way state variables – macroeconomic variables, in particular – influence the yield curve dynamics.[7] Besides telling more about the economic fundamentals than latent variables, the use of observable variables might also provide guidance to the construction of theoretical models of the term structure dynamics. Additionally, the method enables testing hypotheses of economic interest – as a result, instead of pre-specifying the

[4] The risk aversion parameter typically needs to be substantially high, in what Mehra and Prescott (1985) called the equity premium puzzle. Recently, Bansal and Yaron (2004) were albe to match the equity premium using recursive preferences and a plausible value for the risk aversion parameter, but this in turn crucially relies on the high persistence of consumption shocks. See also Schneider and Piazzesi (2006) for the use of recursive preferences and Wachter (2006) for the combined use of a consumption habit utility specification based on Campbell and Cochrane (1999) with a model for the term structure of interests rates. As for the EIS, even its magnitude is subject to controversy, with Bansal, Kiku and Yaron (2007) and Attanasio and Vissing-Jorgensen (2007) estimating it to be above one and Hall (1988) and Campbell (1999) documenting it to be close to zero.

[5] See, for instance, Piazzesi (2005), Ang and Piazzesi (2003), and Bikbov and Chernov (2005).

[6] In practice, one needs to, first, specify the instantaneous interest rate and the prices of risk for the factors assumed to affect the yield curveas functions of state variables (such as economic activity and inflation); second, focus on the asset pricing implications of the structure imposed. Although not usually explored (or even made explicit) in the literature, for each choice of functional form above, in equilibrium there should be an economyconsistent with these choices.

[7] In a number of recent related studies, observable and latent factors coexist: Ang and Bekaert (2004) use one observable (inflation) and two latent factors; Rudebusch and Wu (2004) use two observable (GDP growth and inflation) and two latent; Hördahl, Tristani and Vestin (2006) use three observable (the short rate, GDP growth, and inflation) and one latent one; Ang, Dong, and Piazzesi (2005) use two observable (inflation and GDP growth) and one latent factor. Exceptions include Ang, Piazzesi, and Wei (2004), which use the short rate, the term spread, and GDP growth as their state variables, and Bekaert, Cho, and Moreno (2004), which uses the short rate, the output gap, and inflation. As opposed to what I present in the empirical exercise, these papers do not conduct model comparisons, pre-specifying the state variables they use – this important aspect of the paper is discussed in detail in the sequel.

drivers of the yield curve dynamics, I compare alternative models and select the best among them. Moreover, the explicit link between term structure factors and observable state variables enables policy experiments to be performed. As a result, one can forecast the term structure by using forecasted variables, or perform stress testing of the term structure using scenarios constructed using the state variables. This feature is especially useful to bankers, who are interested in forecasting bond prices and might have a better idea of the expected state of the economy than the expected state of the yield curve. This feature is also of value to financial authorities, as a tool to assess financial stability.[8]

Second, the method is robust to curse of dimensionality problems commonly appearing in traditional models. The curse of dimensionality imposes constraints on the number of yields one can use and, in particular, results in poor measures of the term structure curvature.[9] Here, instead, the dimension of the parameter vector does not increase with the number of yields under study, just with the number of state variables explaining them, very much in the spirit of linear regression, where one loses degrees of freedom by including additional covariates, not more observations.

Third, the identification strategy comes out in a natural way. Essentially, the baseline model needs the state variables driving the term structure to be predetermined with respect to yields. When I incorporate a Taylor rule into the model, the identifying assumption made is also standard, requiring the state variables to be predetermined with respect to the monetary policy instrument.

Fourth, I conduct in- and out-of-sample studies using US data. The in-sample study uses a thorough set of macroeconomic variables to compare alternative specifications of the term structure dynamics and suggests two models which I then use in the out-of-sample exercise: a parsimonious model where the level, slope and curvature factors of the term structure are driven by, respectively, measures of inflation growth, monetary policy, and economic activity[10], and a richer specification where the level is driven by measures of inflation growth and economic activity, the slope by monetary policy and economic activity, and the curvature by fiscal policy growth.[11] The out-of-sample study shows that both specifications consistently outperform the (latent-variable) benchmark model in the study of the yield curve behaviour during the five NBER-dated recessions which occurred in the last three decades. Recessions are of interest not only for being bad states against which economic agents are willing to

[8] Stress testing has become crucial in the risk management toolbox of financial institutions. It is defined in BIS (2000) as "a generic term describing various techniques used by financial institutions to gauge their potential vulnerability to exceptional but plausible events". Due to the fact that standard Value at Risk (VaR) models have been found to be of limited use in measuring exposures to extreme events, stress testing has been incorporated into the risk management routine of financial institutions, and has even been stressed during the ongoing Basel II process as a useful tool in assessing banks' internal models.

[9] One needs at least three yields for the curvature to be defined, but by relying on only five yields, as is often done in the literature, one is unlikely to obtain accurate measures of this factor.

[10] These are, respectively, the Consumer Price Index growth rate, the Fed Funds rate, and the Unemployment Rate.

[11] The level is driven by the Consumer Price Index growth rate and the Unemployment Rate, the slope by the Fed Funds rate and the Unemployment Rate, and the curvature by the growth rate of the ratio between Government deficit and Industrial Production – the proxy measure of GDP at the monthly frequency.

insure, but also for being periods which tend to be preceded by the inversion of the term structure of interest rates, a feature usually difficult to be quickly captured – if at all – by term structure models.

The paper is organized as follows. Section 2 briefly reviews term structure estimation methods and recent developments of the Nelson-Siegel approach. Section 3 presents the model and discusses its identification and implementation. Section 4 presents simulation results, and Section 5 performs an empirical exercise using US data. The Appendix contains an empirical exercise using CRSP interest rate data as a robustness check and discusses strategies for incorporating spatial modelling into the model.

2.2 Yield Curve Estimation

2.2.1 Static Methods

When analyzing the evolution of the yield curve over time, one striking feature is the variety of shapes it can have. These vary from flat ones, where longer term rates are roughly the same as shorter ones, to upward-sloping ones, where longer term rates are higher, but also include 'hump-shaped', inverted, 'spoon-shaped' ones etc. As a result, yield curve fitting methods are expected to be flexible enough to match the different shapes the yield curve can have.

A number of approaches can be used to modelling the term structure of interest rates. First, one may consider models that make explicit assumptions about the evolution of state variables and use either equilibrium or arbitrage methods, which corresponds to modelling dynamic yield curves. According to this class of models, the evolution of the yield curve is modelled as depending linearly on a small number of (arbitrarily chosen) factors. Since in most of the cases the underlying state variable is the short term interest rate, they are frequently labeled as 'short-rate models'. The landmarks of this approach are the papers by Vasicek (1977) and Cox, Ingersoll, and Ross (1985, CIR), both of which use the short rate as the only underlying factor. Subsequent extensions to multi-factor models include the two-factor model of Longstaff and Schwartz (1992), and the three-factor one of Balduzzi, Das, Foresi and Sundaram (1996, BDFS). When it comes to fitting real data, one-factor models perform poorly: the yield curve corresponding to the Vasicek model does not allow a large range of shapes, whereas the ones corresponding to CIR and extensions allowing time-varying parameters such as Hull and White (1990) tend to evolve unrealistically over time. In what regards multi-factor models, there is an understanding that at least three factors are needed to generate a wide variety of yield curve shapes, although even so the fit close to the long end tends to be poor. Moreover, choosing the state variables involves both a certain degree of arbitrariness and a bit of art - direct factors may include the short rate, spot rates of various maturities, forward rates, swap rates, whereas indirect ones may include the short rate volatility, the mean short rate, the latter two rising issues such as the choice of the sample period involved in their calculations. Further, multi-factor models (such as the BDFS) usually lack of explicit formulae and are of difficult calibration to market prices.

Alternatively, one can smooth data obtained from asset prices to describe the static yield curve, usually without taking a view on the factors driving it. This corresponds to fitting, the yield curve as a whole. The analysis starts from information on asset prices, from which one extracts the corresponding yields. As there are only a few maturities available for which there are observations on prices (and, thus, yields), it is interesting to somehow 'connect' those points in order to evaluate instruments with maturities different from those of the yields one has already extracted, usually imposing some degree of smoothness. Among the estimation methods most widely used, there are (regression and smoothing) spline techniques, kernel methods, but also parametric classes of curves, broadly known as the Nelson-Siegel family of curves. Among regression splines one can find several sub-varieties - McCulloch (1971, 1975) used quadratic and cubic splines, Schaefer (1981) employed Bernstein polynomials, whereas Vasicek and Fong (1982) adopted exponential splines. Regression splines have some inconveniences though. One has to take into account the arbitrariness involved, first, in the choice of knot points, second, in the choice of basis functions. Thirdly, splines may oscillate too much and are too sensitive to modelling parameters, with the consequence of fitting poorly at too long and too short maturities. Fourthly, since splines are polynomials, they imply a discount function which diverges as maturity increases rather than converging to zero as required by theory - as a result, implied forward rates also diverge rather than converging to any fixed limit. Fifthly, there is no simple way to ensure that the discount function always declines with maturity i.e. that all forward rates are positive. Although exponential splines are appealing in theory, it is not clear that they perform better than standard splines in practice (Shea, 1985). As for smoothing splines (Fisher, Nychka and Zervos, 1995), they reduce the amount of curvature as one may well desire when uncomfortable with regression splines, but at the expense of a worse fit to the yield curve.

The class of curves first proposed in Nelson and Siegel (1987) is parsimonious and does well in capturing the overall shape of the yield curve, being popular among practitioners and central banks alike (BIS, 2000).[12] Their objective is to describe the yield curve, not being consistent with the absence of arbitrage opportunities. For a sample of N bonds measured at a given point in time, the yield curve

[12] Although one could argue that a three-factor model could be too much of a simplification, Diebold, Rudebusch and Aruoba (2005) find no evidence that extensions of Nelson-Siegel using four or five factors would do better, which is consistent with previous findings of Dahlquist and Svensson (1994)

as a function of time to maturity τ_i is written as

$$y(\tau_i) = \beta_1 + \beta_2 \left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i}\right) + \beta_3 \left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}\right) + u(\tau_i), i = 1, ..., N$$

providing a parsimonious representation of the term structure which is consistent with a well-behaved discount function i.e. continuous, positive and decreasing in τ, taking value 1 when $\tau = 0$ and approaching β_1 as τ grows large.[13] As we justify below, the parameters $\beta_1, \beta_2, \beta_3$ can be interpreted as, respectively, the level, (the negative of the) slope, and curvature components, whereas the parameter λ_t controls the exponential decay of the yield curve: small values produce slow decay and can better fit the curve at long maturities, while large values generate a fast decay and can better fit the curve at short maturities. Moreover, λ also determines where the loading on β_3 achieves its maximum. The loading on β_1 is a constant, implying that an increase in this factor increases all yields equally, which results in a change in the level of the yield curve. The loading on β_2 is a function that starts at 1 but decays monotonically to zero, implying that an increase in β_2 increases short yields more than long yields, resulting in a change in the slope of the yield curve. As for β_3, this is related to the curvature of the term structure, as an increase in β_3 will have little effect on very short or very long yields, but will increase medium-term yields, thus resulting in an increase of curvature of the yield curve. As first described by Diebold and Li (2006), this representation can be related to a dynamic three-factor model of, respectively, level, slope, and curvature, which I describe in the following.

2.2.2 Nelson-Siegel and Beyond

The framework recently proposed in Diebold and Li (2006), and also used in Diebold, Rudebusch, and Aruoba (2006, DRA) reinterprets the Nelson and Siegel (1987) framework as a dynamic latent-factor model.[14] Following Diebold and Li (2006), for every time period t, the yield curve is a function of time-to-maturity τ (or, rather, a combination of exponential functions thereof) and time-varying parameters interpreted as the level, slope, and curvature factors,

$$y_t(\tau_i) = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda_t\tau_i}}{\lambda_t\tau_i}\right) + \beta_{3t} \left(\frac{1 - e^{-\lambda_t\tau_i}}{\lambda_t\tau_i} - e^{-\lambda_t\tau_i}\right) + u_t(\tau_i); i = 1, ..., N, t = 1, ..., T$$

Estimation could in principle be carried out using Nonlinear Least Squares (NLLS) although the usual practice since Nelson and Siegel (1987) – and also followed in Diebold and Li (2006) – has been to fix λ_t to a constant value, compute the factor loadings (regressors), and then use OLS to estimate

[13] Equivalently, it guarantees positive forward rates at all horizons.

[14] These methods still allow arbitrage opportunities, but the authors argue that a market as liquid as the one of US government bonds should not allow much scope for arbitrage opportunities.

$\{\beta_t\}$. The parameter λ_t determines the maturity τ^* at which the loading on the curvature factor achieves its maximum (usually between 2 and 3 years), and Diebold and Li (2006) simply pick a λ_t such that this maximum is achieved at the midpoint between these maturities – 30 months – and set $\lambda^* = \lambda_t = 0.0609$. After computing the sequence $\{\beta_t\}$ of factors and the pricing errors, they model the factors as a univariate AR(1) models and compare the forecasting power of the model out-of-sample with a number of alternatives, with reasonable performance, especially given the simplicity of the model.

The above framework is intuitive and easy to implement, but is still based on latent variables – despite the consensus that changes in the yield curve are exerted by changes in macroeconomic conditions (or, more generally, changes in state variables), the factors in the DL framework remain latent, whereas in DRA latent and observable factors (pre-specified by the researchers) coexist. As a result, it offers no room for comparing alternative views on the main drivers of the term structure dynamics. Moreover, the empirical implementations restrict the dynamics of the time-varying factors in ways that, despite their reasonability, are difficult to be either verified or refuted: whereas Diebold and Li (2006) model the parameters $\{\beta_p\}_{p=1}^3$ as univariate AR(1) processes, Diebold, Rudebusch, and Aruoba (2006), generalize it to a first-order vector autoregression. In any case, the estimation of the stochastic processes driving the level, slope, and curvature factors does not acount for the measurement error coming from the fact that $\{\beta_p\}_{p=1}^3$ are estimated rather than observed, so that any asymptotic statements are likely to be misleading.

2.3 Term Structure Modelling

This section proposes a term structure modelling approach building upon Nelson-Siegel and its reinterpretation by Diebold and Li (2006). The main contrast with respect to the DL model is that here the term structure dynamics is solely driven by the dynamics of observable state variables, as opposed to latent factors. The intuition behind this idea is that if the yield curve moves as a result of changes in relevant state variables, the factors should be somehow linked to these state variables. As a result, one can now, for instance, compare alternative hypotheses on the variables driving the term structure factors and state that level, slope and curvature factors are driven by, say, measures of economic activity, inflation, and monetary policy instrument, respectively.

2.3.1 A Model with State Variables

The DL model writes the yields at time t as a function of the maturity vector $\boldsymbol{\tau}$,

$$y_t(\boldsymbol{\tau}) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_t\boldsymbol{\tau}}}{\lambda_t\boldsymbol{\tau}}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_t\boldsymbol{\tau}}}{\lambda_t\boldsymbol{\tau}} - e^{-\lambda_t\boldsymbol{\tau}}\right) + u_t(\boldsymbol{\tau})$$

further assuming that λ_t is constant over time and setting it to a pre-specified value. Estimation of the parameter vector β_t for every period is carried out using linear least squares, which assumes that the error term does not depend on maturity. This results in a time series of the parameter vector $\{\beta_t\}$ whose dynamics is approximated by univariate first-order autoregressive processes for each of its components. The reasons for fixing λ_t to a pre-specified value are, according to the authors, its lack of straightforward economic intuition and the gains from the use of simple linear techniques when estimating the model. In fact, given the small cross-sectional dimension of the yields dataset they use, fitting a nonlinear model can be a very challenging exercise. The main take-away point is, however, the latent-variable character of the DL model.

The main point of departure from DL in this paper is the link between the dynamics of the latent variables to the one of observable state variables *predetermined* relative to $y_t(\tau)$, which I denote by M_{t-}.[15] This is done by decomposing the parameter vector $\theta_t := (\beta_t', \lambda_t)'$ as a sum of two components: the first, $\overline{\theta} := (\overline{\beta}', \overline{\lambda})'$, being a mean component, and the second being the combination of the state variables M_{t-} and parameters $\underline{\theta} := (\sigma_\beta', \sigma_\lambda')'$ measuring their impact on the latent variables. Thus,

$$\theta_t := \begin{bmatrix} \beta_t \\ \lambda_t \end{bmatrix} = \begin{bmatrix} \overline{\beta} \\ \overline{\lambda} \end{bmatrix} + M_{t-}\begin{bmatrix} \sigma_\beta \\ \sigma_\lambda \end{bmatrix}$$

This specification can be seen as a random-coefficients one for the term structure factors; the factors vary over time given realizations of the state variables. Decomposing the time-varying parameters as above assumes the relation between state variables and term structure factors is deterministic, which might obviously lead to biased results if this assumption does not hold in practice.

The full model reads

$$\begin{aligned} y_t(\tau) &= X_t(\lambda_t)\beta_t + u_t(\tau) \\ \begin{bmatrix} \beta_t \\ \lambda_t \end{bmatrix} &= \begin{bmatrix} \overline{\beta} \\ \overline{\lambda} \end{bmatrix} + M_{t-}\begin{bmatrix} \sigma_\beta \\ \sigma_\lambda \end{bmatrix} \end{aligned}$$

[15]In what follows, given two variables A and B, I write A_{t-} if A is predetermined with respect to B within period t.

where $y_t(\tau)$ is the column vector of yields observed at date t, the nonlinearity of the model comes from the estimation of λ_t in the $N \times 3$ matrix of factor loadings at period t,

$$X_t(\lambda_t) = \begin{bmatrix} 1 & \frac{1-\exp(-\lambda_t\tau_1)}{\lambda_t\tau_1} & \frac{1-\exp(-\lambda_t\tau_1)}{\lambda_t\tau_1} - \exp(-\lambda_t\tau_1) \\ 1 & \frac{1-\exp(-\lambda_t\tau_2)}{\lambda_t\tau_2} & \frac{1-\exp(-\lambda_t\tau_2)}{\lambda_t\tau_2} - \exp(-\lambda_t\tau_2) \\ \ldots & \ldots & \ldots \\ 1 & \frac{1-\exp(-\lambda_t\tau_N)}{\lambda_t\tau_N} & \frac{1-\exp(-\lambda_t\tau_N)}{\lambda_t\tau_N} - \exp(-\lambda_t\tau_N) \end{bmatrix}$$

and the error term is a martingale difference sequence with respect to current and past covariate information and uncorrelated in the maturity domain i.e. $E[u_t(\tau)u_t(\tau)'] = \sigma^2\mathbf{I}$.[16]

This model is more costly to be estimated from the numerical point of view, but this cost is offset by having the dynamics of $\{\beta_t\}$ driven by state variables. Moreover, there are also gains from modelling the dynamics of $\{\lambda_t\}$, apart from a pure generality argument. If the parameter β_{3t} governs the intensity of the curvature of the yield curve, the parameter λ_t governs the locus of its 'tilting point' or, alternatively, where the loading associated to the factor β_{3t} attains its maximum, thus making it unnatural to be disconnected to the analysis of the term structure curvature.

In what regards identification, the argument goes as follows. Data is observed at the monthly frequency, but recorded at different moments within a given month – the state variables M_{t-} are observed at the beginning of each month, whereas the yields are observed at the end of the corresponding month.[17] As a result, the state variables M_{t-} are predetermined with respect to the yields.

Important features of the method are its robustness to errors in variables, its parsimony, and its robustness to the curse of dimensionality. First, as opposed to DL, where (i) the extraction of the $\{\beta_t\}$ sequence of parameters relies solely on the cross-sectional dimension of the data; (ii) the estimation of the AR(1) models for factor dynamics relies solely on the time series dimension of the data; and (iii) the estimation of the factor dynamics uses estimates of $\{\beta_t\}$ as if they were data, incurring in measurement error problems, estimation here relies on both the time series and the cross-sectional dimension of the data and is done in one step. Thus, by working on both T and N, the asymptotic results tend to be much more accurate. Moreover, the fact that the estimation is done simultaneously avoids the measurement error coming from the fact that $\{\beta_p\}_{p=1}^3$ are estimated rather than observed in DL.

[16] Although restrictive, alternative estimation strategies allowing for spatial dependence ie. dependence across yields of close enough maturities were also tried, but without much success.

[17] In the empirical exercise using US data, the state variables are observed at the beginning of each month, whereas the yields are taken from the last working day of each month.

Second, parsimony results from the fact that the ultimate parameters of interest are time-invariant. Third, as opposed to traditional VAR models such as in Evans and Marshall (2002), the number of parameters to be estimated does not increase with the number of yields, even after imposing zero restrictions that imply exogeneity of macro variables with respect to yields.

Finally, when it comes to simulate the movements of the term structure – or out-of-sample forecasting, more generally – one just needs to plug-in updated (or forecasted) values of M_{t-} and compute the resulting yields forecasts; alternative models, such as DRA, which contain both latent and observable factors, would need to rely on extra assumptions on the latent part to do so.

2.3.2 Implementation

In this section I discuss the implementation of the model

$$\begin{aligned} y_t(\tau) &= X_t(\lambda_t)\beta_t + u_t(\tau), t = 1,..,T \\ \begin{bmatrix} \beta_t \\ \lambda_t \end{bmatrix} &= \begin{bmatrix} \overline{\beta} \\ \overline{\lambda} \end{bmatrix} + M_{t-} \begin{bmatrix} \sigma_\beta \\ \sigma_\lambda \end{bmatrix} \end{aligned}$$

where $y_t(\tau)$ is the vector of yields observed at date t, and $u_t(.)$ is the error term, both of dimension $N \times 1$, $X_t(.)$ is $N \times 3$, β_t and $\overline{\beta}$ are 3×1, λ_t and $\overline{\lambda}$ are scalars, σ_β and σ_λ are, respectively, $k_\beta \times 1$ and $k_\lambda \times 1$, and $M_{t-} = \begin{bmatrix} M_{\beta t-} & 0_{3\times k_\lambda} \\ 0_{1\times k_\beta} & M_{\lambda t-} \end{bmatrix}$ is $4 \times k(= k_\beta + k_\lambda)$.[18]

The model consists of N yield observations for each one of the T periods, k state variables per period, and $k+4$ parameters to be estimated, regardless of the number of yields or time periods in the sample – the dimension of the parameter vector grows only with the number of state variables in the model (say, at most three per factor, so that most likely $k \leq 12$). Since $\lambda_t = \overline{\lambda} + M_{\lambda t-}\sigma_\lambda$, one can write $X_t(\lambda_t) = X_t(\overline{\lambda}, \sigma_\lambda)$ but should bear in mind that both τ and $M_{\lambda_{t-}}$ are also arguments of $X_t(.)$ but are omitted for convenience.

The assumption that the error term $u_t(\tau)$ is a martingale difference sequence with respect to current and past covariate information implies conditional moment restrictions of the form $E[u_t(\tau)|W_t] = 0$, where W_t is a vector of instruments including current and past covariate information. In particular, for every period t, one can use unconditional moments of the form $E[W_t' u_t(\tau)] = 0$, whose sample

[18]In particular, $M_{t-} = diag\{m_t^{\beta_1}, m_t^{\beta_2}, m_t^{\beta_3}, m_t^{\lambda}\}$. In the general case, $M_{t-} = \begin{bmatrix} M_{\beta t-}, & M_{\lambda t-} \end{bmatrix}$ is $4 \times k(= k_\beta + k_\lambda)$.

counterpart is

$$
\begin{aligned}
0 &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} u_{it}w_{it} \\
&= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(y_t(\tau_i) - X_t(\tau_i,\theta_\lambda)\overline{\beta} - X_t(\tau_i,\theta_\lambda)M_{\beta t-\sigma_\beta}\right) w_{it}
\end{aligned}
$$

with $\theta := (\theta'_\beta;\theta'_\lambda)' = (\overline{\beta}', \sigma'_\beta; \overline{\lambda}, \sigma'_\lambda)'$ and where it should be noted that $M_{\lambda t-}$ appears inside $X_t(.,.)$ and is omitted for convenience.

Before defining the estimation problem, stack the yields by period to form the $NT \times 1$ vector $y = [y_1(\tau)', y_2(\tau)', ..., y_T(\tau)']'$, the $X_t(.)$ and $M_{\beta t-}$ matrices to form the $NT \times 3$ matrix $X(\theta_\lambda) = [X_1(\theta_\lambda)', X_2(\theta_\lambda)', ..., X_T(\theta_\lambda)']'$ and the $NT \times k$ matrix

$XM(\theta_\lambda) = [(X_1(\theta_\lambda)M_{\beta 1-})', (X_2(\theta_\lambda)M_{\beta 2-})', ..., (X_T(\theta_\lambda)M_{\beta T-})']'$ of dimension , define

$Z_t(\theta_\lambda) = [X_t(\theta_\lambda), X_t(\theta_\lambda)M_{\beta t-}]$ and its stacked version which is of dimension $NT \times (3+k)$, $Z(\theta_\lambda) = [Z_1(\theta_\lambda)', Z_2(\theta_\lambda)', ..., Z_T(\theta_\lambda)']'$, and let $W = [W'_1, ..., W'_T]$ be an instrument matrix of dimension $NT \times r(\geq k+4)$. I use a Generalized Method of Moments estimator $\widehat{\theta}$ of θ, which is such that the quadratic distance between $G_{NT}(\theta) = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} u_{it}w_{it}$ from zero is minimized:

$$
\begin{aligned}
\widehat{\theta} &= \arg\min_{\theta\in\Theta} [G_{NT}(\theta)]' A_{NT} [G_{NT}(\theta)] \\
&= \arg\min_{\theta\in\Theta} \left[\frac{1}{NT}W'u\right]' A_{NT} \left[\frac{1}{NT}W'u\right] \\
&= \arg\min_{\theta\in\Theta} \left[\frac{1}{NT}W'\left[y - Z(\theta_\lambda)\theta_\beta\right]\right]' A_{NT} \left[\frac{1}{NT}W'\left[y - Z(\theta_\lambda)\theta_\beta\right]\right]
\end{aligned}
$$

where A_{NT} is an $NT \times NT$, possibly random, positive semi-definite weighting matrix with rank at least $k+4$, and the last line shows that nonlinearity comes from the subset of parameters $\theta_\lambda = (\overline{\lambda}, \sigma'_\lambda)'$ governing the locus of the tilting point of the yield curve.

One particular case of the above estimator is when $W_t = \left[Z_t(\theta_\lambda), \frac{\partial Z_t(\theta_\lambda)\theta_\beta}{\partial\theta_\lambda}\right]$ and $A_{NT} = \mathbf{I}_{NT}$, which results in the Nonlinear Least Squares estimator. Here, the associated covariance matrix is given by

$$
\Omega = E\left(\nabla'_\theta A_0 \nabla_\theta\right)^{-1} E\left(\nabla'_\theta A_0 V_0 A_0 \nabla_\theta\right) E\left(\nabla'_\theta A_0 \nabla_\theta\right)^{-1}
$$

where the asymptotic variance is $V_0 = \lim_{N,T\to\infty} \frac{1}{NT} Var\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N} u_{it}w_{it}\right)$, $\nabla_\theta := \frac{\partial G_{NT}(\theta)}{\partial\theta} = \left(\nabla_{\overline{\beta}}, \nabla_{\sigma_\beta}, \nabla_{\overline{\lambda}}, \nabla_{\sigma_\lambda}\right)$, with details given in the Appendix, and A_0 is a positive semi-definite matrix such that $A_{NT} \to^p A_0$.

The estimation problem raises a number of issues. First, the model is robust to curse of dimensionality issues, as the dimension of the parameter vector does not increase with the number of yields, but with the number of state variables, which is kept at a manageable size. As a result, one does need to restrain the number of yields used when estimating the yield curve, a fact which brings the undesirable consequence of poorly measuring the term structure curvature and, as a result, poorly estimating the connection between this factor and any state variables associated to it.

Second, more than just allowing the comparison of alternative specifications, one can test competing theories about variables driving the term structure dynamics using inference tools.

Finally, and in contrast with most of the literature, the estimation makes use of both the cross-sectional and time series dimensions of the data, resulting in much faster convergence of the parameter estimates.[19] This is of special interest given issues commonly raised against VAR models used in the analysis of monetary policy: Rudebusch (1998), for instance, points out that the use of quarterly data, together with the relatively frequent changes in monetary policy in the postwar period results in either short time series or misspecified VAR models, thus making inference unreliable: using quarterly data, the twenty years of the 'Greenspan era' correspond to only 80 observations.

2.4 Finite-Sample Performance

This section presents a simulation study investigating the finite-sample performance of the estimation method. To do so, I generate state variables M_{t-}, regressors X_t, population parameter values, and errors to generate the variables y_t. For every experiment, I compute the results of 500 replications, with time-series and cross section dimensions given by, respectively, T $(= 10, 50, 100)$ and N $(= 25, 50, 100)$.

The state variables M_{t-} are constructed by taking the exponent of independent standard Gaussian random variables, the regressors X_t are standard Gaussian random variables, whereas the error terms u_t are Gaussian variables with a variance of 0.2.

In what follows, I consider the model

$$\begin{aligned} y_t(\tau) &= X_t(\lambda_t)\beta_t + u_t(\tau) \\ \begin{bmatrix} \beta_t \\ \lambda_t \end{bmatrix} &= \begin{bmatrix} \overline{\beta} \\ \overline{\lambda} \end{bmatrix} + M_{t-} \begin{bmatrix} \sigma_\beta \\ \sigma_\lambda \end{bmatrix} \end{aligned}$$

[19]The following Section illustrates the finite-sample proprties of the method and the convergence in both the maturity and time dimensions of the data.

TABLE 2.1. Simulation Results for Single-Variable Factor Specification

Panel A	**T = 10**		
	N = 25	**N = 50**	**N = 100**
$\overline{\beta}_1$	1.075	0.987	0.993
	[0.642]	[0.323]	[0.112]
$\overline{\beta}_2$	0.925	1.016	1.013
	[0.675]	[0.389]	[0.208]
$\overline{\beta}_3$	0.786	0.942	0.946
	[1.136]	[0.765]	[0.628]
$\overline{\lambda}$	0.053	0.051	0.049
	[0.030]	[0.019]	[0.011]
σ_{β_1}	0.995	0.999	1.003
	[0.102]	[0.077]	[0.042]
$\sigma_{\beta 2}$	1.004	0.999	0.995
	[0.119]	[0.105]	[0.082]
σ_{β_3}	1.064	1.036	1.024
	[0.442]	[0.334]	[0.259]
σ_{λ}	0.011	0.010	0.010
	[0.011]	[0.006]	[0.004]

Note: Standard errors are reported within square brackets.

with each factor driven by one state variable. As in the empirical exercise, I make the curvature-related factors β_{3t} and λ_t i.e. the curvature intensity and the location where the curve tilts are driven by the same state variable, so that $M_{t-} = diag\{m_{1t-}, m_{2t-}, m_{3t-}, m_{3t-}\}$. In all the experiments, $\overline{\beta} = (1,1,1)'$, $\sigma_{\beta} = (1,1,1)'$, $\overline{\lambda} = 0.05$, and $\sigma_{\lambda} = 0.01$.

The simulation results reported in Tables 2.1-3 (with standard errors inside square brackets) show fast convergence of the θ parameter estimates to their population values, with increasing precision in both N and T. For the closest case to the smallest subset of data used in the empirical section, where $N = T = 50$, the biases are negligible. In what concerns precision, estimates for $\overline{\beta}_1$ and σ_{β_1} tend to be more precisely estimated than their counterparts because the corresponding factor loadings do not involve any parameters to be estimated, thus having no uncertainty.

2.5 Application

2.5.1 *The Data*

The data set used comprises end-of-month yields from US bonds from January, 1970 to December, 2003 and US macroeconomic variables obtained from the US Federal Reserve macroeconomic database

TABLE 2.2. Simulation Results for Single-Variable Factor Specification cont'd

Panel B		**T = 50**	
	N = 25	**N = 50**	**N = 100**
$\overline{\beta}_1$	0.998	0.989	1.000
	[0.458]	[0.134]	[0.055]
$\overline{\beta}_2$	1.015	1.007	0.997
	[0.470]	[0.146]	[0.095]
$\overline{\beta}_3$	0.895	1.016	0.996
	[0.742]	[0.340]	[0.274]
$\overline{\lambda}$	0.050	0.050	0.050
	[0.015]	[0.006]	[0.005]
σ_{β_1}	1.000	1.002	1.000
	[0.040]	[0.028]	[0.014]
$\sigma_{\beta 2}$	0.996	0.999	1.000
	[0.045]	[0.033]	[0.025]
σ_{β_3}	1.033	0.995	1.002
	[0.165]	[0.094]	[0.075]
σ_{λ}	0.010	0.010	0.010
	[0.003]	[0.002]	[0.001]

Note: Standard errors are reported within square brackets.

TABLE 2.3. Simulation Results for Single-Variable Factor Specification cont'dd

Panel C		**T = 100**	
	N = 25	**N = 50**	**N = 100**
$\overline{\beta}_1$	1.003	1.000	0.998
	[0.367]	[0.108]	[0.045]
$\overline{\beta}_2$	0.999	1.002	1.004
	[0.373]	[0.129]	[0.075]
$\overline{\beta}_3$	0.956	0.985	0.985
	[0.577]	[0.269]	[0.203]
$\overline{\lambda}$	0.050	0.050	0.050
	[0.010]	[0.004]	[0.003]
σ_{β_1}	1.000	1.000	1.000
	[0.033]	[0.021]	[0.010]
$\sigma_{\beta 2}$	0.999	1.000	1.000
	[0.035]	[0.026]	[0.018]
σ_{β_3}	1.011	1.006	1.003
	[0.111]	[0.064]	[0.049]
σ_{λ}	0.010	0.010	0.010
	[0.002]	[0.001]	[0.001]

Note: Standard errors are reported within square brackets.

– the FRED – and observed at the monthly frequency.[20] For every given period, the macroeconomic variables used are predetermined with respect to the interest data used.[21]

[20] The dataset is available from http://research.stlouisfed.org/fred2/.

2.5.1.1 Interest Rates

The interest rate data used consists on the December 2003 version of the unsmoothed Fama-Bliss yields described and thoroughly discussed in Bliss (1997).[22] It includes all available issues up to that date, implying that the range of available maturities from which the term structures are estimated will not be uniform throughout the sample period i.e. I use an unbalanced panel of yields ranging from 42 to 134 observations per period. The average number of yields for the full sample is 86.944, with a standard error of 26.854, the number of periods in the full sample is $T = 408$ months, and the longest maturity used in the study is 60 months. The main features in the data are the average upward-sloping yield curve, the fact that yield volatility tends to decrease with maturity whereas persistence tends to increase with maturity.

2.5.1.2 Macroeconomic Variables

Based on the existing literature, I consider a measures of inflation, economic activity, monetary policy, and fiscal policy. The inflation measures used are the CPI (Consumer Price Index For All Urban Consumers: All Items, seasonally adjusted), PPI1 (Producer Price Index: Finished Goods, seasonally adjusted), PPI2 (Producer Price Index: All Commodities, not seasonally adjusted), PPI3 (Producer Price Index: Industrial Commodities, not seasonally adjusted), and PCE (Personal Consumption Expenditures: Chain-type Price Index, seasonally adjusted) – all measured in growth rates.

The measures of economic activity used are HOUST (Housing Starts: Total: New Privately Owned Housing Units Started, seasonally adjusted), INDPRO (Industrial Production Index, seasonally adjusted), EMP (Civilian Employment, seasonally adjusted) – all measured in growth rates – plus TCU (Capacity Utilization: Total Industry, seasonally adjusted), HELP (Index of Help Wanted Advertising in Newspapers, seasonally adjusted) and UR (Unemployment Rate, seasonally adjusted), measured in levels.

The monetary policy instruments used are FF (Federal funds effective rate), NONBR (Non-Borrowed Reserves of Depository Institutions, seasonally adjusted – the monetary aggregate the Fed targeted during the period from October, 1979 to October, 1982), and M1 (Money Stock, in Billions of Dollars, seasonally adjusted).

All the above variables are recorded at the monthly frequency, and were obtained from the FRED database. Finally, following Dai and Philippon's (2005) recent finding that fiscal policy affects the term structure, I introduce the variable DEBT, which is their quarterly fiscal policy variable interpolated to

[21] For instance, when using the yield curve of 31 March, 1970, I make sure I only use variables dated prior to that e.g. 1 March, 1970. In particular, the variables in level used date from 1 March, 1970, and the variables in growth rate are the increment from 1 February 1970 to 1 March, 1970.

[22] I thank Robert Bliss for making his data available.

the monthly frequency and divided by INDPRO, a *proxy* variable for GDP at the monthly frequency. Table 2.4 summarizes the macroeconomic variables used.

2.5.2 On The Economic Determinants of the Yield Curve

This section starts by selectively reviewing the literature addressing the relation between macroeconomic variables and the yield curve factors, thus paving the way for the empirical strategy I implement next. It goes without saying that with a set of macroeconomic variables as big as the one available from the FRED, there are countless alternative specifications to be compared ($15^3 = 3375$ using only the contemporaneous variables described above), so that a pragmatic starting point would be to consider specifications based on the existing literature and summarized in Table 2.5. The evidence documented in the literature is used to construct alternative configurations of M_{t-} which are then compared. For the sake of parsimony, I devote a section to single-variable (SV) specifications – the ones where each factor is driven by one state variable only – before addressing the general multi-variable (MV) case. I then use the 'best' SV and MV specifications in the out-of-sample comparison with the benchmark Diebold-Li model.

Much of the work in macro-finance gained momentum in the late 1990s (see Diebold, Piazzesi, and Rudebusch, 2005, and references therein for the latest account on the literature). One of the early papers is Evans and Marshall (1996) – to which Evans and Marshall (1998) also relates – where, using a VAR framework, the authors study the impact of shocks of measures of monetary policy, employment and inflation on the nominal term structure of interest rates. Their results suggest that the main effect of both employment and inflation measures is to induce a parallel shift of the yield curve, whereas (short-run) fluctuations in the slope and curvature of the yield curve are primarily attributed to the monetary policy shocks.

Also within the VAR framework, but imposing no-arbitrage restrictions, Ang and Piazzesi (2003) construct inflation and economic growth indices which they address as macro factors. By a factor representation of the pricing kernel they obtain a tractable way to examine how those macro factors affect the yield curve dynamics. However, in their study macro factors are able to explain only the short end and the middle of the yield curve. Due to difficulty to deal with the long end they introduce latent factors, now allowing the pricing kernel to be driven by both macro and latent factors. By relying on a Gaussian assumption and on the affine specification, they find that the slope and curvature factors can be explained by the macro factors, whereas the level factor can be only dealt with by using latent factors. In a related paper, but within a different framework, Piazzesi (2005) finds that monetary policy shocks change the slope of the yield curve, since they affect short rates more than long ones.

More recently, Diebold, Rudebusch and Aruoba (2006) examine the correlations between Nelson-Siegel factors and macroeconomic variables under a VAR framework and find that the level factor is highly correlated with inflation and the slope factor is highly correlated with real activity, whereas the curvature factor does not appear to be related to any of the macroeconomic variables used.

2.5.3 In-Sample Analysis

I start estimating SV specifications, where each factor is driven by one state variable only. These can be seen either as a parsimonious way of approaching the problem or as a first step before considering more complex (and difficult to compute) specifications for M_{t-}, besides providing additional out-of-sample benchmarks for those more complex specifications. A simplifying assumption made throughout the exercise is the curvature intensity β_{3t} and the parameter governing the location of the tilting point of the yield curve are the same.

Given two competing specifications with the same number of variables, I compare them using the Mean Absolute Error criterion (both the average and the median of the MAE's across time). The MAE is of special interest here for providing a model selection criterion, an idea of goodness-of-fit, and of mispricing of the specifications. Table 2.6 reports results of selected specifications from an exercise designed to select the best forecasting variables from the different categories.[23]

The preliminary results in Table 2.6 provide a number of insights on the forecasting ability of the state variables. First, the economic activity variable doing the best job at explaining the level factor is UR (see specifications 15-20), the unemployment rate; in what regards the inflation variables, their performance is less clear, but CPI and PCE tend to provide the lowest MAE's (see specifications 1-5).

As for explaining the slope, the best monetary policy variable is FF (see specifications 6-8), whereas the best economic activity variable is UR (see specifications 21-26). Finally, the best monetary policy variable explaining curvature is FF (see specifications 27-29), and the best economic activity variable is UR (see specifications 9-14).

Given the above findings plus the recent evidence that fiscal policy does play a role at explaining the curvature factor of the term structure (Dai and Philippon, 2005), we also include the variable DEBT in our empirical exercise together with the ones already mentioned. As a result, we estimate SV specifications using the three choices for the state variables explaining the level factor, the two choices explaining the slope factor, and the three choices explaining the curvature factor, being left

[23]The results are robust with respect to choice of selection criterion used – using the minimum value of the criterion function, the AIC or the BIC criteria gives the same results.

with 18 alternative specifications to examine. Table 2.7 reports the results of the comparison using the MAE criterion.

Table 2.7 shows a clear dominance of the specifications for which inflation (either CPI or PCE) explains the level, monetary policy (FF) explains the slope, and economic activity (UR) explains the curvature of the term structure. Interestingly, the fact that inflation is the key driver of the level factor holds regardless whether CPI or PCE are used, although the literature tends to prefer the latter (see DRA and Duffee, 2005). However, although in line with the literature, it does not exactly match any of the papers listed above.

In what follows, I refer to the best SV specification (specification 7 in Table 2.7) as SV – see Table 2.8 for the corresponding parameter estimates. The parameter estimates for the SV model show the positive impact of CPI on the level of the term structure, the impact of the monetary policy instrument FF on the slope (actually defined as $-\beta_{2t}$), and the impact of UR on both the intensity and the locus of the curvature, all of them found to be significant using Newey and West (1987) standard errors to account for the time dependence in the data. Interestingly, neither the CPI nor the UR are revised, which makes them even more attractive as predetermined variables with respect to yields. When coupled with the real time Taylor rule proposed in Evans (1997), the findings are consistent with what one would intuitively expect, in the sense that the yield curve tends to invert for values of FF above the Taylor rule, but remaining upward-sloping for values below the threshold.

Based on the findings in the literature, SV specifications are likely too simple to provide a satisfactory account of the term structure dynamics. The next step is thus to study the more general MV specifications. Based on the results reported in Table 2.6, I employ a general-to-specific approach starting with a specification where CPI, PCE and UR drive the level, FF and UR drive the slope, and DEBT, FF and UR drive the curvature. The alternative specifications compared in Table 2.9 show that several coefficients in the larger models are statistically insignificant. The model with the smaller BIC and with all of the parameters statistically significant is specification 7 – which I from now on refer to as MV –, which has the level driven by CPI and UR, the slope by FF and UR, and the curvature by DEBT. Albeit more parsimonious than the full model the average MAE is only slightly larger.

The parameter estimates for the MV model are reported in Table 2.10. The findings reported in Table 2.10 are in line with previous results in that economic activity and inflation drive the level factor, economic activity and monetary policy drive the slope, and fiscal policy drives the curvature factor. However, the performance of the model in terms of MAE is very similar to the SV model.

The parameter estimates – all of which significant – show the upward impact of inflation on the term structure, as expected. The parameters related to the slope also have the expected sign, with FF affecting shorter rates more strongly, but UR having the opposite effect.

2.5.4 Incorporating Economic Relations

So far, the model presented considers only state variables which are predetermined with respect to the yield curve, not exploring (i) any interdependence among them; (ii) any forecasts of their future values, both of which are expected to play a role at explaining future realizations of the yield curve. In this section I discuss how to incorporate into the model information on the joint behaviour of the state variables. Intuitively, by informing the model that certain variables are related one should expect to get more accurate results, provided the relation imposed holds.

In this section I inform the model about the joint behaviour of the state variables using a feedback interest rule, or *Taylor rule.* Taylor (1993) suggested a simple formula describing how the US Federal Open Market committee has set the Federal funds rate since 1987 as a response to measures of inflation and output gaps – this relationship has been dubbed the Taylor rule and has been extensively studied and developed since then. Despite its simplicity, the Taylor rule has a number of appealing properties. Woodford (2001) shows how it incorporates several features of an optimal monetary policy in a class of optimizing models, and provides conditions under which the Taylor rule has a stabilizing effect on the economy. More recent developments such as Clarida, Galí and Gertler (2000) propose and estimate a Taylor rule incorporating both forward- and backward-looking elements. The former account for the fact that the monetary authority is considering future paths of the output and inflation gaps when setting the current value of the monetary policy instrument, whereas the latter arises as a consequence of interest rate smoothing conducted by the monetary policy authority.[24] In what follows, I estimate both forward- and backward-looking versions of the Taylor rule. Instead of using quarterly data, as in Clarida, Galí, and Gertler (2000), I use monthly observations and find that, by and large, their results follow through to the monthly frequency.

The results of this section provide the ground for alternative ways of computing out-of-sample forecasts, in the sense that one can plug into the model estimated quantities generated by a model inspired by (or consistent with) economic theory to obtain estimates of the future behaviour of the term structure.

[24] See also Rudebusch (1995).

2.5.4.1 Taylor Rules

In what follows I consider the following specification proposed and estimated (using quarterly US data) in Clarida, Galí, and Gertler (2000), which nests both forward- and backward-looking versions of the Taylor rule and states that the target rate each period is a linear function of the gaps between expected inflation and output and their respective target levels,

$$r_t^* = rr^* + \pi^* + \gamma^\pi \left[E\left(\pi_{t,l_\pi}|\psi_t\right) - \pi^*\right] + \gamma^g E\left(g_{t,l_g}|\psi_t\right)$$

where ψ_t is the information set available at time t, π_{t,l_π} denotes the percent change in the price level between periods t and $t + l_\pi$ (expressed in annual rates), π^* is the target for inflation, $rr^*(= r^* - \pi^*)$ is the long-run equilibrium real rate, with r^* being, by definition, the desired nominal rate when both output and inflation are at their target values. g_{t,l_g} is a measure of the average output gap between periods t and $t + l_g$, with the output gap being defined as the percent deviation between actual GDP and the corresponding target.[25]

Following Clarida, Galí, and Gertler (2000), the *actual* Fed funds rate follows

$$r_t = \rho(L)r_{t-1} + (1-\rho)r_t^*$$

where $\rho(L) = \rho_1 + \rho_2 L + ... + \rho_{l_r} L^{l_r - 1}$ and $\rho = \rho(1) = \sum_{j=1}^{l_r} \rho_j$, which postulates a partial adjustment of the Fed funds rate to the target r_t^*, with ρ being an indicator of the degree of smoothing of interest changes by the monetary policy authority.

Combining the target rate and Fed funds equations results in the Taylor rule

$$r_t = (1-\rho)\left[rr^* + (1-\gamma^\pi)\pi^* + \gamma^\pi \pi_{t,l_\pi} + \gamma^g g_{t,l_g}\right] + \rho(L)r_{t-1} + \varepsilon_t$$

where $\varepsilon_t = (1-\rho)\left(\gamma^\pi \left[E(\pi_{t,l_\pi}|\psi_t) - \pi_{t,l_\pi}\right] + \gamma^g \left[E(g_{t,l_g}|\psi_t) - g_{t,l_g}\right]\right)$ is a linear combination of forecast errors, thus being orthogonal to any variable in the information set ψ_t. As one can only identify the term $rr^* + (1-\gamma^\pi)\pi^*$, but not rr^* or γ^π separately, and the inflation target is of interest, Clarida, Galí and Gertler (2000) assume that the equilibrium real rate rr^* equals its sample average. This specification allows a number of choices regarding the lead/lag periods of inflation and output, l_π and l_g, respectively, and lags for the Fed funds, l_r. The parameters of interest are π^*, γ^π, γ^g, $\{\rho_j\}_{j=1}^{l_r}$, so

[25] Typically, the information set at time t contains past values of the Fed Funds rate and other economic variables, and usually no information on current inflation and output measures.

that the dimension of the parameter vector is $3 + l_r$. It also nests a number of specifications, as shown in Table 2.11.[26].

The regression equation above implies the set of moment conditions

$$E\left(\left[r_t - (1-\rho)\left[rr^* + (1-\gamma^\pi)\pi^* + \gamma^\pi \pi_{t,l_\pi} + \gamma^g g_{t,l_g}\right] - \rho(L) r_{t-1}\right] z_t\right) = 0$$

where z_t is a vector of instruments known when r_t is set ($z_t \in \psi_t$) and π_{t,l_π}, g_{t,l_g}, and r_{t-1} also belong in ψ_t.

The above moment conditions are used to obtain parameter estimates using the Generalized Method of Moments. As in Clarida, Galí, and Gertler (2000), I set the equilibrium rate rr^* to its sample average, so as to be able to identify the inflation target. To make the feedback rule consistent with the SV specification, I replace r with FF, π with CPI, and I also follow Evans (1997)'s implementation of the Taylor rule, replacing the output gap with the unemployment gap using Okun's law, besides setting the natural rate of unemployment to $UR_t^* = UR^* = 6$.[27] Moreover, I assume that current inflation and unemployment are not observed when setting the Fed Funds rate i.e. neither of them belongs in ψ_t. The moment conditions thus become

$$E\left(\varepsilon_t^* z_t\right) = 0$$

where $\varepsilon_t^* = [FF_t - (1-\rho)[rr^* + (1-\gamma^{CPI})CPI^* + \gamma^{CPI} CPI_{t-1,l_{CPI}} + +\gamma^{UR} 3(6 - UR_{t-1,l_{UR}})] - \rho(L) FF_{t-1}]$.

Interestingly, given that in the SV specification the slope is driven by the Fed funds rate, the above specification can be linked to the interest-rate rule proposed in McCallum (1994), according to which the monetary authority reacts to term premia – the slope in particular – when setting the monetary policy instrument.[28]

The parameter estimates of the forward-looking Taylor rule are reported in Table 2.12. Although only the former is statistically significant, the responses to CPI inflation and unemployment rate are consistent with the results in Clarida, Galí, and Gertler (2000), which uses 1960:1-1996:4 data at the quarterly – as opposed to monthly – frequency. The closest inflation target level to their estimates is given by FWTR1, although not significant, and the interest rate smoothing parameter is more

[26] Note that the Taylor rule is usually applied to quarterly data, whereas I consider monthly data.

[27] Arthur Okun observed that a one percent fall in the unemployment rate from its full employment level tended to produce a three percent increase in real GDP relative to trend. See Evans (1997) for discussion and robustness checks.

[28] The McCallum interest-rate rule also allows rationalizing the empirical failure of the expectations hypothesis – see also Kugler (1997) and Gallmeyer, Hollifield, and Zin (2005).

persistent than theirs. The goodness-of-fit of the specifications is very similar and none of them is rejected when testing for overidentifying restrictions.

Forward-looking Taylor rules might give accurate descriptions in-sample, but if the aim is to do out-of sample forecasting, one needs backward-looking ones. Table 2.13 reports estimates for alternative specifications of backward-looking Taylor rules regarding the choice of l_{CPI} and l_{UR}, the horizons at which the monetary policy authority looks when setting the monetary policy instrument.

The results for the backward-looking Taylor rules are robust to alternative horizons, and suggest that – at least at the monthly frequency – the monetary authority looks mostly at past inflation and past values of the monetary policy instrument when setting its current value. The persistence in the Fed funds rate is shown to be high, and even the non-significant parameters γ^{UR} and CPI^* tend to gravitate across a relatively narrow interval, at least for non-zero values of l_{CPI} and l_{UR}. The J-statistics suggest that the horizon at which the Fed looks is at least six months back. When compared to the forward-looking estimates, the responses to inflation seem to be tougher, and both the response to unemployment and the inflation target level are found not to be statistically significant.

2.5.5 Out-of-Sample Analysis

In this section I perform an out-of-sample study by considering five episodes of economic interest: the five NBER-dated US recessions which have entirely occurred during the period 1970-2003. Recessions are of economic interest *per se* being bad states of nature, characterized by reduced economic activity and increased lay-off of workers, thus being events against which economic agents are willing to insure. Moreover, within the term structure literature, recessions are of interest for being periods which tend to be preceded by the inversion of the yield curve, a feature often difficult to be quickly captured – if at all – by term structure models, making the exercise both more interesting and challenging. The recessions considered are described in Table 2.14.[29]

For every month in each of the five recessions, I compare the forecasts of the alternative specifications using two measures of accuracy . I also report results for specifications SV-TR and MV-TR, which incorporate the Taylor rule in an attempt to improve forecasting ability.

The ways I compute the out-of-sample forecasts are as follows:

For the macro-based specifications, assume the estimation sample has observations from periods $t = 1, ..., T$, where $t = 1$ is January, 1970 and $t = T$ is the month preceding the recession of interest. After obtaining parameter estimates using the estimation sample, the yield curve forecast for period

[29] The NBER-dated recession going from December 1969 to November 1970 is not considered here since the dataset starts on January, 1970.

$t^* > T$, denoted as $\widehat{y}_{t^*|T}$ are obtained either from observed or estimated values M_{t^*} of the state variables using the previously estimated parameters – in the case of the SV and MV specifications, I keep the parameter estimates fixed and keep updating the matrix M_{t^*} of state variables every period

When it comes to the SV-TR and MV-TR specifications, I estimate the Taylor rules as above, and just update the information on CPI, UR, and FF every period, thus obtaining a Taylor rule-based estimate of the value of FF the following period.

Finally, for the DL model, I estimate the model for every period $t = 1, ..., T$, compute the AR(1) processes describing the dynamics of each factor, and re-estimate the model at every period $t > t^*$.[30]

As a measure of 'overall accuracy', I compute average MAEs for the entire duration of each recession i.e. for every month t^* of a given recession, I compute

$$OA_{t^*} = \frac{1}{N_{t^*}} \sum_{j=1}^{N_{t^*}} \left|\widehat{y}_{t^*}(\tau_j) - y_{t^*}(\tau_j)\right|, t^* \in \text{Recession}$$

where N_{t^*} refers to the number of yield at period t^* in the recession. The results reported in Tables 2.15-16.

The results in Tables 2.15-16 show that the macro-based specifications consistently outperform the latent variable model. As a matter of fact, DL cannot beat its competitors for any month in recessions R3-5. When comparing SV and MV specifications, the former tends to perform better in the first two or three months of the recessions, being then outperformed by the latter. This suggests that it might take time for all the state variables to work in favour of the MV specification in such periods.

Panel A shows the dominance of the MV-TR model, especially during the second half of the recession. Its performance is followed by the SV-TR model, which suggests that Taylor rules convey information about the future state of the term structure.

Panel B shows the potential effect of a change in policy regime on the forecasting ability of the TR specifications – R2 was the first recession following the monetary policy experiment, right after its introduction.[31] As a result, the SV and MV-TR models perform closely, and the Taylor rule does not seem to provide a substantial gain to the models incorporating it.

Panels C-E show a clear dominance of the MV-TR specification, which might suggest two things. First, that SV specifications are way too simple to describe the term structure dynamics. Second, that incorporating Taylor rules does indeed play a role, improving the accuracy of the forecasts.

[30] See Diebold and Li (2005) for a thorough out-of-sample comparison of their model and previously existing ones.

[31] See Clarida, Gali, and Gertler (2000) for a study of how the Taylor rule changed with this regime change.

As a measure of 'maturity-disaggregated accuracy', I report time-averaged MAEs for fixed maturities i.e. for a given maturity τ_j I calculate

$$MDA_{\tau_j} = \frac{1}{\#t^*} \sum_{t^* \in \text{Recession}} \left|\widehat{y}_{t^*}(\tau_j) - y_t(\tau_j)\right|, j = 1, ..., N$$

where $\#t^*$ denotes the number of periods in the recession. The results are reported in Table 2.17-19.

The results reported in Tables 2.17-19 confirm the view that macro-based specifications outperform the benchmark DL model. Panel A shows the superior performance of the MV-TR specification up to the 36-month maturity, after which the DL specification tends to do better.

Panels B and C show the superior performance of the MV and, to a lesser extent, SV specifications, most likely due to the change in the policy regime resulting from the monetary policy experiment. Panels D and E show a dominance of the MV specification, at least for maturities up to 10-12 months. In Panel D, the better performing specification from the 11-month maturity towards the long end of the curve is MV-TR, whereas in Panel E it is specification SV which performs better between the 24- to 60-month maturities.

Although it is not obvious which macro-based specification performs best throughout the exercises, all of them consistently outperform the latent-variable benchmark.

Conclusion

This paper proposes a term structure model whose factors are uniquely driven by observable – as opposed to latent – state variables. The explicit link between the term structure factors and the state variables allows comparing alternative views on the drivers of its dynamics and competing economic hypotheses.

The method is robust to curse of dimensionality issues commonly appearing in the literature. This happens because instead of increasing with the number of observations (yields) used, the dimension of the parameter vector increases with the number of state variables, which is kept at a manageable size. As a result, the method is in a position to deliver more accurate measures of the curvature factor, thus better explaining intermediate maturities i.e. the 'belly' of the curve.

The estimation method uses both the cross-sectional and time series dimensions of the data, which results in faster convergence of the parameter estimates and more reliable inference. This is in stark contrast with VAR models, which are subject to the criticism that they make researchers choose between either short time series or misspecified models, thus making inference unreliable – a direct consequence of the frequent changes in policy regimes in the postwar period (Rudebusch, 1998).

The empirical exercise uses a comprehensible set of US macroeconomic data to compare alternative specifications of the term structure. In the in-sample study, the baseline (SV) specification is such that the level, slope and curvature factors are driven by, respectively, measures of inflation (CPI growth), monetary policy (the Fed Funds rate), and economic activity (the unemployment rate). The out-of-sample study compares macro-based models to a latent-variable benchmark model for the five NBER-dated recessions which occurred in the last three decades, showing that the former consistently outperforms the latter – a finding which is robust to alternative criteria.

This paper raises a number of questions for future research. First, how does the method perform using alternatives such as expectations variables obtained in consensus forecasts as state variables.

Second, how it performs as a risk management tool, making it appealing to both financial institutions and regulators, especially under the ongoing Basel II process.

Third, how it performs when coupled with VAR models feeding it with macroeconomic variables, or measures such as the Bernanke and Mihov (1998) monetary policy indicator.

Fourth, how it can be adapted to the study of credit risk, either at the country or the corporate level.

Finally, although the method relies on the Nelson-Siegel yield curve fitting method, it is by no means restricted to it. Nelson-Siegel is used here due to its intuitive appeal, well-known properties, and the common understanding that it is a reasonable first-order approximation to the yield curve. Alternative methods can be also used, and are left for future research.

TABLE 2.4. Macroeconomic Variables by Group

Economic Activity	**Inflation**	**Monetary Policy**	**Fiscal Policy**
UR$^{\mathcal{L}}$	CPI	FF$^{\mathcal{L}}$	DEBT
TCU$^{\mathcal{L}}$	PCE	NONBR	
HELP$^{\mathcal{L}}$	PPI1	M1	
IP	PPI2		
EMP	PPI3$^{\mathcal{N}}$		
HOUST			

Note: Variables in levels and not seasonally adjusted are marked with the superscripts $^{\mathcal{L}}$ and $^{\mathcal{N}}$, respectively. The remaining variables are measured in growth rates and are seasonally adjusted.

TABLE 2.5. Macroeconomic Variables Driving Term Structure Factors

Reference	**Level Factor**	**Slope Factor**	**Curvature Factor**
Evans & Marshall (1996)	Employment Inflation	Monetary Policy	Monetary Policy
Ang & Piazzesi (2003)	–	Inflation	Output
Piazzesi (2005)	–	Monetary Policy	–
DRA (2005)	Inflation	Output	–

2.A Appendix A: Covariance Matrix Derivation

As in the text, $\nabla_\theta = \left(\nabla_{\overline{\beta}}, \nabla_{\sigma_\beta}, \nabla_{\overline{\lambda}}, \nabla_{\sigma_\lambda}\right)$, where

$$\nabla_{\overline{\beta}} = X(\theta_\lambda)$$

$$\nabla_{\sigma_\beta} = XM(\theta_\lambda)$$

$$\nabla_{\overline{\lambda}} = \frac{\partial X(\theta_\lambda)}{\partial \overline{\lambda}}\overline{\beta} = \begin{bmatrix} \frac{\partial X_1(\theta_\lambda)}{\partial \overline{\lambda}} \\ \\ \frac{\partial X_t(\theta_\lambda)}{\partial \overline{\lambda}} \\ \\ \frac{\partial X_T(\theta_\lambda)}{\partial \overline{\lambda}} \end{bmatrix} \overline{\beta}$$

with general element

$$\frac{\partial X_t(\theta_\lambda)}{\partial \overline{\lambda}} = \begin{bmatrix} 0 & \phi_{1t-} & \phi_{1t-} + \exp(-\left(\overline{\lambda} + M_{\lambda t-}\sigma_\lambda\right)\tau_1) \\ 0 & \phi_{2t-} & \phi_{2t-} + \exp(-\left(\overline{\lambda} + M_{\lambda t-}\sigma_\lambda\right)\tau_2) \\ \dots & \dots & \dots \\ 0 & \phi_{Nt-} & \phi_{Nt-} + \exp(-\left(\overline{\lambda} + M_{\lambda t-}\sigma_\lambda\right)\tau_N) \end{bmatrix}$$

TABLE 2.6. Preliminary Results for Single-Variable Specifications

Specification	Level	Slope	Curvature	Avg(MAE)	Med(MAE)
1	CPI	FF	M1	1.010	0.830
2	PCE	FF	M1	1.031	0.825
3	PPI1	FF	M1	1.012	0.836
4	PPI2	FF	M1	1.025	0.830
5	PPI3	FF	M1	1.023	0.833
6	PCE	FF	DEBT	1.069	0.869
7	PCE	NONBR	DEBT	1.822	1.395
8	PCE	M1	DEBT	1.823	1.415
9	CPI	FF	UR	0.833	0.696
10	CPI	FF	TCU	1.022	0.851
11	CPI	FF	HELP	1.058	0.893
12	CPI	FF	IP	1.036	0.873
13	CPI	FF	EMP	1.042	0.898
14	CPI	FF	HOUST	1.050	0.889
15	UR	FF	DEBT	0.877	0.736
16	TCU	FF	DEBT	1.041	0.854
17	HELP	FF	DEBT	1.069	0.873
18	IP	FF	DEBT	1.053	0.881
19	EMP	FF	DEBT	1.063	0.851
20	HOUST	FF	DEBT	1.062	0.865
21	PCE	UR	M1	1.629	1.284
22	PCE	TCU	M1	1.817	1.449
23	PCE	HELP	M1	1.668	1.279
24	PCE	IP	M1	1.815	1.448
25	PCE	EMP	M1	1.817	1.437
26	PCE	HOUST	M1	1.811	1.401
27	UR	FF	FF	0.854	0.738
28	UR	FF	NONBR	0.884	0.746
29	UR	FF	M1	0.872	0.739

Note: The last two columns report, respectively, the average and the median MAE across time.

and

$$\phi_{it-} = \frac{\exp(-\left(\overline{\lambda} + M_{\lambda t-}\sigma_\lambda\right)\tau_i)}{\left(\overline{\lambda} + M_{\lambda t-}\sigma_\lambda\right)\tau_i}\overline{\lambda} - \frac{1 - \exp(-\left(\overline{\lambda} + M_{\lambda t-}\sigma_\lambda\right)\tau_i)}{\left[\left(\overline{\lambda} + M_{\lambda t-}\sigma_\lambda\right)\tau_i\right]^2}\tau_i, i = 1, ..., N$$

Finally,

$$\nabla_{\sigma_\lambda} = \frac{\partial X(\theta_\lambda)}{\partial \overline{\lambda}} \begin{bmatrix} M_{\beta_1}\sigma_\beta \\ M_{\beta_2}\sigma_\beta \\ \cdots \\ M_{\beta_N}\sigma_\beta \end{bmatrix} \begin{bmatrix} M_{\lambda_1} \\ M_{\lambda_2} \\ \cdots \\ M_{\lambda N} \end{bmatrix}$$

TABLE 2.7. Further Results for Single-Variable Factor Specifications

Specification	Level	Slope	Curvature	Avg(MAE)	Med(MAE)
1	CPI	FF	DEBT	1.048	0.886
2	PCE	FF	DEBT	1.069	0.869
3	UR	FF	DEBT	0.876	0.729
4	CPI	FF	M1	1.010	0.830
5	PCE	FF	M1	1.031	0.825
6	UR	FF	M1	0.872	0.729
7*	CPI	FF	UR	0.833	0.696
8	PCE	FF	UR	0.836	0.701
9	UR	FF	UR	0.873	0.727
10	CPI	UR	DEBT	1.048	0.886
11	PCE	UR	DEBT	1.468	1.203
12	UR	UR	DEBT	1.521	1.272
13	CPI	UR	M1	1.462	1.199
14	PCE	UR	M1	1.467	1.170
15	UR	UR	M1	1.526	1.287
16	CPI	UR	UR	1.462	1.211
17	PCE	UR	UR	1.467	1.179
18	UR	UR	UR	1.523	1.262

Note: (i) The last two columns report, respectively, the average and the median MAE across time (the smaller quantities of every column are underlined). The superscript * indicates the best specification according to the MAE criterion.

TABLE 2.8. Results for Best Single-Variable Factor Specification

Specification		$(\widehat{\beta}', \widehat{\lambda})'$		$(\widehat{\sigma}'_{\beta}, \widehat{\sigma}_{\lambda})'$		Avg-Med(MAE)
SV:	$\beta_{1t}: CPI$	10.655	[0.091]	55.820	[1.317]	0.833 – 0.696
	$\beta_{2t}: FF$	−10.148	[0.089]	0.892	[0.003]	
	$\beta_{3t}: UR$	−13.661	[0.169]	2.177	[0.019]	
	$\lambda_t: UR$	0.007	[0.001]	0.004	[0.001]	

Note: Newey-West standard errors with 12 lags are reported within square brackets.

2.B Appendix B: Robustness Check Using the CRSP Data

In this Appendix I estimate a simplified version of the model on CRSP data and show that the SV and MV specifications still outperform the latent-variable benchmark even when pre-specifying the parameter λ, as in Diebold and Li (2005). This once again suggests that, besides the advantages discussed in the text, observable state variables do play a role out-of-sample.

TABLE 2.9. Results for Multi-Variable Factor Specifications

Specification	Level	Slope	Curvature	Avg(MAE)	BIC
1	CPI, PCE$^{\varnothing}$, UR	FF, UR	DEBT, FF$^{\varnothing}$, UR$^{\varnothing}$	0.759	49.191
2	CPI, UR	FF, UR	DEBT, FF$^{\varnothing}$, UR	0.759	45.097
3	CPI, PCE$^{\varnothing}$, UR	FF, UR	DEBT, UR$^{\varnothing}$	0.844	45.100
4	CPI, PCE, UR	FF, UR	DEBT$^{\varnothing}$, FF$^{\varnothing}$	0.800	45.099
5	CPI, UR	FF, UR	DEBT, FF	0.972	41.025
6	CPI, PCE$^{\varnothing}$, UR	FF, UR	DEBT, FF$^{\varnothing}$, UR$^{\varnothing}$	0.819	41.004
7*	CPI, UR	FF, UR	DEBT	0.820	36.910

Note: The last two columns report, respectively, the average MAE across time (the smaller quantities in every column are underlined) and Schwarz's BIC model selection criterion. The superscript $^{\varnothing}$ denotes non-significance of the corresponding parameter, whereas the superscript * indicates the best specification according to the BIC.

TABLE 2.10. Results for Multi-Variable Factor Specification

Specification		$(\widehat{\overline{\beta}}', \widehat{\overline{\lambda}})'$		$(\widehat{\sigma}'_{\beta}, \widehat{\sigma}_{\lambda})'$		Avg-Med(MAE)
MV:	$\beta_{1t} : CPI$	−2.017	[0.304]	47.564	[1.330]	0.820
	$\beta_{1t} : UR$			1.791	[0.014]	
	$\beta_{2t} : FF$	2.094	[0.284]	0.886	[0.004]	
	$\beta_{2t} : UR$			−1.727	[0.017]	
	$\beta_{3t} : DEBT$	2.011	[0.551]	−6.029	[0.153]	
	$\lambda_t : DEBT$	0.031	[0.001]	−0.035	[0.002]	

Note: Newey-West standard errors with 12 lags are reported within square brackets.

TABLE 2.11. Specifications Nested within the Interest Rate Feedback Rule

Specification	l_π	l_g	l_r
(Backward) Taylor rule	< 0	< 0	–
(Backward) Taylor rule with interest rate smoothing	< 0	< 0	< 0
Clarida-Galí-Gertler	> 0	> 0	< 0

Note: The symbols l_π, l_g, l_r denote, respectively, the lags (or forward shifts, for negative values) of the inflation, economic activity and interest rate variables.

2.B.1 The Data

The data set used comprises end-of-month price quotes (bid-ask average) of US bonds from June, 1964 to March, 2000 collected by CRSP. Other than the bond yields, all remaining data are from the US Federal Reserve's macroeconomic database – the FRED –, observed at the monthly frequency.

2.B.1.1 Interest Rates

For every period I consider 17 maturities, going up to the 10-year maturity for a total of 430 months. The maturities used are as follows: 1 to 12 months, 24, 36, 48, 60, and 120 months. Although the

TABLE 2.12. Parameter Estimates for Forward-Looking Interest Rate Rule

	FWTR1	FWTR2	FWTR3	FWTR4
l_{CPI}	1	1	1	1
l_{UR}	1	1	1	1
l_{FF}	−1	−1	−1	−1
Instrument lags	1	2	3	4
γ^{CPI}	2.427*	1.929**	1.702**	1.605**
	[1.358]	[0.870]	[0.770]	[0.746]
γ^{UR}	0.454	0.473	0.618	0.615
	[0.498]	[0.381]	[0.394]	[0.404]
CPI^*	2.020	1.640	1.085	0.911
	[2.898]	[3.054]	[3.461]	[4.084]
ρ	0.965***	0.952***	0.948***	0.950***
	[0.002]	[0.002]	[0.002]	[0.002]
R^2	0.993	0.993	0.993	0.993
J-statistic	0.377	7.787	10.654	12.894
df	2	5	8	11

Note: Specification FWTR1 uses a constant, current values of CPI and UR, and lagged values of CPI, UR, and FF as instruments. Specification FWTR2-4 use the same instruments as FWTR1 plus 2-4 lagged versions of CPI, UR, and UR. Newey-West standard errors with 12 lags are reported inside square brackets. Significance at the 10, 5, and 1 percent levels is denoted by superscripts *, **, and ***, respectively.

TABLE 2.13. Parameter Estimates for Backward-Looking Interest Rate Rule

	BWTR1	BWTR2	BWTR3	BWTR4
l_{CPI}	0	−2	−6	−10
l_{UR}	0	−2	−6	−10
l_{FF}	−1	−1	−1	−1
Instrument lags	1	1	1	1
γ^{CPI}	2.264	2.273	2.355*	2.866*
	[2.127]	[1.408]	[1.388]	[1.691]
γ^{UR}	1.158	0.410	0.598	0.352
	[1.380]	[0.504]	[0.431]	[0.416]
CPI^*	1.257	2.270	1.715	2.089
	[6.491]	[3.305]	[2.177]	[1.882]
ρ	0.986***	0.964***	0.951***	0.949***
	[0.001]	[0.002]	[0.002]	[0.002]
R^2	0.993	0.993	0.993	0.993
J-statistic	26.153***	6.326*	1.128	0.860
df	2	2	2	2

Note: Specification BWTR1 uses a constant, current values of CPI and UR, and lagged values of CPI, UR, and FF as instruments. Specification BWTR2-4 use the same instruments as BWTR1 plus 2-4 lagged versions of CPI, UR, and UR. Newey-West standard errors with 12 lags are reported inside square brackets. Significance at the 10, 5, and 1 percent levels is denoted by superscripts *, **, and ***, respectively.

analysis does not require the maturities to be fixed, this greatly simplifies the empirical exercise. Table 2.20 reports some sample statistics of the bond data.

TABLE 2.14. NBER-Dated Recessions Considered

Recession Code	Start Date	End Date	Duration
R1	November, 1973	March, 1975	16 months
R2	January, 1980	July, 1980	6 months
R3	July, 1981	November, 1982	16 months
R4	July, 1990	March, 1991	8 months
R5	March, 2001	November, 2001	8 months

The main features in the data are the average upward-sloping yield curve, the fact that yield volatility tends to decrease with maturity whereas persistence tends to increase with maturity. The autocorrelations of all yields are individually significant up to lag nine (results available upon request).

2.B.1.2 Macroeconomic Variables

Based on the existing literature, I consider a number of measures of inflation, economic activity, monetary policy, and fiscal policy. The inflation measures used are the CPI (Consumer Price Index For All Urban Consumers: All Items), PPI1-3 (Producer Price Index: Finished Goods, All Commodities, and Industrial Commodities, respectively), and PCE (Personal Consumption Expenditures: Chain-type Price Index) - all measured in growth rates; the measures of economic activity used are HOUST (Housing Starts: Total: New Privately Owned Housing Units Started), INDPRO (Industrial Production Index), the HELP index (Index of Help Wanted Advertising in Newspapers), UR (Unemployment Rate), and EMP (Civilian Employment) - both HELP and UR are considered in levels and growth rates; the monetary policy instruments used are FF (Federal funds effective rate), NONBR (Non-Borrowed Reserves of Depository Institutions), and M1 (M1 Money Stock, in Billions of Dollars). All these variables are seasonally adjusted, of monthly frequency, and were obtained from the FRED database. Finally, following Dai and Philipon (2005)'s recent finding that fiscal policy affects the term structure, the fiscal policy variable used is DEBT (Outstanding Credit Market Debt of U.S. Government, State and Local Governments, and Private Nonfinancial Sectors).

2.B.2 In-Sample Analysis

2.B.2.1 Single-Variable Factor Specifications

The empirical implementation starts by investigating specifications where each factor is driven by one state variable only i.e. $M_{t-} = diag\{m_{1t-}, m_{2t-}, m_{3t-}\}$. This can be seen either as a parsimonious way of approaching the problem or as a first step before considering more complex specifications for M_{t-}.

TABLE 2.15. Overall Accuracy of Alternative Specifications - Recessions 1-3

Panel A	**Recession R1**				
	DL	**SV**	**SV-TR**	**MV**	**MV-TR**
1st month	1.31	1.56	*0.65*	1.71	1.00
2nd month	*0.81*	1.16	0.82	1.22	1.13
3rd month	*0.62*	1.17	0.79	1.16	1.08
4th month	0.60	0.56	*0.39*	0.68	0.59
5th month	1.09	0.45	*0.33*	0.52	0.41
6th month	1.48	0.95	0.85	0.87	*0.67*
7th month	1.37	1.21	1.13	*0.94*	1.00
8th month	1.57	1.67	1.60	1.39	*1.15*
9th month	1.74	1.69	1.67	1.32	*1.15*
10th month	1.88	1.12	1.13	0.83	*0.67*
11th month	1.14	1.05	1.06	0.92	*0.63*
12th month	1.38	0.71	0.79	0.61	*0.24*
13th month	1.05	0.34	0.38	0.40	*0.28*
14th month	0.77	*0.37*	0.40	0.57	0.43
15th month	0.99	0.71	*0.65*	1.79	1.07
16th month	*0.73*	1.09	1.03	1.68	1.65
Panel B	**Recession R2**				
	DL	**SV**	**SV-TR**	**MV**	**MV-TR**
1st month	4.48	*0.88*	1.38	1.03	1.26
2nd month	4.57	*1.69*	1.77	1.82	*1.69*
3rd month	4.28	1.48	1.35	1.45	*1.28*
4th month	*1.99*	3.22	2.26	3.25	2.55
5th month	1.81	*0.81*	1.08	0.90	1.18
6th month	1.60	0.91	0.74	0.83	*0.66*
Panel C	**Recession R3**				
	DL	**SV**	**SV-TR**	**MV**	**MV-TR**
1st month	1.81	*0.88*	5.36	1.01	4.98
2nd month	3.50	2.47	2.58	2.47	*1.89*
3rd month	5.18	2.61	2.65	2.58	*2.25*
4th month	4.98	2.32	2.31	2.40	*2.05*
5th month	3.72	1.71	1.68	1.65	*1.52*
6th month	4.84	2.43	2.36	2.22	*2.14*
7th month	4.94	2.32	2.28	*1.93*	2.00
8th month	4.85	1.21	1.26	1.25	*0.90*
9th month	4.58	2.49	2.52	2.39	*2.30*
10th month	4.74	1.03	1.08	*0.91*	*0.91*
11th month	4.79	0.69	0.72	0.74	*0.59*
12th month	5.72	1.54	1.59	1.53	*1.40*
13th month	5.02	1.55	1.58	1.62	*1.51*
14th month	4.47	2.67	2.66	2.53	*2.49*
15th month	2.98	1.76	1.75	*1.59*	1.63
16th month	2.81	1.87	1.86	1.71	*1.69*

Note: The quantities in italics are the smaller values for a given time period and episode.

TABLE 2.16. Overall Accuracy of Alternative Specifications - Recessions 4-5

Panel D	**Recession R4**				
	DL	**SV**	**SV-TR**	**MV**	**MV-TR**
1st month	1.30	0.46	*0.35*	0.40	0.38
2nd month	0.97	0.90	0.93	0.91	*0.82*
3rd month	1.02	0.71	0.72	0.73	*0.64*
4th month	1.11	0.64	0.64	0.67	*0.59*
5th month	1.52	0.58	0.58	0.59	*0.53*
6th month	1.77	0.57	0.57	0.56	*0.51*
7th month	1.86	0.86	0.86	0.87	*0.82*
8th month	1.88	1.00	0.99	0.98	*0.95*
Panel E	**Recession R5**				
	DL	**SV**	**SV-TR**	**MV**	**MV-TR**
1st month	1.79	0.84	3.45	*0.69*	3.75
2nd month	2.52	0.64	0.65	0.60	*0.56*
3rd month	2.98	0.71	0.72	0.78	*0.55*
4th month	3.24	0.93	0.94	0.91	*0.74*
5th month	3.71	0.78	0.79	0.75	*0.71*
6th month	4.03	0.96	0.97	0.90	*0.86*
7th month	4.72	1.29	1.30	1.32	*1.21*
8th month	4.79	1.23	1.23	*1.17*	1.34

Note: The quantities in italics are the smaller values for a given time period and episode.

Table 2.21 shows that the best performing specification has PCE explaining the level, FF explaining the slope, and DEBT explaining the curvature.

2.B.2.2 Multi-Variable Factor Specifications

Based on the findings in the literature and the results obtained for the SV case, I now allow for more state variables to influence the term structure factors. The findings reported in Tables 2.22-23 are in line with previous results in that inflation (actually, two measures of inflation, PCE and PPI1) drives the level factor, monetary policy drives the slope, and fiscal policy drives the curvature factor. The model is surprisingly similar to the SV specification previously obtained, as their differ only by the inclusion of the extra measure of inflation driving the level factor. Tables 2.22-23 display the results.

As opposed to previous findings in the literature, however, no inclusion of economic activity measures was found to improve on the best specification obtained improved the goodness-of-fit of the model. This finding could be rationalized by arguing that economic agents take into account some form of the Taylor rule when looking at the economic variables available to them and analyzing their impact on the yield curve. Hereafter we refer to the best specification for the multi-variable case (with level

TABLE 2.17. Maturity-Disaggregated Accuracy - Recessions 1 and 2

Panel A	**Recession R1**				
	DL	**SV**	**SV-TR**	**MV**	**MV-TR**
1mo	2.45	0.90	0.96	0.90	*0.73*
2mo	1.36	0.86	0.90	0.86	*0.70*
3mo	*0.63*	0.83	0.85	0.82	0.67
4mo	0.95	0.80	0.81	0.80	*0.65*
5mo	1.49	0.57	0.57	*0.54*	0.58
6mo	1.93	0.60	*0.57*	0.58	0.58
7mo	2.23	0.62	*0.56*	0.63	0.59
8mo	2.36	0.66	0.56	0.66	*0.55*
9mo	2.37	0.71	0.59	0.71	*0.58*
10mo	2.27	0.75	0.63	0.76	*0.61*
11mo	2.12	0.78	0.65	0.80	*0.63*
12mo	1.94	0.81	*0.66*	0.90	0.72
24mo	1.27	0.91	*0.76*	1.00	*0.76*
36mo	0.84	0.91	0.80	1.01	*0.79*
48mo	*0.70*	1.02	0.86	1.03	0.77
60mo	*0.93*	2.13	1.95	2.21	2.01
Panel B	**Recession R2**				
	DL	**SV**	**SV-TR**	**MV**	**MV-TR**
1mo	2.89	1.61	2.32	*1.57*	2.17
2mo	2.57	1.60	2.30	*1.55*	2.14
3mo	2.81	1.57	2.26	*1.51*	2.09
4mo	3.09	1.52	2.20	*1.46*	2.02
5mo	3.42	1.45	2.12	*1.39*	1.95
6mo	3.67	1.33	1.98	*1.31*	1.81
7mo	3.94	1.25	1.86	*1.24*	1.69
8mo	4.12	*1.19*	1.74	*1.19*	1.58
9mo	4.19	1.15	1.63	*1.14*	1.49
10mo	4.21	1.13	1.54	*1.09*	1.40
11mo	4.22	*1.03*	1.49	1.10	1.37
12mo	4.22	*1.01*	1.47	1.13	1.33
24mo	3.06	1.02	0.97	1.04	*0.85*
36mo	2.80	*0.92*	0.96	0.93	1.03
48mo	1.89	1.85	*1.56*	1.89	1.68
60mo	4.00	3.38	2.95	3.45	*2.87*

Note: The quantities in italics are the smaller values for a given time period and episode.

being driven by PCA and PPI1, slope driven by FF, and curvature being driven by DEBT) as the MV specification.

TABLE 2.18. Maturity-Disaggregated Accuracy - Recessions 3 and 4

Panel C	**Recession R3**				
	DL	**SV**	**SV-TR**	**MV**	**MV-TR**
1mo	3.73	*1.44*	1.89	1.49	1.73
2mo	1.64	*1.43*	1.88	1.48	1.71
3mo	*0.79*	1.44	1.90	1.48	1.71
4mo	1.65	*1.46*	1.91	1.51	1.71
5mo	2.72	*1.40*	1.85	*1.40*	1.60
6mo	3.57	1.48	1.91	*1.45*	1.66
7mo	4.17	1.55	1.97	*1.50*	1.71
8mo	4.58	1.51	1.93	*1.43*	1.66
9mo	4.84	1.57	1.97	*1.51*	1.72
10mo	5.01	1.62	2.01	*1.58*	1.77
11mo	5.11	1.68	2.05	*1.65*	1.82
12mo	5.18	1.78	2.14	*1.71*	1.83
24mo	4.57	1.37	1.74	*1.24*	1.40
36mo	5.01	2.16	2.32	*2.02*	2.06
48mo	3.63	1.66	1.90	*1.61*	1.71
60mo	*4.28*	4.61	4.81	4.57	4.69
Panel D	**Recession R4**				
	DL	**SV**	**SV-TR**	**MV**	**MV-TR**
1mo	6.90	0.68	0.79	*0.65*	0.75
2mo	5.77	0.71	0.82	*0.69*	0.78
3mo	4.83	0.71	0.82	*0.70*	0.78
4mo	4.07	0.70	0.80	*0.69*	0.76
5mo	3.47	*0.67*	0.78	*0.67*	0.72
6mo	3.00	*0.67*	0.77	*0.67*	0.71
7mo	2.64	*0.66*	0.70	*0.66*	*0.66*
8mo	2.35	0.39	0.45	*0.33*	0.39
9mo	2.13	0.39	0.45	*0.35*	0.39
10mo	1.93	0.40	0.44	*0.37*	0.38
11mo	1.77	0.41	0.44	0.41	*0.39*
12mo	1.63	0.43	0.45	0.44	*0.40*
24mo	1.35	0.57	0.56	0.53	*0.48*
36mo	0.98	*0.47*	0.50	0.49	0.50
48mo	*0.42*	1.25	1.17	1.26	1.11
60mo	0.84	0.69	0.61	0.70	*0.60*

Note: The quantities in italics are the smaller values for a given time period and episode.

2.B.2.3 In-Sample Comparison of Specifications

Once parameter estimates were obtained I am now in a position to compare the in-sample behaviour of the specifications. Table 2.24 reports results in terms of R^2 and MAE quantities according to which parsimony is well rewarded in our context, given how closely the measures of goodness-of-fit are.

TABLE 2.19. Maturity-Disaggregated Accuracy - Recession 5

Panel E	**Recession R5**				
	DL	**SV**	**SV-TR**	**MV**	**MV-TR**
1mo	7.01	1.36	1.73	*1.21*	2.02
2mo	6.45	1.42	1.79	*1.28*	2.07
3mo	5.98	1.45	1.82	*1.31*	2.09
4mo	5.59	1.47	1.84	*1.33*	2.10
5mo	5.28	1.47	1.87	*1.33*	2.09
6mo	5.03	1.51	1.87	*1.36*	2.10
7mo	4.84	1.50	1.47	*1.35*	2.08
8mo	4.69	1.11	1.47	*0.99*	1.71
9mo	4.57	1.10	1.47	*0.97*	1.69
10mo	4.45	1.08	1.47	*0.94*	1.65
11mo	4.35	1.08	1.44	*0.93*	1.63
12mo	4.24	1.06	1.42	*0.91*	1.59
24mo	3.33	*0.54*	0.87	0.61	0.87
36mo	2.79	*0.78*	0.95	0.80	*0.78*
48mo	2.73	0.87	1.24	*0.85*	1.02
60mo	2.35	*1.51*	1.85	1.59	1.64

Note: The quantities in italics are the smaller values for a given time period and episode.

TABLE 2.20. Basic Statistics of Yields

	Mean	**Std. Error**	**Min**	**Max**	**ACF(1)**	**ACF(9)**
1mo	6.136	2.512	2.600	16.360	0.956*	0.959*
2mo	6.315	2.549	2.740	16.170	0.971*	0.781*
3mo	6.467	2.549	2.760	16.030	0.972*	0.789*
4mo	6.545	2.598	2.810	16.100	0.973*	0.793*
5mo	6.627	2.597	2.850	16.190	0.973*	0.798*
6mo	6.688	2.594	2.850	16.520	0.974*	0.799*
7mo	6.727	2.583	2.920	16.170	0.974*	0.800*
8mo	6.780	2.577	2.930	16.300	0.975*	0.800*
9mo	6.829	2.580	2.980	16.360	0.974*	0.799*
10mo	6.852	2.577	3.010	16.400	0.974*	0.799*
11mo	6.876	2.566	3.020	16.390	0.974*	0.799*
12mo	6.922	2.510	3.110	15.810	0.972*	0.795*
24mo	7.130	2.442	3.660	15.640	0.978*	0.815*
36mo	7.282	2.374	3.870	15.560	0.979*	0.829*
48mo	7.401	2.343	3.970	15.820	0.980*	0.835*
60mo	7.464	2.319	3.980	15.000	0.982*	0.847*
120mo	7.535	2.268	4.110	15.210	0.984*	0.852*

Note: Individual significance at the 5% level is denoted by a superscript *

2.B.3 Out-of-Sample Analysis

In this section I perform a small out-of-sample study by considering three episodes. These episodes are of economic interest due to NBER-dated US recessions which have occurred during the periods December, 1969 to November, 1970; January to July, 1980; and July, 1990 to March, 1991. I estimate

TABLE 2.21. Results for Alternative Single-Variable Factor Specifications

	$\widehat{\beta}$	[s.e.]	$\widehat{\sigma}$	[s.e.]	MAE
$L:CPI$	7.292	[0.061]	82.213	[10.585]	1.939
$S:FF$	−2.356	[0.275]	0.441	[0.030]	1.560
$C:M1$	−0.826	[0.119]	−46.889	[6.525]	
$L:PCE$	7.160	[0.063]	139.993	[14.686]	1.911
$S:FF$	−2.577	[0.252]	0.474	[0.026]	1.492
$C:M1$	−0.857	[0.111]	−42.270	[5.930]	
$L:PPI1$	7.595	[0.052]	66.639	[6.849]	1.981
$S:FF$	−2.200	[0.300]	0.420	[0.032]	1.690
$C:M1$	−0.847	[0.127]	−41.326	[6.838]	
$L:PPI2$	7.775	[0.056]	31.179	[4.215]	1.993
$S:FF$	−2.284	[0.300]	0.434	[0.031]	1.709
$C:M1$	−0.875	[0.127]	−38.170	[6.704]	
$L:PCE$	7.356	[0.046]	115.975	[8.269]	1.890
$S:FF$	−2.751	[0.241]	0.496	[0.024]	1.433
$C:FFD$	−1.026	[0.103]	1.472	[0.394]	
$L:PCE$	7.088	[0.047]	126.068	[8.890]	<u>1.848</u>
$S:FF$	−3.245	[0.210]	0.567	[0.019]	<u>1.344</u>
$C:DEBT$	−0.728	[0.103]	−46.227	[5.782]	
$L:PCE$	5.512	[0.186]	510.668	[47.895]	1.946
$S:NONBR$	0.748	[0.309]	−13.162	[4.617]	1.584
$C:DEBT$	−1.264	[0.187]	33.981	[17.389]	
$L:PCE$	5.648	[0.126]	489.530	[24.113]	1.989
$S:M1$	1.162	[0.312]	−97.129	[18.479]	1.696
$C:DEBT$	−1.449	[0.185]	60.586	[17.319]	

Note: Standard errors inside squared brackets. Non-significant estimates at the 5% significance level are marked with ∅. The underlined MAE values are the smallest ones in the Table.

the DL, SV and MV specifications for three subsamples of the data, all of which starting from June, 1964. The first ends in December, 1969, the second in December, 1979, and the third in December, 1989.

The results reported in Table 2.25 show the overall out-of-sample behaviour of the macro-based specifications tend to outperform the benchmark, although the DL model tends to perform better in one-month ahead forecasts in two out of the three episodes considered. Although the MV specification performs better for the first episode, the more parsimonious SV specification seems to be doing a very good job for the second and third episodes considered.

TABLE 2.22. Results for Alternative Multi-Variable Factor Specifications

Panel A	$\widehat{\beta}$	[s.e.]	$\widehat{\sigma}$	[s.e.]	MAE
$L: PCE$	7.237	[0.044]	130.721	[9.040]	
$S: FF$	−3.104	[0.213]	0.553	[0.019]	1.889
$S: NONBR$	−0.817	[0.101]	−8.238	[2.179]	1.427
$C: DEBT$			−33.551	[5.243]	
$L: CPI$	6.922	[0.048]	177.428	[9.190]	
$S: FF$	−2.680	[0.228]	0.494	[0.022]	1.960
$S: NONBR$	−0.724	[0.110]	−10.821	[1.786]	1.624
$C: DEBT$			−45.088	[6.002]	
$L: PCE$			$-23.681^{\varnothing}$	[18.731]	
$L: CPI$	7.058	[0.044]	162.999	[16.663]	1.934
$S: FF$	−2.906	[0.217]	0.523	[0.021]	1.547
$S: NONBR$	−0.814	[0.103]	−9.357	[2.037]	
$C: DEBT$			−31.819	[5.356]	
$L: PCE$			132.707	[9.378]	
$S: FF$	7.243	[0.046]	0.553	[0.019]	1.892
$S: NONBR$	−3.125	[0.214]	−5.391	[2.400]	1.436
$C: DEBT$	−0.749	[0.104]	−41.542	[5.912]	
$C: FFD$			1.666	[0.398]	
$L: PCE$			133.358	[10.116]	
$S: FF$	7.214	[0.047]	0.492	[0.027]	1.924
$S: NONBR$	−4.247	[0.242]	−6.997	[2.567]	1.510
$S: HELP$	−0.700	[0.110]	0.021	[0.003]	
$C: DEBT$			−53.334	[6.599]	
$L: PCE$	7.059	[0.044]	161.146	[7.562]	
$S: FF$	−4.486	[0.244]	0.472	[0.029]	1.903
$S: HELP$	−0.674	[0.112]	0.025	[0.003]	1.458
$C: DEBT$			−54.253	[6.839]	

Note: Standard errors inside squared brackets. Non-significant estimates at the 5% significance level are marked with $\varnothing$. The underlined MAE values are the smallest ones in the Table.

The results reported in Tables 2.26-27 confirm the view that macro-based specifications tend to outperform the benchmark DL model. For Table 2.26, which reports the results for year 1970, this dominance occurs for 12-13 of the 17 maturities considered. Most notably, the cumulative average MAE across maturities of the DL specification for the nine-month horizon is 50% larger than the ones of the macro-based specifications. Overall, the fitting of the macro-based specifications is much better than the DL one for the shorter half of the yield curve by significant orders of magnitude, although this dominance is reversed in favour of the DL specification when it comes to the longer end. A candidate explanation for this fact is the higher persistence and lower volatility of longer yields, as reported in Table 2.20.

TABLE 2.23. Results for Alternative Multi-Variable Factor Specifications cont'd

Panel B	$\widehat{\beta}$	[s.e.]	$\widehat{\sigma}$	[s.e.]	avg($\widehat{\rho}_t$)	MAE
$L: PCE$			54.966	[12.670]		
$L: PPI1$	7.315	[0.054]	30.765	[4.586]		1.866
$S: FF$	−3.204	[0.220]	0.567	[0.021]	0.954	
$S: NONBR$	−0.732	[0.104]	−5.093	[2.374]		1.369
$C: DEBT$			−49.818	[5.753]		
$L: PCE$			237.420	[14.652]		
$L: PPI2$	6.925	[0.055]	−18.473	[3.607]		1.887
$S: FF$	−2.996	[0.209]	0.540	[0.019]	0.957	
$S: NONBR$	−0.811	[0.101]	−8.259	[2.342]		1.431
$C: DEBT$			−34.563	[5.454]		
$L: PCE$			170.695	[20.213]		
$L: PPI3$	7.106	[0.073]	$-1.926^{\varnothing}$	[4.842]		1.892
$S: FF$	−2.839	[0.232]	0.519	[0.022]	0.961	
$S: NONBR$	−0.716	[0.110]	−7.120	[2.707]		1.436
$C: DEBT$			−51.385	[6.372]		
$L: PCE$			58.423	[11.813]		
$L: PPI1$	7.395	[0.054]	16.708	[4.611]		
$S: FF$	−3.214	[0.222]	0.562	[0.021]	0.953	1.866
$S: NONBR$	−0.618	[0.108]	$2.252^{\varnothing}$	[2.593]		1.369
$C: DEBT$			−62.828	[6.214]		
$C: FFD$			2.962	[0.393]		
$L: PCE$	6.885	[0.042]	181.628	[8.315]		
$L: PPI1$	−3.172	[0.206]	21.672	[4.533]	0.957	1.863
$S: FF$	−0.669	[0.104]	0.558	[0.018]		1.356
$C: DEBT$			−55.522	[6.091]		

Note: Standard errors inside squared brackets. Non-significant estimates at the 5% significance level are marked with $\varnothing$. The underlined MAE values are the smallest ones in the Table.

TABLE 2.24. Goodness-of-Fit of Alternative Models

Model	Average R^2	Median R^2	Average MAE	Median MAE
SV	0.704	0.878	1.848	1.344
MV	0.705	0.884	1.866	1.369

The results reported in Table 2.27 are qualitatively similar to the ones of Table 2.26. However, the goodness-of-fit for all specifications tends to be worse than before, probably due to the change in the way monetary policy was being conducted during that period. Finally, Tables 2.26-27 shows a clear dominance of the macro-based specifications and, in particular, of the parsimonious SV specification over the competing alternatives.

TABLE 2.25. Average MAEs Period-by-Period

	1970			1980			1990		
	DL	**SV**	**MV**	**DL**	**SV**	**MV**	**DL**	**SV**	**MV**
1st month	<u>0.15</u>	1.51	1.63	<u>1.14</u>	1.20	1.28	<u>0.52</u>	0.63	1.03
2nd month	1.18	0.50	<u>0.26</u>	3.79	<u>3.20</u>	3.23	0.73	<u>0.10</u>	0.52
3rd month	1.56	<u>0.80</u>	0.86	4.68	<u>2.29</u>	2.34	0.95	<u>0.31</u>	0.85
4th month	0.98	0.86	<u>0.71</u>	0.93	0.87	<u>0.80</u>	1.20	<u>0.83</u>	1.14
5th month	<u>1.13</u>	1.25	<u>1.13</u>	0.91	<u>0.80</u>	0.81	0.91	<u>0.54</u>	0.67
6th month	1.50	<u>1.15</u>	1.23	0.98	0.80	<u>0.77</u>	0.84	<u>0.11</u>	0.45
7th month	1.86	0.72	<u>0.64</u>	<u>0.34</u>	1.11	1.10	0.61	<u>0.30</u>	0.38
8th month	1.97	1.16	<u>1.10</u>	<u>1.50</u>	2.24	2.22	0.70	0.52	<u>0.50</u>
9th month	2.34	<u>0.59</u>	0.89	2.61	2.53	<u>2.46</u>	0.57	<u>0.56</u>	0.68
10th month	2.67	0.42	<u>0.34</u>	3.65	<u>3.20</u>	3.25	<u>0.38</u>	0.66	0.75
11th month	3.86	0.50	<u>0.44</u>	4.96	<u>3.36</u>	3.49	0.21	0.22	<u>0.14</u>
12th month	4.07	<u>0.59</u>	0.65	4.41	<u>2.14</u>	2.36	0.34	<u>0.21</u>	0.40

Note: The underlined quantities are the smaller values for a given time period and episode.

TABLE 2.26. MAEs of Alternative Specifications

Panel A	**3rd month of event**			**6th month of event**			**9th month of event**		
1970	**DL**	**SV**	**MV**	**DL**	**SV**	**MV**	**DL**	**SV**	**MV**
1mo	1.09	0.54	0.70	1.47	0.33	0.42	1.82	0.32	0.39
2mo	0.87	0.57	0.56	1.31	0.48	0.51	1.69	0.43	0.42
3mo	0.92	0.69	0.69	1.18	0.61	0.56	1.58	0.50	0.49
4mo	0.92	0.69	0.68	1.14	0.68	0.63	1.55	0.55	0.54
5mo	0.94	0.70	0.65	1.14	0.71	0.65	1.50	0.64	0.62
6mo	0.95	0.74	0.68	1.18	0.72	0.65	1.51	0.68	0.65
7mo	0.95	0.79	0.74	1.11	0.83	0.77	1.45	0.78	0.76
8mo	0.97	0.83	0.77	1.07	0.92	0.85	1.39	0.88	0.86
9mo	1.03	0.81	0.76	1.08	0.96	0.89	1.39	0.91	0.89
10mo	1.06	0.82	0.77	1.10	0.97	0.90	1.43	0.91	0.89
11mo	1.10	0.82	0.77	1.13	0.98	0.92	1.45	0.92	0.90
12mo	1.11	0.85	0.80	1.13	1.01	0.95	1.46	0.94	0.92
24mo	1.11	1.13	1.10	1.09	1.32	1.26	1.41	1.22	1.20
36mo	1.06	1.29	1.26	1.02	1.47	1.43	1.29	1.41	1.40
48mo	0.83	1.55	1.53	0.83	1.69	1.65	1.09	1.64	1.63
60mo	0.77	1.61	1.59	0.79	1.73	1.69	1.05	1.66	1.66
120mo	0.79	1.52	1.50	0.65	1.81	1.77	0.88	1.75	1.74
Avg	0.97	0.94	0.91	1.08	1.01	0.97	1.41	0.95	0.94
Panel B	**3rd month of event**			**6th month of event**			**9th month of event**		
1980	**DL**	**SV**	**MV**	**DL**	**SV**	**MV**	**DL**	**SV**	**MV**
1mo	3.15	1.37	1.51	2.27	1.55	1.57	2.00	1.36	1.38
2mo	3.56	1.86	1.99	2.39	1.41	1.43	1.98	1.32	1.33
3mo	3.69	2.06	2.19	2.44	1.37	1.37	2.13	1.41	1.41
4mo	3.81	2.26	2.38	2.51	1.47	1.46	2.18	1.53	1.52
5mo	3.88	2.40	2.51	2.51	1.49	1.49	2.19	1.58	1.57
6mo	3.85	2.44	2.54	2.48	1.48	1.48	2.17	1.59	1.58
7mo	3.78	2.44	2.53	2.44	1.45	1.44	2.16	1.58	1.57
8mo	3.63	2.35	2.44	2.37	1.36	1.41	2.11	1.54	1.57
9mo	3.76	2.55	2.63	2.46	1.44	1.54	2.17	1.62	1.67
10mo	3.66	2.50	2.58	2.37	1.40	1.49	2.11	1.60	1.65
11mo	3.69	2.59	2.66	2.37	1.44	1.53	2.10	1.63	1.68
12mo	3.35	2.31	2.37	2.16	1.32	1.31	1.96	1.52	1.50
24mo	2.86	2.34	2.34	1.84	1.48	1.50	1.67	1.74	1.74
36mo	2.33	2.14	2.11	1.42	1.59	1.58	1.40	1.88	1.85
48mo	1.98	2.02	1.96	1.16	1.67	1.64	1.23	1.97	1.93
60mo	1.87	2.07	1.99	1.05	1.90	1.86	1.20	2.20	2.14
120mo	1.66	2.22	2.09	0.97	2.13	2.05	1.11	2.35	2.26
Avg	3.21	2.23	2.28	2.07	1.53	1.54	1.88	1.67	1.67

Note: The underlined quantities are the smaller values for a given time period and episode.

TABLE 2.27. MAEs of Alternative Specifications cont'd

Panel C	**3rd month of event**			**6th month of event**			**9th month of event**		
1990	**DL**	**SV**	**MV**	**DL**	**SV**	**MV**	**DL**	**SV**	**MV**
1mo	0.84	0.55	1.00	0.88	0.42	0.77	0.82	0.47	0.73
2mo	0.92	0.43	0.89	1.03	0.48	0.84	0.96	0.50	0.76
3mo	0.93	0.35	0.81	1.05	0.47	0.83	0.96	0.49	0.75
4mo	0.84	0.37	0.83	0.98	0.46	0.82	0.89	0.50	0.76
5mo	0.83	0.32	0.78	0.95	0.43	0.79	0.87	0.45	0.71
6mo	0.78	0.31	0.77	0.91	0.41	0.77	0.81	0.44	0.71
7mo	0.72	0.29	0.75	0.83	0.39	0.72	0.73	0.44	0.69
8mo	0.67	0.31	0.77	0.78	0.34	0.71	0.69	0.39	0.66
9mo	0.66	0.29	0.75	0.78	0.35	0.71	0.69	0.40	0.66
10mo	0.62	0.28	0.71	0.76	0.37	0.70	0.67	0.40	0.65
11mo	0.57	0.34	0.73	0.72	0.40	0.70	0.63	0.43	0.66
12mo	0.67	0.31	0.77	0.81	0.41	0.77	0.69	0.45	0.72
24mo	0.57	0.36	0.82	0.72	0.45	0.78	0.61	0.45	0.69
36mo	0.59	0.37	0.80	0.73	0.45	0.78	0.64	0.43	0.67
48mo	0.66	0.34	0.81	0.81	0.44	0.82	0.75	0.39	0.66
60mo	0.72	0.33	0.80	0.86	0.43	0.83	0.82	0.37	0.65
120mo	0.91	0.32	0.79	1.01	0.43	0.83	1.03	0.34	0.63
Avg	0.74	0.35	0.80	0.86	0.42	0.78	0.78	0.43	0.69

Note: The underlined quantities are the smaller values for a given time period and episode.

References

[1] Ang, A., and M. Piazzesi (2003), A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables, *Journal of Monetary Economics* 50 (4), 745-787.

[2] Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.

[3] Attanasio, O. P., and A. Vissing-Jorgensen (2003), Stock market participation, intertemporal substitution and risk aversion, *American Economic Review* 93, 383-391.

[4] Balduzzi, P., S. R. Das, S. Foresi and R. Sundaram (1996), A Simple Approach to Three-Factor Affine Models of the Term Structure, *Journal of Fixed Income* 6, 43-53.

[5] Bank for International Settlements (2000), Stress Testing by Large Financial Institutions: Current Practice and Aggregation Issues. Report by the Committee on the Global Financial System, April.

[6] Bansal, R., and A. Yaron (2004), Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481-1509.

[7] Bansal R., D. Kiku, and A. Yaron (2007), Risks for the long run: Estimation and inference, Working Paper.

[8] Bernanke, B. S, and A. S. Blinder (1992), The Federal Funds Rate and the Channels of Monetary Transmission, *American Economic Review* 82(4), 901-921.

[9] Bernanke, B. S., and I. Mihov (1998), Measuring Monetary Policy, *The Quarterly Journal of Economics* 113(3), 869-902.

[10] Bikbov, R., and M. Chernov (2005), No-Arbitrage Macroeconomic Determinants of the Yield Curve, Columbia Business School working paper.

[11] Bliss, R. R. (1997), Testing Term Structure Estimation Methods, Advances in Futures and Options Research, Vol. 9.

[12] Brett, C., J. Pinkse, and M. Slade (2002), Spatial Price Competition: a Semiparametric Approach, *Econometrica*, 70-3, 1111-1153.

[13] Brousseau, V. (2002), The Functional Form of Yield Curves, Working Paper 148, European Central Bank.

[14] Campbell, J. Y. (1986), Bond and stock returns in a simple exchange model, *Quarterly Journal of Economics* 101, 785-804.

[15] Campbell, J. Y. (1993), Intertemporal asset pricing without consumption data, *American Economic Review* 83, 487-512.

[16] Campbell, J. Y. (1996), Understanding risk and return, *Journal of Political Economy* 104, 298-345.

[17] Campbell, J. Y. (1999), Asset prices, consumption and the business cycle, in John B. Taylor, and Michael Woodford, eds.: *Handbook of Macroeconomics* (Elsevier Science, North-Holland, Amsterdam).

[18] Campbell, J. Y. and J. Cochrane (1999), By force of habit: a consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205-251.

[19] Case, A. (1991), Spatial Patterns in Household Demand, *Econometrica* 59 (4), 953-965.

[20] Christiano, L., Eichenbaum, M. and C. Evans (1997), Sticky Price and Limited Participation Models: A Comparison, *European Economic Review 41* (6), 1201-1249.

[21] Christiano, L., Eichenbaum, M. and C. Evans (1999), 'Monetary Policy Shocks: What Have We Learned and to What End?', *Handbook of Macroeconomics*, ed.s Michael Woodford and John Taylor, North Holland.

[22] Clarida, R., J. Galí and M. Gertler. (2000), Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory, *The Quarterly Journal of Economics*, 115, 147-180.

[23] Cochrane, J. H. (1994), Shocks, *Carnegie-Rochester Conference Series on Public Policy* 41, 295-364.

[24] Conley, T. G. (1999), GMM Estimation with Cross Sectional Dependence, *Journal of Econometrics* 92 (1), 1-45.

[25] Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985), A Theory of the Term Structure of Interest Rates, *Econometrica* 53, 385-407.

[26] Dahlquist, M., and L. E. O. Svensson (1996), Estimating the Term Structure of Interest Rates for Monetary Policy Analysis, *Scandinavian Journal of Economics*, 98(2), 163-183.

[27] Dai, Q., and T. Philippon (2005), Fiscal Policy and the Term Structure of Interest Rates, working paper.

[28] Dai, Q., and K. J. Singleton (2000), Specification Analysis of Affine Term Structure Models, *Journal of Finance*, 55 (5), 1943-1978.

[29] de Jong, F. (2000), Time series and cross section information in affine term structure models. *Journal of Business and Economic Statistics* 18, 300–314.

[30] De Long, J. B., and L. H. Summers (1991), Equipment Investment and Economic Growth, *The Quarterly Journal of Economics* 106 (2), 445-502.

[31] Dewachter, H. and Lyrio, M. (2006), Macro Factors and the Term Structure of Interest Rates, Journal of Money, Credit, and Banking, 38 (1), 119-140.

[32] Diebold, F.X. and Li, C. (2006), Forecasting the Term Structure of Government Bond Yields, *Journal of Econometrics* 130, 337-364.

[33] Diebold, Francis X., Li, Calin & Yue, Vivian (2005), Modelling Term Structures of Global Bond Yields, Univ. of Pennsylvania working paper.

[34] Diebold, F.X., Piazzesi, M. and Rudebusch, G.D. (2005), Modeling Bond Yields in Finance and Macroeconomics, *American Economic Review*, 95, 415-420.

[35] Diebold, F. X., G. D. Rudebusch, and S. B. Aruoba (2006), The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach, *Journal of Econometrics* 131, 309-338.

[36] Duffee, G. R. (2006), Term structure estimation without using latent factors, *Journal of Financial Economics*, Journal of Financial Economics 79, 507-536.

[37] Duffie, D. and R. Kan (1996) A yield-factor model of interest rates, *Mathematical Finance* 6, 379-406.

[38] Duffie, D., M. Schroder, and C. Skiadas (1997), A term structure model with preferences for the timing of resolution of uncertainty, *Economic Theory* 9, 3-22.

[39] Dunn, K., and K. Singleton (1986), Modeling the term structure of interest rates under non-separable utility and durability of goods, *Journal of Financial Economics* 17, 27-55.

[40] Epstein, L. G., and S. E. Zin (1989), Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937-968.

[41] Estrella, A., and G. A. Hardouvelis (1991), The Term Structure as a Predictor of Real Economic Activity, *Journal of Finance*, 46, 555-576.

[42] Estrella, A. and F. S. Mishkin (1996), The Yield Curve as a Predictor of US Recessions. Federal Reserve Bank of New York, *Current Issues in Economics and Finance.*

[43] Evans, C. L. and D. A. Marshall (1996), Economic Determinants of the Term Structure of Nominal Interest Rates. Federal Reserve Bank of Chicago working paper.

[44] Evans, C. L. and D. A. Marshall (1998), Monetary Policy and the Term Structure of Nominal Interest Rates: Evidence and Theory, *Carnegie-Rochester Conference Series on Public Policy*, 49, 53-111.

[45] Fisher, M., D. Nychka and D. Zervos (1995), Fitting the Term Structure of Interest Rates with Smoothing Splines, Federal Reserve Bank Finance and Economics Discussion Paper, 95-1.

[46] Fuhrer, J. (1996), "Monetary Policy Shifts and Long-Term Interest Rates," *The Quarterly Journal of Economics*, 1183-1209.

[47] Fuhrer, J., and G. Moore (1995), "Monetary Policy Trade-Offs and the Correlation Between Nominal Interest Rates and Real Output," *American Economic Review* 85, 219-239.

[48] Gallmeyer, M. F., B. Hollifield, and S. E. Zin (2005), Taylor Rules, McCallum Rules, and the Term Structure of Interest Rates, NBER Working Paper 11276.

[49] Hall, R. E. (1988), Intertemporal substitution in consumption, *Journal of Political Economy* 96, 339-357.

[50] Hördahl, P., Tristani, O. and Vestin, D. (2006), A Joint Econometric Model of Macroeconomic and Term Structure Dynamics, Journal of Econometrics, 131(1-2): 405-44.

[51] Hull, J., and A. White (1990), Pricing Interest-Rate-Derivative Securities, *The Review of Financial Studies* 3, (4), 573-592.

[52] Judd, J. P., and G. D. Rudebusch (1998), Taylor's Rule and the Fed: 1970-1997, *Economic Review*, Federal Reserve Bank of San Francisco 3, 3-16.

[53] Kelejian, H. H., and I. R. Prucha (1998), A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances, *Journal of Real Estate Finance and Economics* 17 (1), 99-121.

[54] Kelejian, H. H., and I. R. Prucha (1999), A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model, *International Economic Review* 40 (2), 509-533.

[55] Kelejian, H. H., and I. R. Prucha (2004), Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations, *Journal of Econometrics* 118, 27-50.

[56] Kelejian, H. H., I. R. Prucha, and Y. Yuzefovich (2004), Instrumental Variable Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances: Large and Small Sample Results, Advances in Econometrics, Vol. 18: Spatial Spatiotemporal Econometrics, (K. Pace and J. LeSage eds).

[57] Kugler, P. (1997), Central Bank Policy Reaction and the Expectations Hypothesis of the Term Structure, *International Journal of Financial Economics* 2, 217-224.

[58] Leeper, E. M., C. A. Sims, and T. Zha (1996), What Does Monetary Policy Do?, *Brookings Papers on Economic Activity* 2, 1-78.

[59] Longstaff, F., and E. Schwartz (1992), Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model, *Journal of Finance*, 47, 1259-1282.

[60] McCallum, B. T. (1994), Monetary Policy and the Term Structure of Interest Rates, NBER Working Paper No. 4938.

[61] McCulloch, J. (1971), Measuring the Term Structure of Interest Rates, *Journal of Business*, 44, 19-31.

[62] McCulloch, J. (1975), The Tax-Adjusted Yield Curve, *Journal of Finance*, 30, 811-830.

[63] Mehra, R. and E. C. Prescott (1985), The equity premium: A puzzle, *Journal of Monetary Economics* 15, 145-161.

[64] Nelson, C. R., and A. F. Siegel (1997), Parsimonious Modelling of Yield Curves, *Journal of Business*, 60, 473-489.

[65] Newey, W. K., and K. D. West, (1987), A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55 (3), 703-708.

[66] Ord, J. (1975), Estimation Methods for Models of Spatial Interaction, *Journal of the American Statistical Association* 70, 120-126.

[67] Piazzesi, M. (2005), Bond yields and the Federal Reserve, *Journal of Political Economy* 113 (2), 311-344.

[68] Piazzesi, M. and M. Schneider (2006), Equilibrium yield curves, in Michael Woodford Daron Acemoglu, and Kenneth Rogo , eds.: *NBER Macroeconomics Annual*: 2006 (MIT Press, Cambridge).

[69] Pinkse, J., L. Shen and M. Slade (2006), Dynamic Spatial Probit with Fixed Effects using One Step GMM: an Application to Mine Operating Decisions, Spatial Economic Analysis, Vol. 1, No. 1, 53-90.

[70] Rotemberg, J., and M. Woodford (1997), An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy, *NBER Macroeconomics Annual.*

[71] Rudebusch, G.D. (1995), "Federal Reserve Interest Rate Targeting, Rational Expectations, and the Term Structure," *Journal of Monetary Economics* 35, 245-274.

[72] Rudebusch, G. D. (1998), Do Measures of Monetary Policy in a VAR Make Sense?, *International Economic Review* 39, November 1998, 907-931.

[73] Rudebusch, G.D. (2002), "Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia," *Journal of Monetary Economics* 49, 1161-1187.

[74] Schaefer, S. (1981), Measuring a Tax-Specific Term Structure of Interest Rates in the Market for British Government Securities, *Economic Journal*, 91, 415-431.

[75] Shea, G. S (1985), Interest Rate Term Structure Estimation with Exponential Splines: A Note, *Journal of Finance*, 40, 319-325.

[76] Stock, J. H., and M. W. Watson (2000), Forecasting Output and Inflation: The Role of Asset Prices, *Journal of Economic Literature* XLI, 788-829.

[77] Strongin, S. (1995), The identification of monetary policy disturbances: Explaining the liquidity puzzle, *Journal of Monetary Economics* 35, 463–497.

[78] Taylor, J. B. (1993), Discretion versus Policy Rules in Practice, *Carnegie-Rochester Conference Series on Public Policy* 39, 195-214.

[79] Townsend, R. M. (1994), Risk and Insurance in Village India, *Econometrica* 62 (3), 539-591.

[80] Vasicek, O. (1977), An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, 5, 177-188.

[81] Vasicek, O., and H. Fong (1982), Term Structure Modelling Using Exponential Splines, *Journal of Finance*, 37, 339-341.

[82] Wachter, J. A. (2006), A consumption-based model of the term structure of interest rates, *Journal of Financial Economics* 79, 365-399.

[83] Weil, Ph. (1989), The equity premium puzzle and the risk-free rate puzzle, *Journal of Monetary Economics* 24, 401-421.

[84] White, H. (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica* 48 (4), 817-838.

[85] Woodford, M. (2001), The Taylor Rule and Optimal Monetary Policy, *American Economic Review*, 91 (2), 232-237.

[86] Wu, T. (2002), "Monetary Policy and the Slope Factors in Empirical Term Structure Estimations," Federal Reserve Bank of San Francisco Working Paper 02-07.

3

Estimating The 'Coordinated Effects' of Mergers

3.1 Introduction

This paper empirically evaluates whether firms face incentives to tacitly collude in differentiated product[1] markets and whether these incentives change following a merger. Tacit collusion may appear when the same firms repeatedly interact in the same markets, since they may have an incentive to set high prices because they expect that by not doing so, their competitors will lower their own prices in the future. Mergers could facilitate - or enhance - the likelihood of collusion by reducing the number of firms that need to interact in a given market, thus facilitating the convergence of a focal point. Both academics[2] and competition agencies[3] have recognized that mergers can potentially facilitate tacit collusion by changing the incentives firms face when setting prices. Recent examples where antitrust authorities have invoked the theory of coordinated effects when attempting to block mergers include the Nestlé-Perrier, Kali and Salz , Gencor-Lenrho, Airtours and Sony-BMG cases in the EU jurisdiction, Safeway in the United Kingdom and ATP, Arch Coal, Cruises and Hospital Corporation in the US.

---

[1] In contrast with homogeneous products, differentiated products are not identical to other products in the same industry. According to the *principle of differentiation*, in the words of Tirole (1988, p.278): "[uniproduct] firms generally do not want to locate at the same place in the product space . The reason is simply the Bertrand paradox: Two firms producing perfect substitutes face unbridled price competition (at least in a static framework.) In contrast, product differentiation establishes clienteles ("market niches", in the business terminology) and allows firms to to enjoy some market power over these clienteles. Thus, firms usually wish to differentiate themselves from other firms." In this paper I deal with multiproduct firms, so that the above holds for products, instead of firms.

[2] Chamberlin (1929) argued this point informally, while Stigler (1964) and Friedman (1971) formalized this intuition in the theory of repeated games. Following Stigler (1964), in order to sustain collusion firms must be able to (i) come to an agreement (which can be difficult when products are complex and differentiated), (ii) monitor each others' behaviour (in order to detect undercutting) and, of course, (iii) enforce collusive behavior collectively by punishing the cheating firms. See Aumann (1986, 1989) and Mertens (1987) for surveys.

[3] See for example the EU, UK or US Horizontal Merger Guidelines.

In this paper, I take the most basic textbook model of tacit collusion[4] to data in order to evaluate the way in which an actual merger impacts on firms incentives and ability to tacitly collude. In particular, I follow and expand Friedman's (1971) model of tacit collusion to data and use it to inform a merger investigation. In doing so, I generalize the standard textbook model of tacit collusion in a number of important ways. Specifically, I allow for product differentiation, multi-product firms, competitive fringes and multi-market contact. However, I recognize that even so, Friedman's model makes a number of important and certainly unrealistic assumptions, most notably that firms benefit from complete information. Despite that fact, since Friedman's model forms the basis of analysis in both textbooks and, at least in significant part, law - via the 'Airtours' tests introduced by the Court of First Instance (CFI) in the Airtours judgment in Europe, I nonetheless consider that empirically exploring the economics of this benchmark model is a useful contribution. Future empirical research will, of course, need to explore coordinated effects merger simulation in incomplete information settings.

The intuition behind the coordinated effects of mergers can most easily be developed using Friedman's (1971) result that firms will be willing to coordinate whenever their share of monopoly profits is greater than their returns to not-coordinating. Specifically, if N is the number of firms, δ is the discount rate and firms follow grim strategies, then essentially every modern industrial organization textbook establishes the relationship that collusion will be sustainable with N symmetric firms if $V^{\text{Collusion}} = \pi^{\text{Monopoly}}/N(1-\delta) > \pi^{\text{Defection}} + \delta\pi^{\text{Nash}}/(1-\delta) = V^{\text{Defection}}$, where π^{Nash} represents the profits to a firm under static Nash behavior (with $\pi^{\text{Nash}} = 0$ for the case of a homogeneous products Bertrand competition stage game) and $\pi^{\text{Defection}}$ represents the payoff to a cheating firm who is assumed to receive it for one period. Since each firm's payoff to collusion is declining in N, this relationship suggests that generically mergers (reductions in N) will make collusion easier to sustain since each firms share of collusive profits increases. In what follows, I show that this intuitive result is, in large part, misleading and, in fact, the elementary version of this theory will generally predict the opposite - that mergers will make tacit collusion harder to sustain, not easier. I establish this result in Proposition 1, my core theoretical contribution, and discus the intuition for my results there.

This research builds most directly on three significant literatures. First, I build on the empirical literature on the unilateral effects of mergers, and in particular merger simulation. This literature has evolved over the last two decades following the work of Davidson and Deneckere (1985), Farrell and Shapiro (1990), Baker and Bresnahan (1999), Hausman et al (1994) and Nevo (2000). The Bertrand

[4] See, for instance, Tirole (1988, Section 6.3)

differentiated product model wherein firms compete in prices and market outcomes are characterized as the Nash equilibrium of the game provides the benchmark model in this literature.[5]

Second, I build on the empirical literature attempting to evaluate the conduct of firms using game-theoretic pricing models. Specifically, authors such as Gollop and Roberts (1979), Roberts (1983), Suslow (1986), Bresnahan (1982, 1987), Nevo (2001), Slade (2002) and Salvo (2004) have attempted to evaluate whether observed equilibrium prices are more consistent with collusive or Nash equilibrium pricing. To do so, these papers compare the models of perfect collusion and static Nash equilibrium prices and evaluate which model better predicts observed market outcomes. An important contribution to this literature was provided by Corts (1999) who critiqued this empirical literature on the grounds that the mapping between the empirical test of collusive behavior and the underlying theory of collusion was incomplete. One interpretation of this paper's contribution is that it makes a modest step towards addressing that criticism seriously by taking an actual model of tacit collusion to data.

Third, this paper is also related to both the older and the more recent, and important, theoretical contributions on the coordinated effects of mergers. Specifically, I will follow Friedman (1971) closely, but in studying asymmetric contexts this work is related to Compte, Jenny and Rey (2002), Vasconcelos (2005) and Kühn (2004) have recently studied collusion under asymmetric market structures. While Compte et al (2002) examine coordinated effects in the context of a Bertrand-Edgeworth homogeneous goods model with capacity constraints and calibrate their model with the data from the Nestlé-Perrier case, Kühn (2004) is the first paper to study mergers in differentiated product markets. I will abstract from capacity constraints, since I have no data about them. However, capacity can play an important role in tacit collusion. For example, Lambson (1996) argues that slight asymmetries in capacities can reduce the danger of tacit collusion.

Kühn (2004) studies a richer context than I do, in that he studies collusion under imperfect information, following, for example, Abreu, Pearce and Stachetti (1990). However, I note that while Kühn's information context is richer and more realistic, he must make other assumptions that make his model difficult to imagine using as a basis for informing empirical work - for instance, he considers only the case where the price for every good sold by each firm is the same.

The paper is organized as follows. Section 2 I describe the network server industry. Section 3 sets the framework to evaluate the incentives to collude. In Section 4 I describe the data, whereas Section 5 discusses the implementation of the model. Section 6 describes the results.

[5]As an aside note that, in fact, this model has been used both to evaluate 'unilateral' effects of horizontal mergers and also as the basis for a literature evaluating vertical integration and restraints. (See Brenkers and Verboven (2006), Bonnet and Dubois (2008), Villas-Boas (2007) and most recently DG-COMP's decision in the TomTom-TeleAtlas case (2008).

3.2 The Network Server Industry

3.2.1 Network Servers

The late 1980s and early 1990s saw a paradigm shift in what concerns computer architecture as the mainframe-based system was superseded by the 'client-server' computer architecture that is the dominant nowadays.[6] As a result, instead of centralized computing power and users interacting via 'dumb terminals', processing power became more decentralized and flexible, distributed between PCs with their own operating systems, and with increasingly powerful machines - servers - linking these PCs altogether through networks.

Servers are heavy-duty computers that can both provide resources (such as software) and share resources (such as files) with other computers - the clients - on a computer network.[7] Very much like desktops, servers consist of both hardware, such as the processors, memory and the hard-disk drive, as well as software, such as the operating system and specific applications. Servers are, however, more powerful and expensive machines than desktops costing anywhere from a thousand to several million dollars. With many users simultaneously using their services, they are typically deployed to run round-the-clock to tackle critical computing tasks such as keeping track of a retail chain's sales, a customer database, logging phone calls and reconciling stock trade transactions, thus requiring better resistance to crashes, hacker attacks and other faults. Servers range from high-end Unix servers, with numerous processors and multimillion-dollar price tags, to comparatively inexpensive Intel-based machines running a Linux or Microsoft Windows operating system used, for instance, to power small LANs and low-volume Internet sites.[8]

Servers are now widespread in nearly every sector of the economy. Although the server market is smaller in unit volume than the desktop one, its products command higher margins. Moreover, selling a server can spur related spending on services and other products. Gartner Group (2008) reports the 2007 server market revenue to be of the order of $55bn, amounting to the shipment of almost 9mn units. At the global level, the main players are IBM, HP-Compaq, Sun and Dell, which jointly command over 75% of the market revenues - see Table 3.1. Financial corporations and the communications sector are responsible for, respectively, 25% and 14% of the industry world revenues, whereas Government purchases amount to 11%.[9] Network servers have also been at the centre of a

[6] See Bresnahan and Greenstein (1996) for an economic analysis of the transition.
[7] For a basic discussion of servers see, for example, Sybex (2001).
[8] LAN (local area network) is a local computer network for communication between computers. In particular, a network connecting computers from different offices.
[9] Gartner Group (2007).

TABLE 3.1. Market Shares in the World Server Market

	1999		2001		2002		2007	
Firms	**Sales**	**Units**	**Sales**	**Units**	**Sales**	**Units**	**Sales**	**Units**
IBM	28.3	17.3	29.0	15.0	31.1	14.3	31.1	14.5
HP	14.4	12.3	26.6	9.7	25.2	30.1	28.3	29.8
Compaq	14.0	28.0	16.0	23.3	–	–	–	–
Sun	14.0	5.2	15.4	5.8	15.1	6.0	10.8	3.8
Dell	4.1	11.7	6.5	16.1	7.5	18.5	11.4	21.4
Others	25.2	24.9	6.5	30.1	21.1	31.1	13.9	27.1
Market Revenue	USD 48.5 bn		USD 47.0 bn		USD 43.1 bn		USD 54.8 bn	
Market Shipments	3.4 mn units		4.4 mn units		4.6 mn units		8.8 mn units	

Note: Data from Gartner Group (2000, 2002, 2003, 2008). Figures reported under sales and quantity are, respectively, revenue- and quantity-based market shares, illustrating the different presence of the players across market segments. Market shipments report the number of units sold in a given year worldwide.

number of recent competition policy debates, most notably the European Commission's decision in 2004 against Microsoft.[10]

3.2.2 Consolidation in the Network Server Industry

HP's acquisition of Compaq was the highest profile merger during a period of marked consolidation in the server industry. The second half of the 1990s was marked by a series of mergers and acquisitions in the computer industry and, in particular, in the network server industry - see Table 3.2. In particular, August 1997 and June 1998 saw Tandem and DEC being acquired by Compaq - the latter was the largest acquisition in computer industry at that time, creating the second largest computing company in the world in terms of revenue.[11] In June 1999, the Fujitsu-Siemens 50-50 joint venture was announced, with the aim of creating Europe's largest IT supplier and second PC vendor, offering a greater range of products by combining Fujitsu's consumer and professional LAN-based PCs with Siemens' servers and enterprise solution products.

Market consolidation since the mid-1990s may have resulted from the falling prices of servers. Market observers attribute falling prices to both demand shocks[12] and a process of commoditization. Evidence

[10] In March 2004, the European Commission found Microsoft guilty of violating the EU competition legislation and fined it in EUR 497mn, the biggest-ever fine in an EU competition case. The EC argued that Microsoft took advantage of its Windows virtual monopoly, unfairly leveraging its dominance over PC operating systems into other markets, especially the market for servers and media player software. In the media player front, Microsoft was accused of bundling by shipping Windows with its Windows Media Player, whereas in the server market the allegations were due to inter-operability issues ie. Microsoft was allegedly not allowing easy interaction between computer servers using a Windows OS and applications from vendors other than Microsoft itself.

[11] See http://web.archive.org/web/20050331034315/http://h18000.www1.hp.com/corporate/history.html for details of the Compaq history.

[12] The industry benefited from the wave of 'Millenium Bug' new equipment purchases that took place at the end of the 1990s, suggesting positive demand shocks in the late 1990's but that demand was temporary and the removal of this source of demand together with the availability of quality used equipment from bankrupt Internet companies in the early 2000s and uncertainties regarding the economic outlook led to a temporary slowdown in ICT spending and, in particular, a negative impact on the market

TABLE 3.2. Key Events in the Network Server Industry 1996-2002

	Date	Event
	1996	
	Febuary	Cray acquired by SGI (Silicon Graphics)
	February	Packard Bell acquired by NEC, *inter alia*
	July	Sun acquires Cray Business System Division from SGI
	1997	
23	June	Tandem acquired by Compaq
11	August	AST Research acquired by Samsung
	1998	
	June	DEC (Digital) acquired by Compaq
	1999	
	April	Eckhard Pfeiffer resigns as Compaq's CEO
17	June	Joint venture creates Fujitsu Siemens
	August	Data General announced being acquired by EMC
	September	Intergrah exits PC and server business
24	September	Sequent acquired by IBM
	2000	
2	March	SGI sells Cray to Tera Computer
4	April	Tera Computer renamed Cray Inc
	2001	
4	September	Announcement of HP and Compaq plans to merge
	September	Moody's downgrades HP debt
	September	Standard and Poors put HP on negative outlook
	December	HP-Compaq merger cleared by Canadian Competition Bureau
	2002	
	January	HP-Compaq merger cleared by EU DG Competition
	March	US FTC clears HP-Compaq merger
19	March	HP shareholders approve merger
20	March	Compaq shareholders approve merger
3	May	Compaq becomes part of HP
7	May	New HP-Compaq officially launched

of competition can be seen in both the high-end server arena, where in early 2001 Sun introduced its newest line at half the price of comparable IBM products and at the lower specification end of the market, where Dell and Compaq, the top two Windows server sellers, were reportedly involved in a price war to keep market share.[13] Van Reenen (2006) estimates the yearly, quality-adjusted price fall

for servers. IDC (2002) reports that 'The fallout from the dot-com bubble and the "perfect storm" in the IT industry that preceded it caused the worldwide server market to decline by nearly 20% in 2001'.

[13]Gartner Group (1999), documents that US revenues dropped by 4.3%, whereas shipments grew by 15.9% from 1997Q4 to 1998Q4.

for servers to be of the order of 30% between 1996 and 2001 in the US (and about 22% in Western Europe).[14]

The process of consolidation involved a number of acquisitions of smaller competitors by the bigger players - IBM, Sun, HP, Compaq and Dell. Acquisitions include those of Sequent, Data General, DEC and Tandem by the bigger players and the Fujitsu-Siemens combined server operation, in an attempt to be able to take on the other big players, at least in Europe.

The process of consolidation involved a number of acquisitions of smaller competitors by the bigger players - IBM, Sun, HP, Compaq and Dell. Acquisitions include those of Sequent (by IBM), Data General (by EMC), DEC and Tandem (both by Compaq) and the Fujitsu-Siemens combined server operation, in an attempt to be able to take on the other big players, at least in Europe.

3.2.3 The HP-Compaq Merger

Compaq was an established player in the computer industry by the time it entered the server market in about November 1989, producing its one-millionth server in November 1996, By 2001Q1 it ranked first in worldwide server sales.

Hewlett Packard (HP) had built a strong reputation in the inkjet and laser printer markets, which it introduced in 1984.[15] In the 1990s, it expanded their computer product line, which initially had been targeted at university, research, and business customers, to reach consumers. HP also grew through a number of acquisitions, from Apollo Computer in 1989, to Convex Computer in 1995, before acquiring Compaq.

HP and Compaq formally announced their plans to merge on September 4, 2001 - see Table 3.2. According to HP, their aim was to create a \$ 85bn global technology leader capable of challenging IBM's worldwide leading position. The cost of the deal was approximately \$25bn, and the parties argued the cost-savings from synergies would amount to \$2.5bn in savings, with an estimated 15,000 job cuts. They further argued the combined entity would become the market leader in servers, storage, management software, printing and imaging, and PCs, being in a position to offer the end-to-end solutions that customers demand. The combined entity would also double HP's profitable and growing services business, enhance its R&D and extend its customer base worldwide. The estimated cost-savings would add between \$5 and \$9 in present value to each HP share and increase earnings per

[14] Even before making any adjustments for quality increase, van Reenen (2006) reports price falls of about 10% in the US in the period 1996-2001. Interestingly, he finds distinct price trends for low-end and high-end servers: whereas the former fell by about 30% per annum in 1996-2001, the latter fell by about 20%, suggesting that they belong to different market segments. This finding is supported by anecdotal evidence from market observers, according to which the high-end servers would face high demand from e-business, web applications and customer databases, not to mention computer intensive uses such as supply-chain management, weather forecast, video web applications, the design of molecules etc. (See Gartner Group (2001).)

[15] See http://www.hp.com/hpinfo/abouthp/histnfacts/ for details of HP's history.

share by 13% during the first year following the merger. Moreover, according to HP, by improving its profitability the new company would have the financial strength to extend its successful imaging and printing franchise into new promising categories such as digital imaging and digital publishing.

The prevailing view at the time of the merger was that Compaq could help HP with additional market share in both desktops and servers, besides enabling a 'substantial additional presence' in the service market, given its prior acquisition of DEC. According to Reuters, "Broad hints from HP and a look at market share led industry observers to conclude that HP would use Compaq's well-regarded storage system and its NT servers, low-end network computers that run Microsoft Windows, in stitching together its product line-up. The combined company was also expected to use HP's own high-end Unix servers and Compaq's mainframe-style computer, the (Tandem) Himalaya."[16]

Such a global merger requires clearance by a number of antitrust authorities. Despite the concern that the merger would be blocked by antitrust authorities, the merger was ultimately cleared in all markets where it underwent scrutiny. The merger first received clearance from the Canadian Competition Bureau in December 2001. The European Commission approved the merger in early 2002, after focusing on the activities of the combined company in the markets for PCs, servers, PDAs, storage solutions and services. And in March 2002, the merger was cleared by the FTC. The merger was completed in March 2002. Due to the sharp decline in share prices, the deal valued at $25bn when the merger was announced was then worth only $19bn, approximately.

3.3 Evaluating the Incentives to Collude

3.3.1 Qualitative Analysis of Coordinated Effects in the HP-Compaq Merger

Chamberlin (1929) argued informally that when the same firms repeatedly interact, oligopolists may have an incentive to implicitly or tacitly collude and the result would be prices above competitive levels. Stigler (1964), economic theory, antitrust merger guidelines and court judgments (in particular Airtours) all suggest that whether firms can tacitly sustain collusive prices depends on a number of factors. Specifically, Stigler (1964) argues that in order to sustain collusion firms must be able to (i) come to an agreement (which can be difficult when products are complex and differentiated), (ii) monitor each others' behaviour (in order to detect undercutting) and, of course, (iii) enforce collusive behavior collectively by punishing the cheating firms.

Agreement. The server industry involves a significant number of large players, but with a material competitive fringe of smaller firms. The distribution of brands and market shares varies across time

[16] See http://www.reuters.com/article/filmNews/idUSN0245554920020322.

and markets, but participants in my dataset include AST, Acer, Apple, Compaq, Data General, Dell, Digital, Fujitsu, Gateway, HP, Hitachi, IBM, Micron, Mitsubishi, NCR, NEC, Siemens, SGI, Sun, Toshiba, Unisys and VA Linux. Most firms are multi-product firms and so the number of products is substantial. In my dataset, a single geographic market has a maximum of 222 products for sale in a single quarter.[17] At face value, such a large number of differentiated products appear to make the problem of agreeing on a tacitly collusive outcome very difficult in this industry. On the other hand, this is a setting where a considerable amount of the product heterogeneity may in fact not be greatly valued by customers in the sense that each firm may face a substantially elastic demand for its product, as consumers will easily substitute across providers – at least at a given point in the quality spectrum.[18]

For the purposes of this paper I will suppose that firms can achieve such an agreement, despite the considerable complexity even without the ability to meet to exchange information and/or audit each others accounts. However, I make this assumption cautiously and I do note that this is potentially an important caveat for my results, as well as an important area for future research. Ideally, I would like to be able to capture the way in which the number of dimensions of an agreement may affect the likelihood of collusion in the way in which the model I build cannot. Naturally, as in other areas of microeconomics, the way in which prices are related to product characteristics may provide such a dimension-reducing solution for tacit colluders (Lancaster, 1961, Gorman, 1955).[19]

Monitoring. One of the Airtours[20] conditions is sufficient market transparency to ensure that all tacitly colluding oligopolists would become aware 'sufficiently precisely and quickly' of the way in which other members' market conduct is evolving. The industry has a considerable amount of publicly available market information, since companies such as IDC and Gartner publish detailed quarterly information on shipments (quantities) and revenues by product by geographic market. This degree of transparency is not universal across markets, but it is a considerably greater degree of transparency than was available in the Sony/BMG merger that was the subject of the Impala judgements by the CFI

[17] The maximum number of products observed in our dataset was in Q3 of 2000 in the EU market area.

[18] There is clearly a big quality difference between a high and a low-end server. However, it is less clear that consumers perceive significant differences across providers at a given point in the product quality spectrum.

[19] In the famous case involving GE and Westinghouse in electrical turbines, a published pricing book with formulae were used to help to map product characteristics to prices. Similarly, in the US airline industry the prospect of 'per-mile' pricing was allegedly used as a potential simplifying tool to facilitate a proposed tacitly collusive arrangement.

[20] Case T-342/99, Airtours v. Commission, CFI Judgment of June 6, 2002 (See in particular paragraph 62.) "[F]irst, each member of the dominant oligopoly must have the ability to know how the other members are behaving in order to monitor whether or not they are adopting the common policy. As the Commission specifically acknowledges, it is not enough for each member of the dominant oligopoly to be aware that interdependent market conduct is profitable for all of them but each member must also have a means of knowing whether the other operators are adopting the same strategy and whether they are maintaining it. There must, therefore, be sufficient market transparency for all members of the dominant oligopoly to be aware, sufficiently precisely and quickly, of the way in which the other members' market conduct is evolving;"

and subsequently by the ECJ[21], but even this information must be treated carefully as at least some information is collated from voluntary reports from companies who may, on occasion, have incentives to either over- or under-report sales.

Naturally, in addition to the industry data available, there are industry associations, publications and conferences which each provide opportunities for informal communication and/or intelligence gathering. Furthermore there is a considerable degree of multi-market contact and firms will often meet in forums, such as standard-setting organizations, as well as sometimes explicitly and legitimately cooperating through joint ventures.[22]

Enforcement. There are two aspects to stability of tacit collusion, internal and external stability which, respectively, apply to the ability of those tacitly colluding to sustain the collusive outcome(s) and the inability of those not tacitly colluding to gain by disrupting it. In the main, in this paper I focus on techniques that can help evaluate this aspect of a tacitly collusive theory of harm. One of the Airtours conditions[23] was that tacit collusion must be sustainable over time and the CFI noted that some notion of retaliation was "inherent" in this condition.[24] Retaliation can take a number of forms, from a reversion to Nash equilibrium to temporary price wars (See Green and Porter (1984) and Porter (1983)'s analysis of the railroad Joint Executive Committee in the 1880s.)

Two introductory observations are worthy of note in relation to the HP/Compaq merger. First, while I have noted that there are a number of features of the marketplace that potentially facilitate collusion in the server industry and other significant features hindering it, neither the DOJ/FTC nor DG Competition reported material consideration of a theory of harm of coordinated effects when analyzing the HP-Compaq case. Second, there were no significant objections to the merger from rivals such as Dell, IBM and Sun.

Before moving on to discuss consider enforcement in the context of the server industry, I pause to note that I am not able to consider whether there is an important role for capacity constraints in

[21] In Sony/BMG and the appeals by Impala, the question before the courts was whether or not the fact that sales numbers were published weekly in Billboard Magazine was sufficient to establish a level of transparency that would enable a competition authority to consider blocking the merger, even though transaction prices between retailers and music distributors were not published.

[22] One such case is the NEC-Mitsubishi joint venture of computer monitors and LCD panels, agreed upon in September 1999, see http://findarticles.com/p/articles/mi_m0CGN/is_3750/ai_55805203; Another such example is the collaboration between HP and NEC to develop internet protocol servers for the Japanese market agreed upon in July 1999, see http://www.hoise.com/primeur/99/articles/monthly/AE-PR-08-99-19.html.

[23] See also, Gencor v Commission, para 276.

[24] Case T-342/99, Airtours v. Commission, CFI Judgment of June 6, 2002 (See in particular paragraph 62.) "[S]econd, the situation of tacit coordination must be sustainable over time, that is to say, there must be an incentive not to depart from the common policy on the market. As the Commission observes, it is only if all the members of the dominant oligopoly maintain the parallel conduct that all can benefit. The notion of retaliation in respect of conduct deviating from the common policy is thus inherent in this condition. In this instance, the parties concur that, for a situation of collective dominance to be viable, there must be adequate deterrents to ensure that there is a long-term incentive in not departing from the common policy, which means that each member of the dominant oligopoly must be aware that highly competitive action on its part designed to increase its market share would provoke identical action by the others, so that it would derive no benefit from its initiative (see, to that effect, Gencor v Commission, paragraph 276);"

facilitating or hindering tacit collusion in this paper, since I do not have data on either inventories or production capacities. I do so while noting that it is nonetheless possible that there is an important role for capacity constraints, perhaps resulting from the shortage of silicon (used to manufacture processors), processors themselves, or other components, such as memory and hard disks.

Finally, I note the issue of future market growth can be relevant to an evaluation of the likelihood of tacit collusion in an industry. At the end of the period of the dataset, the 'perfect storm' in the IT industry led many observers to believe that the market for servers was expected to shrink. In fact, research by Gartner Group in August 2001 argued that companies had overspent \$1bn on application server technology since 1998.[25] Their report estimated that an additional \$2bn was expected to be overspent in the following two years, and recommended that companies be cautious when acquiring server technology, pointing out that application server vendors were encouraging customers to purchase higher-end technology that they did not need.

3.3.2 The Benchmark Model

In this section, I define the framework used for exploring internal stability of a tacitly colluding group of oligopolists. For the most part, the framework is exactly that developed by Friedman (1971) and familiar to all students of industrial organization since that time.[26] I have, of course, had to generalize it appropriately to allow for differentiated products and multi-product firms. Readers familiar with unilateral effects merger simulation will also recognize considerable overlap with the framework typically examined in that literature, a fact that I view as a considerable strength of the approach I am suggesting.

3.3.2.1 The Stage Game

I begin by studying the stage-game of the dynamic model. The stage game is simply a standard differentiated product Bertrand pricing game, identical to that used in the unilateral effects merger simulation literature by Werden and Froeb (1994), Berry (1994), Hausman et al (1994), Berry, Levinsohn and Pakes (1995) and Nevo (2002). Specifically, I suppose the existence of J products in the market of interest and study the pricing game wherein each firm f of the F active ones produces a

[25] See http://www.gartner.com/5_about/press_releases/2001/pr20010820b.html.

[26] The model builds upon a vast literature in the last four decades. Notable contributions to the literature of repeated games include Abreu (1986), who studies symmetric Cournot repeated games, and Brock and Scheinkman (1985) and Lambson (1987), who investigate symmetric Bertrand repeated games. Davidson and Deneckere (1984) study how mergers impact collusion using trigger strategies and exogenous market sharing rules, starting from a setting with symmetric capacities and Bertrand competition.

subset of products, $\Im_f \subset \{1, ..., J\}$, and chooses the prices of those products to maximize their profits:

$$\max_{\{p_j | j \in \Im_f\}} \sum_{j \in \Im_f} (p_j - c_j) D_j(p)$$

$$\begin{aligned} s.t.\ D_j(p) &\geqslant 0, j \in \Im_f \\ p_j &\geqslant 0, j \in \Im_f \end{aligned}$$

where c_j is the marginal cost of product j, assumed constant. Provided equilibrium prices of all goods in the market are positive and all goods are sold in positive quantities as is universally assumed in the existing empirical literature (and so the constraints for this program do not bind in equilibrium), the first-order conditions are given by:

$$D_k(p) + \sum_{j \in \Im_f} \frac{\partial D_j(p)}{\partial p_k} (p_j - c_j) = 0$$

I now introduce the 'ownership matrix' which, for every product in the market, assigns the firm producing it. Define the matrix Δ of dimension J by J and typical element

$$\Delta_{jk} = 1\{\text{both } j \text{ and } k \text{ produced by the same firm, } j, k = 1, ..., J\}$$

where $1\{.\}$ is the indicator function. Notice that changing ownership structure in merger simulations amounts to changing this ownership indicator matrix – in particular, a monopoly is denoted by setting every element of Δ equal to one. Using the ownership indicators, the firm's first order condition may be simply rewritten as:

$$D_k(p) + \sum_{j=1}^{J} \Delta_{jk} \frac{\partial D_j(p)}{\partial p_k} (p_j - c_j) = 0, k = 1, ..., J$$

The (implicit) solution to this set of equations, $p^{NE} = (p_1^{NE}, ..., p_J^{NE})$, provides the prices at which each firm is maximizing its profits given the prices of others, and hence is the Nash equilibrium price to the stage game. Notice that there is one of these first-order conditions from firm f's objective function for every $k \in \Im_f$. Since every product is owned by some firm, I obtain a total of J first-order conditions, one for every product provided each firm chooses its prices to maximize its own profits.

3.3.2.2 The Repeated Game

I now consider the above analysis as a stage game within the broader context of an infinitely repeated game. Following the repeated game literature, each firm is assumed to maximize its net present value of profits, and I require that at each point in the game tree the firm makes choices which are optimal given that it reached that node of the game tree, so that I study sub-game perfect equilibria of the repeated game[27].

The repeated game literature has developed a considerable range of possible punishment mechanisms, including 'simple penal codes', as presented in Abreu (1988), and optimal punishment mechanisms à la Abreu, Pearce and Stacchetti (1990). While such punishment mechanisms may well have a role in the analysis of tacit co-operation, I begin with the simple approach, suggested by Friedman (1971), of Nash reversion, which has a number of desirable properties. First, Abreu (1988, Theorem 5) shows that simple strategies - a constant sequence of the same static Nash equilibrium - suffice to achieve any feasible subgame-perfect equilibrium payoff. Second, if any deviation is followed by a Nash reversion, the punishments are automatically ensured to be credible (Friedman, 1971). However, Nash reversion is usually not the most severe punishment to defection - this happens only if the Nash equilibrium of the stage game coincides with the minmax payoffs, as in my setting. However, understanding Nash reversion appears likely to be a useful and tractable benchmark, leaving more sophisticated forms of punishment for further research.

As in Friedman, I consider the incentives to collude or defect when each player adopts grim strategies and consider the sustainability of a tacitly collusive equilibrium. To do so, I introduce the following notation. Denote the one period Nash equilibrium payoffs to firm f as π_f^{NE} and the one period returns to collusion by firm f as π_f^{Coll}. Similarly denote the one period gain to firm f from defection when all other firms are playing collusively as π_f^{Def}.

A defector, when rivals are playing grim strategies, earns his one period defection payoff and then subsequently receives only his Nash equilibrium profits. Thus, the net anticipated return to defection today for firm f is

$$V_f^{Def}(\delta_f) = \pi_f^{Def} + \frac{\delta_f \pi_f^{NE}}{1-\delta_f}$$

while her payoff to collusion today and in all subsequent periods given that rivals continue to collude is

$$V_f^{Coll}(\delta_f) = \frac{\pi_f^{Coll}}{1-\delta_f}$$

[27] See Selten (1965).

Hence, firm f has no incentive to deviate from collusive pricing provided that:

$$\begin{aligned} V_f^{Coll}(\delta_f) &> V_f^{Def}(\delta_f), f = 1, ..., F \\ \frac{\pi_f^{Coll}}{1-\delta_f} &> \pi_f^{Def} + \frac{\delta_f \pi_f^{NE}}{1-\delta_f}, f = 1, ..., F \end{aligned}$$

In order to examine the incentives to collude using grim strategies, I must therefore consider the returns achieved by each firm in the three pricing scenarios – 'Collusion', 'Nash equilibrium pricing' and 'defection'. Since I discussed above how to compute Nash equilibrium and (perfect) collusion profits, it remains to show how to calculate the payoff to defection.

Is is worth mentioning that the only components in this equation that are not evaluated in a unilateral effects merger simulation are (i) the payoff to defection π_f^{Def} and (ii) the discount factor δ_f. The former, like the Nash and collusive equilibrium payoffs, depends directly on the nature of the static profit function for each firm and therefore may be easily calculated using the methodologies developed for the analysis of data generated by static pricing games, as I detail below.

These F incentive compatibility constraints play the key role in defining the set of situations in which tacit collusion is individually rational for each firm and hence feasible. Depending on the parameters of interest, one may wish to consider that these inequalities define a set of discount factors, a set of demand and cost conditions or, alternatively, a set of collusive prices – which, as we shall see, may or may not be the perfectly collusive prices we associate with maximizing industry profits.

In an antitrust case, the discount factor could usually be taken from internal documents specifying the companies required rate of return. Alternatively, if companies are listed, an asset pricing model could be used to infer an appropriate discount rate for payoffs. Thirdly, and more closely paralleling the theoretical literature, I can report the range of discount factors for which collusion could be sustained under any given industry structure.

3.3.2.3 The Payoff to Defection

Following the theoretical literature on repeated games, I define the payoff to defection to firm f as the maximum amount of profits that could be achieved given its rivals' prices (ie., treating them as fixed). In the case most directly of interest, where firm f is deciding whether or not to defect from the collusive (industry profit maximizing) tacit agreement, other firms will be choosing their prices to be the collusive prices, and so the static payoff to firm f when defecting is:

$$\pi_f^{Def} := \max_{\{p_j | j \in \Im_f\}} \sum_{j \in \Im_f} (p_j - c_j) D_j(\underline{p}_f, \underline{p}_{-f}^{Coll})$$

$$\begin{aligned} s.t.\ D_j(\underline{p}_f, \underline{p}_{-f}^{Coll}) &\geqslant 0, j \in \Im_f \\ p_j &\geqslant 0, j \in \Im_f \end{aligned}$$

For the case of linear demand equations, this non-linear maximization problem is a quadratic objective function subject to linear constraints and so is easy to solve, even for large problems, using standard methods. Generally, it must be solved numerically but even in those cases it is a simpler mathematical object to evaluate than the Nash Equilibrium computed in unilateral effects merger simulations, since it involves only an optimization: the calculation of a 'defection' price simply involves finding the location on the deviating firms' best response function at the point where it's rivals are charging collusive prices.

3.3.3 Extensions

3.3.3.1 Competitive Fringe

To study the effects of a competitive fringe, suppose the existence of F_{df} dominant firms and F_{cf} firms constituting a competitive fringe, so that the total number of firms in the market is given by $F = F_{df} + F_{cf}$. Assume also that there are two subsets of products, $\Im_{df} \subset \{1, ..., J_{df}\}$ and $\Im_{cf} \subset \{J_{df+1}, ..., J_{df+cf}\}$ where $J_{df} + J_{cf} = J$, which owned by, respectively, the dominant firms and the competitive fringe and partition the price vector into components corresponding to dominant firms and the competitive fringe, $p = (\underline{p}^{df\prime}, \underline{p}^{cf\prime})'$. In this modified pricing game, the competitive fringe faces a flat demand curve, so that the solution to the problem

$$\max_{\{p_j | j \in \Im_{cf}\}} \sum_{j \in \Im_{cf}} (p_j - c_j) D_j(p)$$

$$\begin{aligned} s.t.\ D_j(p) &\geqslant 0, j \in \Im_{cf} \\ p_j &\geqslant 0, j \in \Im_{cf} \end{aligned}$$

and the associated first-order condition,

$$D_k(p) + \sum_{j \in \Im_{cf}} \frac{\partial D_j(p)}{\partial p_k}(p_j - c_j) = 0, k \in \Im_{cf}$$

is given by

$$p_j = c_j, j \in \Im_{cf}$$

resulting in the zero-profit condition for this subset of firms. For the remaining F_{df} dominant firms, the first-order condition is obtained in a similar way, but the demand curve faced by those firms is downward-sloping, so

$$D_k(p) + \sum_{j \in \Im_{df}} \frac{\partial D_j(p)}{\partial p_k}(p_j - c_j) = 0, k \in \Im_{df}$$

For a firm f among the dominant firms, the static defection profits are obtained as the solution to the problem

$$\pi_f^{Def} := \max_{\{p_j | j \in \Im_f\}} \sum_{j \in \Im_f} (p_j - c_j) D_j(\underline{p}_{-f}^{df}, \underline{p}_{-f}^{df,Coll})$$

$$\begin{aligned} s.t.\ D_j(\underline{p}_{-f}^{df}, \underline{p}_{-f}^{df,Coll}) &\geqslant 0, j \in \Im_f \\ p_j &\geqslant 0, j \in \Im_f \end{aligned}$$

whereby the defecting firm f only takes into account the prices set by the other dominant firms operating in the market, denoted by $\underline{p}_{-f}^{df}$, which are assumed to be collusion prices. It then follows that, among the remaining F_{df} dominant firms, each firm f faces an incentive-compatibility constraint similar to the ones in the standard model, whereby it has no incentive to deviate from collusive pricing provided that:

$$\begin{aligned} V_f^{Coll}(\delta_f) &> V_f^{Def}(\delta_f),\ f = 1..., F_{df} \\ \frac{\pi_f^{Coll}}{1-\delta_f} &> \pi_f^{Def} + \frac{\delta_f \pi_f^{NE}}{1-\delta_f},\ f = 1..., F_{df} \end{aligned}$$

3.3.3.2 Partial Coalitions

The second approach I take is to consider what happens when the number of firms that tacitly cooperate is fewer than the full set of firms in the industry. That is, I suppose that a number of firms remain outside the tacitly cooperating coalition of firms so that the tacitly cooperating firms constitute only a partial coalition. I have considered that full cooperation would involve collusive prices that maximize industry profits, the profits that would result if every product were under the ownership of one firm. In this case, I assume that tacitly co-operative profits are those that result from the Nash equilibrium outcome that would have occurred if only all members of the coalition had merged. Doing so allows us to (i) consider what happens to the incentive compatibility of tacitly cooperating as the number of tacitly cooperating firms falls and (ii) consider what happens to the incentive to tacitly collude of those outside the tacitly collusive arrangement. This case is similar to that provided by the competitive fringe, except that those firms not tacitly cooperating are allowed to optimally (but statically) exploit market power endowed by the actions of the cartel. The problem can be stated as follows. Assume the existence of F_c firms taking part in the cartel and F_{nc} firms not taking part in the cartel and acting strategically, so that the total number of firms in the market is given by $F = F_c + F_{nc}$. Assume also that there are two subsets of products, $\Im_c \subset \{1, ..., J_c\}$ and $\Im_{nc} \subset \{J_{c+1}, ..., J_{c+nc}\}$ where $J_c + J_{nc} = J$, which is owned by, respectively, the cartel and non-cartel firms. Now each firm faces the first order condition

$$D_k(p) + \sum_{j=1}^{J} \Delta_{jk} \frac{\partial D_j(p)}{\partial p_k}(p_j - c_j) = 0, k = 1, ..., J$$

where the ownership matrix now reflects the existence of a subset of firms taking part in the cartel ie. Δ_{jk} takes value one not only if products i and j are produced by the same firm, but also if i and j belong to $\Im_c$, the set of products produced by the cartel.

To define the defection profits I partition the price vector as $p = (\underline{p}^{c\prime}, \underline{p}^{nc\prime})'$, for cartel and non-cartel firms, respectively. The static defection profits are now obtained as the solution to the problem

$$\pi_f^{Def} := \max_{\{p_j | j \in \Im_f\}} \sum_{j \in \Im_f} (p_j - c_j) D_j(\underline{p}_f^c, \underline{p}_{-f}^{c,Coll}, \underline{p}^{nc})$$

$$\begin{aligned} s.t.\ D_j(\underline{p}_f^c, \underline{p}_{-f}^{c,Coll}, \underline{p}^{nc}) &\geqslant 0, j \in \Im_f \\ p_j &\geqslant 0, j \in \Im_f \end{aligned}$$

with f denoting the defecting firm and $-f$ the other firms constituting the cartel. The corresponding set incentive-compatibility constraints is given by

$$\begin{aligned} V_f^{Coll}(\delta_f) &> V_f^{Def}(\delta_f),\ f = 1..., F_c \\ \frac{\pi_f^{Coll}}{1-\delta_f} &> \pi_f^{Def} + \frac{\delta_f \pi_f^{NE}}{1-\delta_f},\ f = 1..., F_c \end{aligned}$$

3.3.3.3 Multimarket Contact

If firms interact repeatedly, not only over time but also across $M > 1$ markets, be it geographical markets or market niches, collusion is more likely to happen (Bernheim and Whinston, 1990). Essentially, multimarket contact pools the incentive-compatibility constraints of the different markets where firms interact.

In the standard model, I define $\pi_{f,m}^{Def}, \pi_{f,m}^{Coll}, \pi_{f,m}^{NE}$, as, respectively, the defection, collusion and Nash static profits and $\delta_{f,m}$ as the discount factor[28] of firm f operating in market m, I can define the value functions $V_{f,M}^{Def}(\underline{\delta}_f)$ and $V_{f,M}^{Coll}(\underline{\delta}_f)$ as the sums across markets of the value functions to defection and collusion:

$$\begin{aligned} V_{f,M}^{Def}(\underline{\delta}_f) &= \sum_m V_{f,m}^{Def}(\delta_{f,m}) = \sum_m \pi_{f,m}^{Def} + \frac{\delta_{f,m}\pi_{f,m}^{NE}}{1-\delta_{f,m}} \\ V_{f,M}^{Coll}(\underline{\delta}_f) &= \sum_m V_{f,m}^{Coll}(\delta_{f,m}) = \sum_m \frac{\pi_{f,m}^{Coll}}{1-\delta_{f,m}} \end{aligned}$$

The multimarket incentive-compatibility constraint for collusion for firm f then reads

$$V_{f,M}^{Coll}(\underline{\delta}_f) > V_{f,M}^{Def}(\underline{\delta}_f)$$

To study the effects of a competitive fringe coupled with multimarket contact, consider a firm f belonging to the subset of dominant firms and define $\pi_{f,m}^{Def}, \pi_{f,m}^{Coll}, \pi_{f,m}^{NE}$, as, respectively, the defection, collusion and Nash static profits and $\delta_{f,m}$ as the discount factor of dominant firm f operating in market m. I can then define the value functions $V_{f,M}^{Def}(\underline{\delta}_f)$ and $V_{f,M}^{Coll}(\underline{\delta}_f)$ as the sums across markets of the value functions to defection and collusion:

$$\begin{aligned} V_{f,M}^{Def}(\underline{\delta}_f) &= \sum_m V_{f,m}^{Def}(\delta_{f,m}) = \sum_m \pi_{f,m}^{Def} + \frac{\delta_{f,m}\pi_{f,m}^{NE}}{1-\delta_{f,m}}, \; f = 1..., F_{df} \\ V_{f,M}^{Coll}(\underline{\delta}_f) &= \sum_m V_{f,m}^{Coll}(\delta_{f,m}) = \sum_m \frac{\pi_{f,m}^{Coll}}{1-\delta_{f,m}}, \; f = 1..., F_{df} \end{aligned}$$

and the multimarket incentive-compatibility constraint for collusion for dominant firm f then reads

$$V_{f,M}^{Coll}(\underline{\delta}_f) > V_{f,M}^{Def}(\underline{\delta}_f), \; f = 1..., F_{df}$$

To study the effects of the cartel coupled with multimarket contact, consider a firm f belonging to the cartel and define $\pi_{f,m}^{Def}, \pi_{f,m}^{Coll}, \pi_{f,m}^{NE}$, as, respectively, the defection, collusion and Nash static profits and $\delta_{f,m}$ as the discount factor of firm f of the cartel operating in market m. I can then define the valuefunctions $V_{f,M}^{Def}(\underline{\delta}_f)$ and $V_{f,M}^{Coll}(\underline{\delta}_f)$ as the sums across markets of the value functions to defection

[28] We allow for different discount factors across markets for the sake of generality, since firms might have different discount rates for e.g. emerging vis-a-vis developed markets.

and collusion:

$$\begin{aligned} V_{f,M}^{Def}(\underline{\delta}_f) &= \sum_m V_{f,m}^{Def}(\delta_{f,m}) = \sum_m \pi_{f,m}^{Def} + \frac{\delta_{f,m}\pi_{f,m}^{NE}}{1-\delta_{f,m}}, \ f = 1...,F_c \\ V_{f,M}^{Coll}(\underline{\delta}_f) &= \sum_m V_{f,m}^{Coll}(\delta_{f,m}) = \sum_m \frac{\pi_{f,m}^{Coll}}{1-\delta_{f,m}}, \ f = 1...,F_c \end{aligned}$$

and the multimarket incentive-compatibility constraint for collusion for each firm f taking part in the cartel reads

$$V_{f,M}^{Coll}(\underline{\delta}_f) > V_{f,M}^{Def}(\underline{\delta}_f), \ f = 1...,F_c$$

3.4 Data

I have used data from the International Data Corporation's (IDC) Quarterly tracker database.[29] This enables us to analyze the evolution of price, revenue and unit sold of every server since the first quarter of 1996 through the first quarter of 2001 in three major regions (USA, Western Europe, Japan). In principle IDC covers the population of models. It gathers revenue and characteristics data from vendors in all the main regions and then cross checks the company totals with global HQs and its own customer surveys. Transaction prices (called "street prices" by IDC) are also estimated on a region-specific, quarterly, model-by-model basis based on discussions with industry participants. These prices take into account the various discounts offered off the list price as well as trade-ins.

Looking in a cross section I define an observation in a region as a vendor-family-modeloperating system. A model (I use model and brand interchangeably) is distinguished within a family (with some grouping). So a typical example would be Sun Microsystem's (vendor) Ultra-Enterprise (family) 1000HE series (model) running UNIX (OS). There are 33 separately identified vendors most of whom will have only two or three families (IBM has the most models and seventeen individual "families").

One obvious concern is that the IDC data only has basic model characteristics. To address this substantial time was invested in collecting extra data on server characteristics and matching them into the IDC data. The IDC Quarterly Tracker was used as my "population" and matched in new server characteristics by name and by time period. Several sources were used to obtain these data, including the trade publication Datasources, company web pages, back issues of computer magazines and their web pages and major resellers. The final dataset covered 60% of the IDC models and over 80% of the revenues of all servers.

[29] 7th June 2001 version. For a full description of this database and the recent evolution of the market see IDC (1998,2000b)

The characteristics I have include the number of rack slots, the chip architecture (RISC, CISC or IA32), motherboard type (e.g. Symmetric Parallel Processing - SMP, Massively Parallel Processing - MPP), the types of operating system used (Windows, UNIX, Netware, OS390/OS400, VMS, Linux and others), vendor dummies and , whether the system is rack-optimised.[30]

I also collected GDP, exchange rate, consumer price and PPP information from the OECD. Input cost indices were collected from various sources. Input prices for servers include the FRB's (Federal Reserve Board) quality adjusted price index for semi-conductor chips (Aizorbe, 2000), BLS (Bureau of Labor Statistics) quality adjusted price indices for hard disk drives and other secondary electronic input products. All prices are in 1996 US dollars.[31]

Raw prices have fallen rapidly among servers (at about 15% per annum) and quality has improved dramatically. IBM, Compaq, HP, and Sun Microsystems have been the leading server hardware vendors throughout my sample period. Dell Computers entered in late 1995 and has grown dramatically to also be a major player by the end of the period.

I have also used stock market and balance sheet data to compute the firms' discount factors. The former are from CRSP and COMPUSTAT Global, whereas the latter are from COMPUSTAT Global, as well as interest rate data from the Bank of Japan. Finally, I also use the SMB, HML and momentum factors from Ken French's webpage[32].

3.5 Model Implementation

To implement the model, I need to estimate a set of incentive-compatibility constraints. The ingredients needed for that are

1. A demand model to calculate the firms' static profit functions under alternative ownership schemes; and

2. An asset pricing model to obtain the firms' discount rates.

Intuitively, the demand model chooses parameter values such that the distance between model-based (coming from utility maximization - see below) and observed market shares is minimized - see Berry (1994) for a thorough discussion. Since price is (potentially) endogenous, instrumental variable tech-

[30] Rack-mounted servers are designed to fit into nineteen inch racks. They allow multiple machines to be clustered or managed in a single location.

[31] The US CPI inflation rate from January/1996 to December/2007 is 36.03%, whereas the one from March/2001 to December/2007 is 19.20%.

[32] Available from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html

niques whithin a GMM framework are employed; computation of the asset pricing model is standard, using regression techniques.

To compute the profit functions I also need estimates of the marginal costs of each product, and I obtain those using a standard technique in that I back out marginal costs from pre-merger prices. Simply put, one assumes that the first-order condition of the firm's problem holds, so that the only unknowns are the marginal costs of the products in the market, for which the equation is solved for.[33]

In what follows I describe the demand model, the identification strategy and the calculation of discount factors.

3.5.1 Demand Estimation

My random-coefficients logit model follows Berry, Levinsohn and Pakes (1995, BLP hereafter), whereby one starts with a microeconomic model of rational behaviour for individual consumers which is then aggregated, generating the (aggregate) market demands. Individual heterogeneity here is model in a way not to restrict substitution patterns *a priori*, so that the resulting cross-product elasticities are obtained as a result of their distance in characteristic space - as opposed to ,eg., the standard logit specification.

The conditional indirect utility u of individual i when consuming product j from market m is given by

$$u_{ijm} = \sum_{k=1}^{K} x_{jmk}\beta_{ik} + \xi_{jm} + \varepsilon_{ijm},\ \ i = 1, ..., I;\ j = 1, ..., J;\ m = 1, ..., M$$

where x_{jmk} are observed product characteristics such as price, memory, speed and storage, ξ_{jm} represent unobserved (by the econometrician) product characteristics, assumed observed by all consumers. The consumer-specific coefficients can be decomposed as $\beta_{ik} = \lambda_k + \sigma_k' v_i$, where λ_k measures the impact of characteristic k on mean utility, the unobservable consumer attributes are denoted by v_i, the vector σ_k measures the impact of unobservable consumer attributes on characteristic k. Finally, ε_{ijmt} is an individual and option-specific idiosyncratic component of preferences, assumed to be a mean zero Type I Extreme Value random variable independent from both the consumer attributes and the product characteristics.

The specification of the demand system is completed with the introduction of an outside good, since some consumers decide not to buy any server. The conditional indirect utility from the outside option is:

$$u_{i0} = \xi_{0m} + \sigma_0 v_i + \varepsilon_{i0}$$

[33] Recall that pre-merger prices are observed in the data, so marginal costs would be the only unknowns in this system of equations.

where ε_{i0} is a mean zero individual market and time specific idiosyncratic term and v_i is an individual specific component reflecting heterogeneity in tastes.

More generally[34], the utility of consumer i when consuming product j is given by

$$u_{ij} = u(x_j, p_j, v_i; \theta), \; j = 0, 1, ..., J$$

where x_j is a vector of observable (by the econometrician) produt characteristics, p_j is the price of product j (kept separate from the other ones given its endogeneity), v_i represents consumer-level heterogeneity, J is the number of available products and θ is the parameter vector of interest, measuring the impact of preferences on the utility of consumers. The subset of preferences leading to the choice of product j is given by

$$A_j(\theta) = \{v_i : u_{ij} > u_{ik}, \text{ for all } i, \text{ and for } k \neq j\}$$

and ties are assumed to have zero probability.

Intuitively, once $A_j(\theta)$ is determined for every product $j = 1, ..., J$, one has to count the probability that a product is chosen, so that the market shares can be computed. Letting $f(v)$ be the distribution of preferences in the population, the choice probabilities, which coincide with the model-generated market shares (assuming $f(.)$ is correctly specified) are given by

$$s_j(x, p; \theta) = \int_{v \in A_j(\theta)} f(v) d(v)$$

where the integral is typically computed using simulation. after computing the vector of model-generated market shares, one has to solve a system relating model-generated and observed market shares, $s^{obs} = s(\xi, ...; \theta)$. In general, this system has no closed-form solution, so numerical methods have to be deployed. Fortunately, the solution has been shown to exist and is a unique – this is the famous contraction mapping result, see Berry (1994) and BLP (1995).

The reader will note that this is a demand specification which (i) treats each individual server acquisition as a separate choice (see, for example, Hendel, 1999) and (ii) abstracts away from explicitly modeling inter-temporal substitution (see, for example, Nevo and Hendel (2006) and Gowrisankaran and Rysman (2005) for an alternative approach). On the first point, I do note that many servers are purchased by businesses and this may therefore potentially be a strong assumption. However, since I have no information about the numbers of servers purchased by purchaser, and little indication that

[34] I now follow BLP (1995) closely.

such effects ultimately make the aggregate sales of servers 'lumpy' in the way that might concern us, I do not think the approach unreasonable. On the second issue of intertemporal substitution, again the approach seems not unreasonable in light of the relatively short lives of computers, although abstracting in this way clearly represents a pragmatic modeling approximation to actual consumer choice behaviour in the industry. More general demand specifications could certainly be possible in future work and may even, in fact, have important implications for the incentives faced by oligopolists.

3.5.2 Identification Strategy

Following the literature I treat price as endogenous in the demand specification and this section describes the identification strategy for demand. First, I note that there is considerable variation in prices in the dataset. Sales prices have fallen rapidly for servers (at about 15% per annum) and quality has improved dramatically. To estimate my model, besides the exogenous characteristics, I use four classes of instruments, for a given product in a given regional market (US, Japan and Western Europe). The first class were the so called "Hausman instruments", proposed by Hausman, Leonard and Zona (1994) and consisting of prices of a product in other regions instrumenting its price in a given region.

The second was the class of "Distance instruments", where one uses the price of the closest product with respect to a certain characteristic as the instrument of the price of a given product; within this category, I consider the prices of (i) upper neighbours; (ii) lower neighbours; and (iii) the average between upper and lower neighbours considering characteristics such as CPU capacities and number of CPUs. Their justification comes from the fact that, when setting the price of a given product, firms takes into account the characteristics of the competing products in the market – for instance, firms spend considerable amounts on market research, surveys and conjoint analysis studies. For illustration, consider a server characteristic such as memory: when setting the price of a (new) product, the firm looks at the products in the market and sets its price taking into account the prices of products with memory capacities similar to the one of the product it is launching ie. its product's immediate neighbours. In particular, it looks at the memory capacities of its lower quality and higher quality neighbours, from which the product is expected to suffer more intense competition.

The third was the class of "BLP instruments" (following BLP, 1995), a set of polynomial basis functions of exogenous variables exploiting the three-way panel structure of the data, consisting of the number of firms operating in the market, the number of other products of the same firm and the sum of characteristics of products produced by rival firms. firms; for the other two regional markets, this class consists of the number of other products of the same firm and the sum of characteristics of rival firms.

Finally, I also consider a class of instruments also used in Ivaldi and Lörincz (2005), following Bresnahan, Stern and Trajtenberg (1997) in their study of the PC market and which can be seen as a generalization of the BLP instruments. Besides using the exogenous characteristics of the products, I use a set of polynomial basis functions of exogenous variables within a group and/or subgroup.[35] For the same market, I calculate the number of other products of the same firm and the number of firms in the same group; and the number of other products of the same producer in the same group and same subgroup. From the other two regional markets, I calculate the number of rival firms producing products in the same group; and the number of rivals producing products in the same group and same subgroup. BST instruments implicitly assume a form of localized competition among products, and this seems consistent with anecdotal evidence from the server industry, characterized by a number of market niches and highly heterogeneous products.

3.5.3 Estimating Discount Factors

In this section I detail how to compute the discount factors used to calculate the net present value of the profits of a firm:

$$V_f(\delta_f) = \pi_f + \delta_f \pi_f + \delta_f^2 \pi_f + ... = \frac{\pi_f}{1 - \delta_f}$$

In an antitrust case, the discount factor could be taken from internal documents specifying the firms' discount rate. An alternative way of treating discount factors, which most close parallels the theoretical literature, is to calculate the range of discount factors for which collusion could be sustained under any given industry structure. Here, however, I make use of the fact that the firms of interest are listed and rely on asset pricing models to infer an appropriate discount rate for payoffs. In the empirical implementation of the model I also compare the 'collusion threshold' of the discount rate to the ones obtained using stock market data.

Our object of interest will be the cost of capital (or discount rate) r_f^* of firm f, related to the discount factor in the following way:

$$\delta_f = \frac{1}{1 + r_f^*}$$

If a firm faces no risk and has free access to the financial markets, the discount rate corresponds to the market interest rate. In practice, however, firms are risky and issue sock and debt to finance their projects, so the weighted average cost of capital (WACC) takes the cost of equity and the (after-tax)

[35] Group refers to processor architecture and can take on the following: UP, SMP and MPP, for uni-processor, symmetric multi-processor, and massively parallel processing, respectively. Sub-group refers to operating systems used; these are Linux, Netware, Windows NT, IBM OS400, IBM OS 390, VMS, Unix, Other.

cost of debt into account, weighted by the proportion of debt and equity in the value of the firm:

$$r_f^* = \frac{D}{V}(1-\tau)r_f^d + \frac{E}{V}r_f^e$$

where D/V and E/V are, respectively, the ratio of debt and equity to the value of the firm, r_f^d is the cost of debt, r_f^e is the cost of equity and τ is the marginal corporate tax rate. In practice, D is the book-value of debt, E is the market value of equity (number of shares outstanding times share price), and $V = E + D$.

The cost of debt r_f^d is obtained as the ratio between interest expenses over debt for a given firm, whereas the cost of equity is obtained from an asset pricing model . The CAPM is an asset pricing model according to which a stock's systematic risk is measured by its beta, the slope of the regression of excess stock returns on excess market returns:

$$E(r_f - r) = \beta_f \times E(r_M - r)$$

where r_f is the return of the stock of firm f, r is the risk-free rate and r_M is the return of the market portfolio. Asset pricing models such as the CAPM provide the expected rate of return of the firms' stocks. If investors expect a given rate of return from the stock of a firm, the cost of capital rule says that the firm should have been using the same discount rate to compute the net present value of its projects[36]. The corresponding discount rate of asset f is then given by

$$r_f^e := E(r_f) = r + \beta_f \times E(r_M - r)$$

The CAPM does not, however, price other relevant risk factors. In other terms, there are empirical regularities in asset returns which the CAPM is unable to capture, such as the fact that stocks from small companies tend to outperform stocks of large ones. Fama and French (1993) propose to incorporate factors associated with size and book-to-market, resulting in asset pricing models with superior performance when compared to the CAPM. Their model reads:

$$E(r_f - r) = \beta_f E(r_M - r) + s_f r_{SMB} + h_f r_{HML}$$

where r_{SMB} and r_{HML} are returns to, respectively, the SMB (Small-Minus-Big) and the HML (High-Minus-Low) factors. Carhart (1997) extends the Fama-French model by adding a fourth factor that

[36] See Berk and DeMarzo (2007) for an introduction to capital budgeting.

captures the Jegadeesh and Titman (1993) momentum anomaly. The resulting model (referred to as FFM, for Fama-French plus Momentum) has four risk factors and reads

$$E(r_f - r) = \beta_f E(r_M - r) + s_f r_{SMB} + h_f r_{HML} + m_f r_{MOM}$$

The cost of equity obtained from the general FFM model then reads:

$$r_f^e := E(r_f) = r + \beta_f E(r_M - r) + s_f r_{SMB} + h_f r_{HML} + m_f r_{MOM}$$

3.6 Results

3.6.1 Demand Estimation

3.6.1.1 Market Segments

Although market research companies and regulators consider different market segments, these are defined in an informal way, typically involving price thresholds. IDC, for instance, defines the volume market segment as the one constituted of servers priced below \$25,000, the mid-range segment as the one with servers priced between \$25,000 and \$500,000, and the high-end segment as the one with servers priced above \$500,000. In a number of rulings, including the Compaq-DEC, Fujitsu-Siemens and the HP-Compaq ones, the DG Competition delineated the server markets on the basis of price bands into separate markets for entry-level servers (below \$100,000), medium level servers (\$100,000 - \$999,999) and large servers (above \$1,000,000).

Recently, however, Ivaldi and Lörincz (2005) investigated the server market and found evidence of many sub-segments within the volume segment using both the traditional SSNIP test and their alternative FERM one. Based on their analysis, I focus on two market segments, namely that consisting of servers priced below \$4,000 and the one consisting of servers priced between \$4,001 and \$10,000, with 5,537 and 8,799 observations, respectively.[37]

Only four players – Compaq, Dell, HP and IBM – have market shares above 1% in all three regions in the 0-4 market segment (all of them happen to have more than 5%), and the associated C_4 concentration ratios are 51%, 69%, and 61.5% for US, EU and Japan, respectively. In the 4-10

[37] Ivaldi and Lörincz (2006) studied the US, European and Japanese markets separately conducting two versions of the SSNIP test (5% and 10%) and their FERM test (10% and 15%). The 10% SSNIP and 15% FERM tests suggested the same first price threshold at USD 4,000 (the 5% SSNIP test suggested USD 3,000 for Japan and the 10% FERM test suggested USD 3,000 for the US). The 10% SSNIP test also suggested a second price threshold of USD 10,000 for both the US and Europe, and USD 9,000 for Japan. When it comes to the following price thresholds, the methods disagree markedly: the 10% SSNIP test suggests a threshold at USD 1mn for all three regions, but the 15% FERM test suggests a total of 5 market segments for the US and 6 for both the EU and Japan, but the threshold are markedly different. We thus decided to consider a market segment with servers priced below USD 4,000 and another with servers priced between USD 4,000 and USD 10,000. (We refer to these market segments as 0-4 and 4-10, respectively, in the text.)

segment, only five players – the ones above plus Sun – have market shares above 1% in all three regions, but only three of them command more than 5% of market share in all three regions (Compaq, Dell and IBM). The associated C_4 and C_5 concentration ratios are, respectively, 88%, 82%, 39% and 94%, 86%, 46% for US, EU and Japan, suggesting a much higher concentration than in the 0-4 segment.

3.6.1.2 Demand Estimates

Other than price, I consider characteristics such as CPU capacity ("CPU_CAP"), number of CPUs ("CPU_COUNT"), the number of extra racks available in the server ("UR", an important measure of scalability), and a dummy indicating whether the server is rack-optimized ("RACK"). I also consider the interaction of (i) the number of extra racks available in the server, CPU architecture fixed-effects (CISC, RISC or IA32, for Intel 32-bit architecture), and operating system fixed-effects (Linux, Netware, Windows NT, Openvms, OS/400, Unix, Other) with (ii) processor architecture (UP for uniprocessor, SMP for symmetric multi-processor, MPP for massive multi-parallel processor), region (US, EU, JP), and time fixed-effects. Finally, I also interact firms and regions with time fixed-effects. The variety of interaction terms allow flexibility enough to distinguish, for instance, the effects of the number of extra racks (also CPU architecture and operating system) over time for uni-processors and symmetric multi-processors separately; and how the penetration of different firms in different regions evolves over time eg. the evolution of the Japanese players in Japan (or Fujitsu in Japan and Fujitsu-Siemens in Europe) given the advances of Dell. Although I considered variables such as number of CPUs and their capacities in early specifications, they were found to be either insignificant or collinear with other characteristics, thus dropped from the estimation.

I consider four classes of instruments to account for price endogeneity. These were Hausman instruments, distance instruments, BLP instruments and BST instruments. The Hausman, BLP and BST instruments for the 0-4 segment are compared in Table 3.3, which shows the sensitivity of the results to the choice of instruments. The BLP instruments are the only to perform well, delivering a negative and significant price coefficient (β_{price}) and highly elastic own price effects. For the 4-10 segment, however, the only to perform well are the BST ones, that also deliver highly elastic own price affects, albeit of smaller magnitude than those of the 0-4 segment – see Table 3.3. The intuition for having different instruments working for different price intervals has to do with the increased complexity and differentiation of servers as their prices increase: low-end servers are fairly standardized, undifferentiated, products, since there are less options of processor type, architecture, number of processors, OS etc a vendor can use when designing and building such a model. In practice, this will make the BST instruments (also used by Ivaldi and Lörincz, 2005), which are based on interactions between variables such as processor type (dummies for uni-processor, symmetric multi processing and multi-parallel processing, for instance) tend to collapse and become collinear with the intercept.

Based on the findings for the standard logit specifications, I estimate a number of random coefficients logit models. For each of the price segments considered, I estimate demand models using BLP and BST instruments for the 0-4 and 4-10 price segments, respectively, and random draws generated by a Lognormal distribution, which has positive support and generates only negative own price elasticities. In all cases reported, first-stage regression of price on the instruments were found to have F-statistics significant at the one percent level.

For the 0-4 price segment, as reported in the last column of Table 3.3, the price coefficients are negative and significant, but not the price dispersion coefficient (σ_{price}), which makes the random-coefficients logit collapse into the standard logit model, a finding that suggests that consumers care equally about price for products within this segment. The own price elasticities for the BLP-LN specification, with BLP instruments and lognormal random draws, are within the range between -11 and -40 (-13 to -40 in the US, -17 to -40 in the EU and -11 and -40 in Japan), whereas average ones are around -30 and median ones around -31.

As in the results for the 0-4 segment, the dispersion coefficient for price was not significant for the 4-10 segment – see Table 3.3. The elasticities obtained for this market segment were somewhat lower, with medians between -21 and -23, but still extremely high.

For the 0-4 segment, for instance, the results show the downward time trend for UP (uni-processor) motherboards, whereas for the 4-10 segment one can observe an overall positive time effect of SMP and especially MPP motherboards coupled with the CISC architecture in contrast with the negative effects of the RISC architecture when interacted to the same variables. The impact of both UP and SMP motherboards over time turns from positive to negative during the sample for the 4-10 market segment, but this change not only occurs first, but is also more dramatic for the former.

3.6.2 Discount Factors

I calculated the discount factors using balance sheet and asset return data. Whenever the calculations returned discount factors greater than one, I inputed discount factors using the firm country's term structure of interest rates and the one-year holding period as benchmarks.

Generally speaking, larger firms are expected to have lower betas, thus higher discount factors, due to smaller cost of capital and, by-and-large, this is what happens for US companies, as reported in Table 3.4. IBM, for instance, has a lower cost of capital than Dell, resulting in discount factors of 0.970 and 0.950, respectively.

TABLE 3.3. Results for Alternative Demand Specifications for the 0-4 Price Segment – World Market

		Standard Logit			RC Logit
Panel A: 0-4 Segment	**OLS**	**Hausman**	**BLP**	**BST**	**BLP-LN**
Mean price	0.539	22.138	−13.168	5.547	−12.189
	(3.158)	(3.543)	(−2.262)	(3.216)	(−2.150)
	***	***	**	***	**
Price Dispersion					0.155
					(0.010)
					–
J-statistic p-value	NA	–	–	–	–
100% elastic own price effects?	No	No	Yes	No	Yes
Median own price effects					
EU			−31.290		−30.917
US			−30.163		−29.804
JP			−32.284		−31.921

		Standard Logit			RC Logit
Panel B: 4-10 Segment	**OLS**	**Hausman**	**BLP**	**BST**	**BST-LN**
Mean price	−0.079	−0.135	−1.631	−1.877	−3.979
	(−1.398)	(0.233)	(−1.434)	(−2.029)	(−2.872)
	–	–	–	**	***
Price Dispersion					0.978
					(1.600)
					–
J-statistic p-value	NA	–	–	–	–
100% elastic own price effects?	No	No	No	Yes	Yes
Median own price effects					
EU				−10.602	−22.727
US				−9.808	−21.145
JP				−11.159	−23.796

Note: The Table reports estimates of the (mean) price coefficient of logit model, t-statistics (reported within brackets) and the p-value of the price coefficient and the J-statistic of overidentifying restrictions, whenever applicable. The last column displays results for a random-coefficients logit with BST instruments and Lognormal random draws. The symbols – (resp. *, **, ***) denote not-significant (resp. significant at the 10%, 5%, 1% significance level). It also reports whether 100% of the own price elasticities were elastic and, whenever this is the case, reports the median own price elasticity.

The results reported in Table 3.4 also show that American companies are those with the smallest discount factors, which corresponds to a higher cost of capital. The findings for Japan can be understood as a consequence of the years of economic stagnation in the Japanese economy during the 1990s and early 2000s, when interest rates were kept at impressive low levels and market returns were consistently negative.

TABLE 3.4. Discount Factors for Firms in the Server Industry

Firm	Discount Factor
Compaq	0.967
Dell	0.957
Gateway	0.962
HP	0.967
IBM	0.972
Micron	0.963
NCR	0.971
SGI	0.976
Sun	0.959
Unisys	0.971
Fujitsu	0.989
Hitachi	0.989
Mitsubishi	0.990
NEC	0.988
Toshiba	0.989

Note: The calculation of discount factors was performed as described in Section 5.3. The Appendix describes the calculations in detail.

3.6.3 Evaluating the Incentives to Collude

I now investigate whether there are incentives to collusion in the network server industry and whether they increase as a result of the HP-Compaq merger. Schematically, there are two cases to consider, depending on whether there was collusion before the merger. If there is no collusion prior to the merger, the question is whether the merger creates the incentives to collude. If, however, there is collusion prior to the merger, the question is whether the merger enhances the incentives to collude, following the decrease in the number of firms interacting in the market.

A typical structural merger analysis would detect little market power in both the 0-4 and 4-10 price segments, as already hinted by the high elasticities obtained in the demand model. As Table 3.5 reports, price-cost margins are less than, respectively, 0.05 and 0.10, so that the argument for blocking a merger on the grounds of unilateral effects would be weak. Indeed, as reported in Table 3.6, even the price changes resulting from a pre-merger Nash-Bertrand equilibrium to a post-merger collusive equilibrium would be at least weak, less than one and four percent for the 0-4 and 4-10 price segments, respectively, regardless of the geographical market.

The profit functions of the stage game for each region and market segment, reported in Table 3.7 and 38 for the 0-4 and 4-10 segments, respectively, confirm the coexistence of a handful of big players interacting in several (regional and niche) markets as well as a number of small players, often operating in only one market. The static gains from defecting are extremely small – for instance, IBM and Dell would gain less than $20,000 in a quarter by defecting in each of the markets in the 0-4 segment.

Moreover, the profit increases following the merger are modest, in that they increase about seven percent in the US from a pre-merger Nash equilibrium to a post-merger collusive equilibrium, and less than that for the remaining geographical markets – see Table 3.7.

Defecting does indeed become more attractive in the 4-10 price segment – now the static gains faced by IBM and Dell in the US and EU markets would be short of $1mn in a given quarter. Moreover, profit increases from a pre-merger Nash equilibrium to a post-merger collusive equilibrium are in the range 9-15% in the US, about 13% in the EU, and in the range 20-23% in Japan for the non-merging parties – see Table 3.8.

The above analysis lacks the study of the incentives to colluding *vis-a-vis* defecting for each firm operating in a given market. In what follows, I empirically examine both the existence of incentives to collude and how they change following the HP-Compaq merger. In other terms, I investigate the sustainability of collusive equilibria by estimating a set of incentive-compatibility constraints and comparing, for each of the market participants, how the value function to colluding compares to the one of defecting.

In my standard model, I evaluate the incentive-compatibility (IC) constraints to collusion considering price segments (0-4 and 4-10) and geographic markets (US, EU and Japan) separately. The next step is to assume the existence of a handful dominant firms coexisting with a competitive fringe; in what I call the CF model, the number of IC constraints in a given market equals the number of dominant firms. Finally, I consider the setting whereby every firm operating in the market has some market power, but only a small subset of them takes part in a cartel. For each case, I aggregate the IC constraints across geographic markets and also across both geographic markets and price segments in what I call the multimarket (MM) versions of the model.

3.6.3.1 Incentive Compatibility Constraints: The Benchmark Model

In what follows, I investigate the sustainability of collusive equilibria by examining my estimated set of incentive-compatibility constraints and comparing, for each of the market participants, how the value function to colluding compares to that associated with the option of defecting. In this sub-section I examine the benchmark model, while in later sections I examine how my benchmark results change as a result of (i) a competitive fringe, (ii) multi-market contact and (iii) coalitions of players.

In my benchmark model, I evaluate the incentive-compatibility (IC) constraints to collusion considering price segments (0-4 and 4-10) and geographic markets (US, EU and Japan) separately. Notably, I find that at discount factors calculated using each firms' WACC, the net present value (NPV) from collusion is always greater than the NPV of defection. This result, that the returns to collusion are

TABLE 3.5. Price-Cost Margins by Firm and Geographical Market

	0-4 Market Segment		4-10 Market Segment	
Firm	**Coll**	**Nash**	**Coll**	**Nash**
Panel A: US				
Compaq	0.039	0.035	0.077	0.050
HP	0.029	0.025	0.085	0.051
NEC	0.037	0.032	0.051	0.026
IBM	0.032	0.028	0.076	0.046
AST	0.024	0.020	–	–
Data General	–	–	0.047	0.024
Dell	0.034	0.030	0.083	0.062
Gateway	0.049	0.042	–	–
Hitachi	0.038	0.032	–	–
Micron	0.045	0.039	–	–
NCR	–	–	0.055	0.029
SGI	0.026	0.022	–	–
Sun	0.032	0.027	0.063	0.035
Toshiba	0.034	0.029	0.085	0.049
Unisys	0.026	0.022	–	–
VA Linux	0.033	0.028	–	–
Panel B: EU				
Compaq	0.040	0.036	0.064	0.045
HP	0.029	0.025	0.069	0.041
Fujitsu	0.026	0.022	0.069	0.041
NEC	0.037	0.031	0.084	0.050
IBM	0.041	0.035	0.068	0.042
AST	–	–	0.090	0.054
Data General	0.027	0.022	–	–
Dell	0.033	0.028	0.079	0.050
Gateway	0.031	0.026	0.093	0.056
SGI	0.029	0.024	–	–
Sun	–	–	0.064	0.036
Toshiba	0.030	0.025	0.081	0.047
Unisys	–	–	0.062	0.034
VA Linux	0.034	0.028	0.081	0.048
Panel C: JP				
Compaq	0.040	0.034	0.073	0.042
HP	0.058	0.048	0.070	0.038
Fujitsu	0.029	0.024	0.075	0.045
NEC	0.026	0.022	0.072	0.043
IBM	0.036	0.031	0.073	0.040
AST	–	–	0.092	0.053
Dell	0.039	0.033	0.080	0.046
Gateway	0.035	0.029	0.084	0.046
Hitachi	0.027	0.022	0.069	0.037
Mitsubishi	0.031	0.025	0.079	0.044
NCR	–	–	0.080	0.044
Sun	–	–	0.068	0.036
Toshiba	0.034	0.027	0.068	0.037

TABLE 3.6. Model-Implied Average Percentual Price Changes Following HP-Compaq Merger

Firms	0-4 Market Segment		4-10 Market Segment	
	$p_{pre}^{NE} \to p_{post}^{Coll}$	$p_{pre}^{NE} \to p_{post}^{NE}$	$p_{pre}^{NE} \to p_{post}^{Coll}$	$p_{pre}^{NE} \to p_{post}^{NE}$
Panel A: US				
Compaq	0.439	0.051	3.008	0.116
HP	0.423	0.128	3.808	0.663
NEC	0.581	0	2.623	0
IBM	0.456	0	3.243	0.004
AST	0.379	0	–	–
Data Gen	–	–	2.455	0
Dell	0.432	0	2.270	0.013
Gateway	0.741	0	–	–
Hitachi	0.595	0	–	–
Micron	0.699	0	–	–
NCR	–	–	2.752	0
SGI	0.421	0	–	–
Sun	0.504	0	2.933	0.002
Toshiba	0.532	0	3.926	0.001
Unisys	0.419	0	–	–
VA Linux	0.513	0	–	–
Panel B: EU				
Compaq	0.500	0.137	1.982	0.242
HP	0.434	0.165	3.013	1.142
Fujitsu	0.427	0	3.022	0.011
NEC	0.666	0	3.702	0.005
IBM	0.628	0.001	2.796	0.018
AST	–	–	3.926	0.004
Data Gen	0.496	0	–	–
Dell	0.522	0	3.149	0.017
Gateway	0.568	0	4.042	0.004
SGI	0.53	0	–	–
Sun	–	–	2.953	0.005
Toshiba	0.555	0	3.652	0.003
Unisys	–	–	2.981	0.001
VA Linux	0.621	0	3.674	0.003
Panel C: JP				
Compaq	0.728	0.048	3.340	0.195
HP	1.053	0.111	3.411	0.341
Fujitsu	0.533	0	3.256	0.007
NEC	0.472	0	3.084	0.007
IBM	0.549	0	3.512	0.003
AST	–	–	4.308	0.001
Dell	0.636	0	3.739	0.003
Gateway	0.706	0	4.081	0.001
Hitachi	0.528	0	3.455	0.002
Mitsubishi	0.614	0	3.869	0.001
NCR	–	–	3.942	0.001
Sun	–	–	3.378	0.002
Toshiba	0.651	0	3.339	0.002

TABLE 3.7. Pre- and Post-Merger Static Profits – 0-4 Price Segment

Firm	Pre-Merger Coll	Pre-Merger Def	Pre-Merger Nash	Post-Merger Coll	Post-Merger Def	Post-Merger Nash	%Change $\pi^{NE}_{pre} \rightarrow \pi^{Coll}_{post}$	%Change $\pi^{NE}_{pre} \rightarrow \pi^{NE}_{post}$
Panel A: US								
Compaq	3527	3553	3293					
HP	1053	1067	984	4580	4606	4278	7.1	0.0
NEC	73	74	68	73	74	68	6.9	0.2
IBM	1313	1329	1227	1313	1329	1229	7.0	0.2
AST	147	150	137	147	150	138	7.1	0.2
Dell	2431	2455	2269	2431	2455	2273	7.1	0.2
Gateway	387	393	363	387	393	363	6.8	0.2
Hitachi	4	4	4	4	4	4	6.9	0.2
Micron	93	94	87	93	94	87	6.8	0.2
SGI	2	2	2	2	2	2	7.0	0.2
Sun	124	126	116	124	126	116	6.9	0.2
Toshiba	111	112	104	111	112	104	6.9	0.2
Unisys	45	46	42	45	46	42	7.0	0.2
VA Linux	283	287	264	283	287	264	7.0	0.2
Panel B: EU								
Compaq	2328	2350	2238					
HP	1409	1429	1357	3737	3755	3600	4.0	0.1
Fujitsu	813	827	783	813	827	787	3.8	0.5
NEC	83	85	80	83	85	81	3.7	0.5
IBM	1068	1084	1029	1068	1084	1034	3.7	0.5
Data Gen	0	0	0	0	0.1	0	3.8	0.5
Dell	923	938	889	923	938	893	3.8	0.5
Gateway	49	50	47	49	50	47	3.7	0.5
SGI	14	15	14	14	15	14	3.8	0.5
Toshiba	13	13	13	13	13	13	3.7	0.5
VA Linux	4	4	4	4	4	4	3.7	0.5
Panel C: JP								
Compaq	434	443	425					
HP	272	278	267	706	718	692	2.0	0.1
Fujitsu	495	506	485	495	506	486	2.0	0.1
NEC	568	580	556	568	580	557	2.0	0.1
IBM	1079	1095	1056	1079	1095	1057	2.1	0.1
Dell	840	855	823	840	855	824	2.0	0.1
Gateway	59	60	58	59	60	58	1.9	0.1
Hitachi	215	221	211	215	221	211	1.9	0.1
Mitsubishi	89	91	87	89	91	87	1.9	0.1
Toshiba	192	197	189	192	197	189	1.9	0.1

Note: Figures are in thousands 1996 US dollars per quarter, rounded to the nearest thousandth. The last two columns report the percentage changes in profits following a merger between HP and Compaq in two situations: (i) from a pre-merger Nash equilibrium to apost-merger collusive equilibrium; and (ii) from a pre-merger to a post-merger Nash equilibrium. The post-merger values for HP report the percentage change of the post-merger values of the merged entity, HP-Compaq, with respect to the combined pre-merger values of HP and Compaq.

positive and non-negligible both pre- and post-merger, appears to be robust across all three geographic regions and both market segments.

TABLE 3.8. Pre- and Post-Merger Static Profits – 4-10 Price Segment

	Pre-Merger			Post-Merger			%Change	
Firm	Coll	Def	Nash	Coll	Def	Nash	$\pi_{pre}^{NE} \rightarrow \pi_{post}^{Coll}$	$\pi_{pre}^{NE} \rightarrow \pi_{post}^{NE}$
Panel A: US								
Compaq	7816	8776	7026					
HP	1388	1686	1258	9204	10219	8304	11.1	0.2
NEC	1	1	1	1	1	1	9.1	0.9
IBM	4383	5141	3964	4383	5142	3992	10.5	0.7
Data Gen	2	3	2	2	3	2	8.9	0.9
Dell	19221	20015	16702	19221	20015	16806	15.0	0.6
NCR	305	405	279	305	405	281	9.3	0.8
Sun	1843	2299	1677	1843	2299	1690	9.8	0.8
Toshiba	31	38	28	31	38	28	10.3	0.7
Panel B: EU								
Compaq	19097	20276	16390					
HP	3878	4633	3440	22976	23946	20002	15.9	0.9
Fujitsu	3220	3964	2868	3220	3964	2964	12.2	3.4
NEC	672	812	596	672	812	613	12.6	2.9
IBM	6571	7689	5814	6571	7689	5999	13.0	3.2
AST	352	418	312	352	418	321	12.6	2.7
Dell	6718	7662	5944	6718	7662	6110	13.0	2.8
Gateway	1	2	1	1	2	1	12.6	2.6
Sun	1261	1597	1124	1261	1597	1162	12.2	3.4
Toshiba	62	75	55	62	75	57	12.6	2.8
Unisys	25	33	23	25	33	23	12.2	3.3
VA Linux	52	62	46	52	62	47	12.6	2.8
Panel C: JP								
Compaq	2390	2870	1962					
HP	1246	1578	1022	3636	4235	2992	21.9	0.3
Fujitsu	3163	3769	2579	3163	3769	2603	22.6	0.9
NEC	3748	4384	3043	3748	4384	3071	23.1	0.9
IBM	1220	1551	1001	1220	1551	1010	21.8	0.9
AST	24	30	20	24	30	20	20.5	0.8
Dell	1272	1548	1052	1272	1548	1060	20.9	0.8
Gateway	1	1	1	1	1	1	21.0	0.8
Hitachi	667	872	547	667	872	552	21.8	0.9
Mitsubishi	414	523	341	414	523	344	21.2	0.8
NCR	13	16	11	13	16	11	20.6	0.8
Sun	1047	1353	858	1047	1358	866	22.0	0.9
Toshiba	976	1242	804	976	1242	811	21.4	0.9

Note: Figures are in thousands 1996 US dollars per quarter, rounded to the nearest thousandth. The last two columns report the percentage changes in profits following a merger between HP and Compaq in two situations: (i) from a pre-merger Nash equilibrium to a post-merger collusive equilibrium; and (ii) from a pre-merger to a post-merger Nash equilibrium. The post-merger values for HP report the percentage change of the post-merger values of the merged entity, HP-Compaq, with respect to the combined pre-merger values of HP and Compaq.

For example, in the low-end server market segment, the results presented in Table 3.9 show that the net gains from colluding are significant in all geographic markets. Prior to the merger, a firm such as Dell would obtain gains of \$4.7m pre-merger in the US and almost \$1m and \$0.5m in Europe and Japan respectively, whereas the gains of IBM in the three markets would be about \$2.5m, \$1.1m and

\$0.7m. On the other hand, by acquiring Compaq, the net gains perceived by HP as a result of collusion increase from \$2m to \$8.8m. Similar results are presented for the \$4-10,000 market segment in Table 3.10.

However, strikingly, the merger is predicted to make the net gains from colluding smaller in each market area for every firm. In fact, the reason is intuitive – the merger (absent efficiencies) increases Nash equilibrium profits for all players in the market and thus reduces the incentive to collude.[38] This result is a general one to the extent that optimally collusive prices do not change with market structure, neither will each non-merging firms' defection payoffs, and so fairly generically unilateral effects of a merger will tend to narrow the incentive to coordinate. The only potential exceptions to this can appear through the merging firms' incentives to collude. I capture this result in Proposition 1.

[38] In the present setting, (cost) efficiencies are materialized as reductions in marginal costs. Thus, "10% efficiencies" means that a pre-merger marginal cost of c will, as the result of the merger, become $0.9c$.

Proposition 1 *In the benchmark model with tacit collusion, if $\delta_f, f \in \Im^{non-merging}$, are fixed, the merger does not generate efficiencies and the firms produce weakly substitutable goods*[39]*, then for any ownership structure before and after the merger, Δ^{pre} and Δ^{post}:*

1. $V_f^{collusion}(\Delta^{pre}) = V_f^{collusion}(\Delta^{post})$ *for all* ***non-merging*** *firms* $f \in \Im^{non-merging}$;

2. $V_f^{defection}(\Delta^{pre}) \leq V_f^{defection}(\Delta^{post})$ *for all* ***non-merging*** *firms* $f \in \Im^{non-merging}$;

3. *The incentive for all* ***non-merging*** *firms to tacitly collude always (weakly) narrows post-merger*

   $V_f^{collusion}(\Delta^{pre}) - V_f^{defection}(\Delta^{pre}) \geq V_f^{collusion}(\Delta^{post}) - V_f^{defection}(\Delta^{post})$, *for all* $f \in \Im^{non-merging}$;

4. *If $\delta_f^{pre} \geq \delta_f^{post}$ for all $f \in \Im^{merging}$, then the aggregate returns to the* ***merging firms*** *tacitly colluding pre-merger are no smaller than the returns post-merger:*

$$\sum_{f \in \Im^{merging}} V_f^{collusion}(\Delta^{pre}) \geq \sum_{f \in \Im^{merging}} V_f^{collusion}(\Delta^{post}).$$

   If $\delta_f^{pre} = \delta_f^{post} = \delta$ for all $f \in \Im^{merging}$, then for the ***merging firms***

$$\sum_{f \in \Im^{merging}} V_f^{collusion}(\Delta^{pre}) = \sum_{f \in \Im^{merging}} V_f^{collusion}(\Delta^{post}).$$

   That is the returns to tacit collusion for the merging parties are greater pre-merger than post-merger provided the cost of capital does not decrease following a merger;

5. *For the* ***merging*** *firms, aggregate static defection payoffs*

$$\sum_{f \in \Im^{merging}} \pi_f^{def}(\Delta^{post}) \geq \sum_{f \in \Im^{merging}} \pi_f^{def}(\Delta^{pre})$$

   are (weakly) greater post-merger than pre-merger, as are aggregate post-merger Nash equilibrium profits,

$$\sum_{f \in \Im^{merging}} \pi_f^{NE}(\Delta^{post}) \geq \sum_{f \in \Im^{merging}} \pi_f^{NE}(\Delta^{pre}).$$

[39] Two products are weakly substitutes if the effect of a change in the price of one product on the sales volume of the other product is non-negative.

If $\delta_f^{pre} = \delta_f^{post} = \delta$ *for all* $f \in \Im^{merging}$*, then for the* ***merging firms***

$$\sum_{f \in \Im^{merging}} V_f^{def}(\Delta^{post}) \geq \sum_{f \in \Im^{merging}} V_f^{def}(\Delta^{pre}).$$

Proof.

1. Since the vector of the industry profit maximizing tacitly collusive prices are independent of market structure and $V_f^{\text{coll}} := \frac{\pi_f(p^{\text{coll}})}{1-\delta_f}$, we have immediately that $V_f^{\text{coll}}(\Delta^{pre}) = V_f^{\text{coll}}(\Delta^{post})$ for all non merging firms $f \in \Im^{non-merging}$, provided the discount factors $\delta_f, f = 1, ..., F$ remain unchanged post-merger;

2. Turning to the defection payoff, for any non-merging firm $\pi_f^{def}(\Delta^{pre}) = \pi_f^{def}(\Delta^{post})$ while if a merger occurs between firms producing substitutes then absent efficiencies, all non-merging firms are (weakly) better off so that $\pi_f^{NE}(\Delta^{post}) \geq \pi_f^{NE}(\Delta^{pre})$ for all non-merging firms, $f \in \Im^{non-merging}$. Since $V_f^{def} := \pi_f^{def} + \frac{\delta_f \pi_f^{NE}}{1-\delta_f}$ we can therefore write $V_f^{defection}(\Delta^{pre}) \leq V_f^{defection}(\Delta^{post})$ for all non-merging firms, $f \in \Im^{non-merging}$;

3. Follows directly from (1) and (2);

4. Next notice that for, any $f \in \Im^{merging}$, since $V_f^{\text{coll}} := \frac{\pi_f(p^{\text{coll}})}{1-\delta_f}$ and $\sum_{f \in \Im^{merging}} \pi_f^{NE}(\Delta^{post}) \geq \sum_{f \in \Im^{merging}} \pi_f^{NE}(\Delta^{pre})$ we can write $\sum_{f \in \Im^{merging}} V_f^{\text{collusion}}(\Delta^{pre}) \geq \sum_{f \in \Im^{merging}} V_f^{\text{collusion}}(\Delta^{post})$ provided $\delta_f^{pre} \geq \delta_f^{post}$ for all $f \in \Im^{merging}$;

5. Finally notice that aggregate static defection payoffs $\sum_{f \in \Im^{merging}} \pi_f^{def}(\Delta^{post}) \geq \sum_{f \in \Im^{merging}} \pi_f^{def}(\Delta^{pre})$ are (weakly) greater post-merger than pre-merger, since the enlarged firm has greater flexibility to cheat by undercutting, so cannot be made worse off in so doing than the sum of its constituent parts solving their analogous problems. Similarly, aggregate post-merger Nash equilibrium profits will be higher for the merged firm since any merger will (weakly) increase Nash equilibrium prices towards collusive levels. Finally we note that, if $\delta_f^{pre} = \delta_f^{post} = \delta$ for all $f \in \Im^{merging}$, then $\frac{\delta_f}{1-\delta_f} \sum_{f \in \Im^{merging}} \pi_f^{NE}(\Delta^{post}) \geq \frac{\delta_f}{1-\delta_f} \sum_{f \in \Im^{merging}} \pi_f^{NE}(\Delta^{pre})$ and so

$$\sum_{f \in \Im^{merging}} \left[\pi_f^{def}(\Delta^{post}) + \frac{\delta_f}{1-\delta_f} \pi_f^{NE}(\Delta^{post})\right] \geq \sum_{f \in \Im^{merging}} \left[\pi_f^{def}(\Delta^{pre}) + \frac{\delta_f}{1-\delta_f} \pi_f^{NE}(\Delta^{pre})\right]$$

which is analogous to $\sum_{f \in \Im^{merging}} V_f^{def}(\Delta^{post}) \geq \sum_{f \in \Im^{merging}} V_f^{def}(\Delta^{pre})$. QED

∎

Finally note that there is a general result available for the impact of a merger on the incentive to collude versus defect for the merging parties in my particular special case, since the cost of capital estimates found that $\delta_{HP} = \delta_{Compaq} = 0.967$.[40] It follows from part 4 of Proposition 1 that for the merging parties the aggregate payoff to collusion does not change following the merger, $\sum_{f \in \Im^{merging}} V_f^{\text{collusion}}(\Delta^{pre}) = \sum_{f \in \Im^{merging}} V_f^{\text{collusion}}(\Delta^{post})$, and this result can be seen in Tables 9 and 10. Similarly, it follows from part 5 of Proposition 1 that the aggregate payoff to defection increases following the merger $\sum_{f \in \Im^{merging}} V_f^{def}(\Delta^{post}) \geq \sum_{f \in \Im^{merging}} V_f^{def}(\Delta^{pre})$. Again, this result can be seen in each of the panels in Tables 9 and 10. As a result, it follows immediately that in this case the net incentive to collude decreases post merger: $V_f^{\text{collusion}}(\Delta^{pre}) - V_f^{defection}(\Delta^{pre}) \geq V_f^{\text{collusion}}(\Delta^{post}) - V_f^{defection}(\Delta^{post})$. Although the magnitude of the decrease depends on the estimates on the demand side of the model, this result establishes that the direction of the change does not, rather it relies on the empirical finding that $\delta_{HP} = \delta_{Compaq}$.

Next I compare the critical discount factors of the firms operating in a given market. The critical discount factor of a firm, δ_f^*, is the minimum discount factor sustaining collusion and is obtained be equating the value to colluding to the value to defecting ie. $\delta_f^* = \frac{\pi_f^{Def} - \pi_f^{Coll}}{\pi_f^{Def} - \pi_f^{NE}}$. As reported in Table 3.11 and discussed below, the critical discount factors, above which collusion is incentive-compatible, are substantially lower than those coming from asset pricing, especially for the 0-4 price segment. I also establish a theoretical result for the critical discount factors.

[40] Any small differences in the results below stem from the fact that the discount factors differ in the fourth decimal place.

Proposition 2 ***Critical Discount Factors***. *Define the critical discount factor of firm f for a given market structure* Δ *as* $\delta_f^* = \frac{\pi_f^{Def}(\Delta)-\pi_f^{Coll}(\Delta)}{\pi_f^{Def}(\Delta)-\pi_f^{NE}(\Delta)}$.

1. *If we consider perfect collusion both before and after the merger and the merger generates no efficiencies, then* $\pi_f^{Coll} = \pi_f^{Coll}(\Delta^{pre}) = \pi_f^{Coll}(\Delta^{post})$ *for all* ***non-merging*** *firms* $f \in \Im^{non-merging}$, $\pi_f^{Def} = \pi_f^{Def}(\Delta^{pre}) = \pi_f^{Def}(\Delta^{post})$ *for all* ***non-merging*** *firms* $f \in \Im^{non-merging}$ *and, provided the products in the market are weak substitutes,* $\pi_f^{NE}(\Delta^{pre}) \leq \pi_f^{NE}(\Delta^{post})$ *for all* ***non-merging*** *firms* $f \in \Im^{non-merging}$*. In that case, critical discount factors for non-merging firms weakly increase following a concentration:*

$$\delta_f^{*pre} = \frac{\pi_f^{Def} - \pi_f^{Coll}}{\pi_f^{Def} - \pi_f^{NE}(\Delta^{pre})} \leq \frac{\pi_f^{Def} - \pi_f^{Coll}}{\pi_f^{Def} - \pi_f^{NE}(\Delta^{post})} = \delta_f^{*post} \text{ for all } f \in \Im^{non-merging}$$

2. *For the* ***merging*** *parties, we have* $\delta_f^{*pre} = \frac{\pi_f^{Def}(\Delta^{pre})-\pi_f^{Coll}(\Delta^{pre})}{\pi_f^{Def}(\Delta^{pre})-\pi_f^{NE}(\Delta^{pre})}$ *and*

$$\delta_f^{*post} = \frac{\sum_{f\in\Im^{merging}} \pi_f^{Def}(\Delta^{post}) - \sum_{f\in\Im^{merging}} \pi_f^{Coll}(\Delta^{post})}{\sum_{f\in\Im^{merging}} \pi_f^{Def}(\Delta^{post}) - \sum_{f\in\Im^{merging}} \pi_f^{NE}(\Delta^{post})}$$

for $f \in \Im^{merging}$.

As reported in Table 3.11 and discussed below, the critical discount factors, above which collusion is incentive-compatible, are substantially lower than those coming from the asset pricing model, especially for the 0-4 price segment. Applying the results in Proposition 2, part (1), notice that the critical discount factor for non-merging firms always increases from pre-merger to post-merger levels. This move is consistent with the merger making it harder for non-merging firms to tacitly coordinate. There is no analogous general result for the merging firms, and indeed, one interesting feature of Table 3.11 is that the critical discount factors for the merging parties fall post-merger. That is, they move in the opposite direction to the critical discount factors of the non-merging parties.

However, when evaluating such moves it is important to keep in mind that all the critical discount factors I estimated using the benchmark models are substantially below those which finance methods suggest would be appropriate. The two are closest in the $4-10,000 segment where threshold values are above 0.8 in the US and slightly below that in the other markets. In each segment, I note that the merging firm (HP-Compaq) tends to have a low critical post-merger discount factor compared to rivals, with the exception of Dell, a fast growing firm with above-average margins in the industry, attributed by many to its distinctive business model.

3.6.3.2 Incentive Compatibility Constraints: The Benchmark Model with Multimarket Contact

The effect of multimarket contact essentially amplifies the results of the standard model. Given non-binding incentive-compatibility constraints, their aggregation across geographical markets puts an increased wedge between the value of tacitly colluding when compared to the value of defecting. Table 3.12 shows that for the 0-4 market the pre-merger gains to tacitly colluding are in excess of \$9.6m in present value terms for Compaq, \$3.6m for HP (making a total of \$13.3m), \$6m for Dell and \$4.2m for IBM, whereas the corresponding post-merger values are marginally lower at \$13.2m for HP-Compaq while Dell and IBM's incentives similarly decrease by fairly small amounts. For the same reasons as in the Benchmark case, the gains from colluding decreased for all non-merging parties.

In the 4-10 market segment, the gains from collusion both pre- and post-merger are more substantial in both absolute and relative terms. For example, the results suggest that a player such as Dell would gain almost \$93m post-merger by sustaining the collusive outcome. The gains from Fujitsu and NEC are also noticeable, especially when compared to IBM's post-merger \$32m, and come mostly from their strong position in the Japanese market. However, as in the benchmark case, all non-merging firms see a decrease in their relative incentive to collude post merger relative to the situation pre-merger. In my example, the merging parties also see a decrease in their relative payoff to collusion following the merger.

An analysis of the critical discount factors, reported in Table 3.13, shows that collusion across the three geographical regions is sustained provided discount factors are above 0.6 in the 0-4 segment, whereas a higher value (above 0.8) is would suffice in the 4-10 price segment. Qualitatively therefore these results are the same as for the single market contact case (see Table 3.11.) In particular, the merging parties do find that collusion is 'easier' to sustain post-merger than pre-merger in the sense that their critical discount factor falls post-merger. This movement stands in contrast to the non-merging parties who each find that their critical discount factors increase as a result of the merger. As in the benchmark case, the reason for the latter effect is that non-merging firms static collusive and defection payoffs do not change pre- and post-merger, while their Nash equilibrium payoffs – now in each market - unambiguously rise.

Finally, and for completeness, Table 3.14 reports the results of aggregating across both geographical markets and market niches and shows similar results both in terms of value functions and discount factors.

3.6.3.3 Incentive Compatibility Constraints: The Competitive Fringe Model

So far I have assumed all firms operating in the market have a non-negligible degree of market power. As the server industry is characterized by a number of fringe competitors, I also report results of such a market configuration. In this model, to collude means to maximize the dominant firms' profits with the competitive firms choosing prices at marginal cost. I estimate alternative versions of the model, with four (Compaq, HP, Dell, IBM), five (plus Sun) and six (plus Fujitsu-Siemens) dominant firms and obtained very similar results, so I only report results for the competitive fringe (CF hereafter) model with four dominant firms.

By and large, the qualitative results for the CF model were close to those of the standard one, as reported in Table 3.15.[41] The slight increases in both collusive and defective static profits result in increased values to colluding and defecting, but the former dominate the latter. As a result, the incentives to collude for the non-merging parties reduce following the simulated HP-Compaq merger, but are still positive and sizeable. When compared to the standard model, the pre-merger incentives to collude in the CF model are larger, and their reduction following the HP-Compaq merger also decrease, but to a lesser extent. As before, the figures in the 4-10 segment are substantially larger than those in the 0-4 one.

Finally, Table 3.16 reports the critical discount factors in the CF model. Comparing the results in Table 3.11 and Table 3.16, critical discount factors are, unsurprisingly, significantly larger than in the standard case. Larger critical discount factors reflect the fact that the presence of a competitive fringe makes coordination harder to sustain (in this particular sense emphasized by the theoretical literature.).

Accounting for multi-market contact in the competitive fringe model, involves incorporating two forces acting, generally, in different directions. In the main, multi-market contact helps cooperation while a competitive fringe generally makes it more different. In my empirical example, the net impact is that firms retain the ability to coordinate both pre- and post-merger (see Table 3.17). The bottom panel of Table 3.17 reports critical discount factors which, in comparison with those reported in Table 3.11, clearly indicate that multi-market contact does not fully offset the disadvantages of the competitive fringe so that critical discount factors remain substantially above those for the benchmark model in Table 3.11. Table 3.18 reports the results of allowing for multi-market contact across the price segments as well as across regions.

---

41 The price changes from (i) a pre-merger Nash equilibrium to a post-merger collusive equilibrium; and (ii) a pre-merger to a post-merger Nash equilibrium, are only slightly larger than those reported for the standard model, so are not reported.

3.6.3.4 Incentive Compatibility Constraints: Partial Coalitions

So far, I have studied what happens (i) when all firms have market power and might collude (the standard model); and (ii) there are only a few firms with market power which might collude plus a competitive fringe. In either case, collusion is incentive-compatible both pre- and post-merger, despite decreasing slightly following a simulation of the HP-Compaq merger. One might, however, argue that there might be a small number of firms colluding and a number firms not colluding yet with market power. In such a case, the colluding firms will form a cartel facing strategic outsiders. What I do is to consider a number of arrangements of the cartel to investigate which arrangements form a stable coalition among the firms. The starting point is the standard model, which implicitly assumes that all firms are members of the cartel. As the cartel configuration becomes more selective the value to colluding, albeit still positive and showing the same pattern as in the standard model, decreases. The value to colluding is still positive when the firms in the cartel (pre-merger) are the 'big four' ie. Compaq, HP, Dell and IBM. However, when the number of firms is as small as three, the results start to change.

I first consider a cartel formed by Compaq, HP and IBM, followed by a cartel formed by Compaq, HP and Dell. The first thing to notice when investigating the cartel formed by Compaq, HP and IBM, as reported in Table 3.19, is that collusion is not incentive-compatible in the US and EU when considering the 0-4 price segment – prior to the merger, the gains to colluding are still positive, but they become negative following the merger. Alternatively, the critical discount factors for IBM are greater than one. For the 4-10 price segment, collusion is not incentive-compatible in the EU and Japanese markets instead, and once again the critical discount factors for IBM are greater than one. When aggregating across geographical markets within either the 0-4 or the 4-10 price segments, the value to colluding turns from positive to negative. As a result, multimarket contact is not enough to soften competition in this case.

The cartel formed by Compaq, HP and Dell has a similar behaviour, in that collusion is not incentive-compatible in the EU for the 0-4 price segment following the simulation of the HP-Compaq merger – see Table 3.20. For the 4-10 segment, this is true for both EU and Japan. When aggregating across geographical markets, collusion is incentive-compatible for the 0-4 price segment both prior and following the HP-Compaq merger, but is incentive-compatible both before and after the merger in the 4-10 price segment. When aggregating across price segments, the larger 4-10 segment prevails and collusion is incentive-compatible.

As a result, I obtain a suggestion that a coalition formed by the big four players is stable both before and after the simulation of the HP-Compaq merger. This is not necessarily the case when the coalition includes only three of those big players. When IBM is assumed to be the third member of the coalition, collusion is not incentive-compatible and breaks down; if the third member of the coalition is assumed to be Dell, collusion is still incentive-compatible at the aggregate level, despite breaking down in individual markets – a consequence of multimarket contact softening competition.

3.6.3.5 Incentive Compatibility Constraints: Side Payments

I then investigate what happens if I allow for side payments, that means to say, if the firm(s) which benefit(s) from the collusive outcome can make transfers to the other members of the cartel so as to make collusion more easily sustainable. I focus on the case where the payments are such that the discount factors are equalized across firms. This is the 'Balanced temptations equilibrium' proposed in Friedman (1971), who introduced this equilibrium as a mechanism for achieving a cooperative outcome in a non-cooperative setting. The idea is that all firms would, in a particular sense, be equally tempted to defect from the tacitly cooperating group of firms, specifically, that they would have the same discount factor. He argued that this was the lowest discount factor capable of sustaining cooperation on the Pareto frontier of the set of feasible profits of the industry. He argued that this was a less extreme solution than that suggested by the maximization of industry profits.

One simple way to implement this equilibrium is by allowing for side payments across firms up to the point that discount factors are uniform, so that firms which benefit from a collusive solution are allowed to make transfers to others and in so doing will generate incentives for other firms to take part in the collusive agreement. In particular, side payments may make cooperation incentive-compatible for every player, thus making the collusive agreement more stable. Specifically, for any given $\delta = \delta_1 = ... = \delta_F$ I can calculate, the net side payments $\lambda_f, f = 1, .., F$ required for each firm to sustain cooperation:

$$\delta_f = \frac{\pi_f^{Def}(\Delta) - \pi_f^{Coll}(\Delta) - \lambda_f}{\pi_f^{Def}(\Delta) - \pi_f^{NE}(\Delta)}.$$

Since side-payments must add to zero, $\sum_{f=1}^{F} \lambda_f = 0$, I can determine both δ and the set of λ's. As always, I can consider both full coalitions and partial coalitions.

Table 3.21 reports the results and shows that the side payments required are generally quite small, less than \$300,000 per quarter in each of the markets and market segments I considered (less than 5% of the static profits). Mechanisms, such as purchases of goods from rival companies, could potentially be used to achieve side payments of this kind of magnitude. The form of the transfers required are themselves interesting. In Panel A, transfers are each from Compaq to its smaller rivals (pre-merger;

and from HP-Compaq post merger) as the small firms must be induced to cooperate. The one interesting exception to this pattern is in Japan in the 4-10 market segment, where Dell is found to need to compensate HP and HP-Compaq.

3.6.3.6 Comparing Discount Factors

One natural question that arises regards the closeness between the critical discount factors and the ones estimated using finance data for a given firm. I address this question by comparing betas instead of discount factors – I back out the betas implied by the critical discount factors and examine whether they are within the 99% confidence interval of the corresponding asset pricing beta. For the great majority of the firms for which I estimate the cost of capital, it is very well approximated by the cost of equity, since firms finance themselves using mostly equity (the only firms financed less than 90% by equity are SGI and Unisys).

The results are reported in Table 3.22 and indicate, in the main, that critical discount factors from my models are, in the main, significantly smaller than the estimated discount factors reflecting the firms cost of capital. In the benchmark model, where are all firms are assumed to have market power and might participate in the collusive scheme, no more than three have an implied beta which falls within the 99% confidence interval for the CAPM beta. By and large, this result is unaffected by multi-market contact and suggests that coordination would appear to be sustainable if the benchmark model is a reasonable approximation to reality – although importantly for any merger assessment that is true both pre- and post-merger. In the main, although critical discount factors rise once I allow for a competitive fringe, indicating that coordination appears harder to sustain in that model, I continue to find that in most instances the critical discount factors lie significantly below the estimated discount factors.

Conclusion

This paper attempts to take the coordination literature following Friedman (1971) seriously and to use it to understand the incentives to tacitly coordinate in a particular differentiated product market, the server market. In doing so, I build on both the literature on repeated games and the literature on unilateral effects merger simulation.

I find that in the benchmark tacitly collusive model (Friedman, 1971) the incentives to collude are substantial, even without assuming sophisticated punishment mechanisms. Although I find that such incentives are affected by the merger between HP and Compaq, and the incentive to cooperate remains substantial, while the merger actively decreases the incentives to tacitly cooperate of non-merging firms in the industry. I show that for non-merging firms, mergers generally reduce the incentive to cooperate in the benchmark model and that this result is not sensitive to the estimates of the differentiated product demand model. Intuitively, the result emerges because the unilateral effects of a merger will mean that Nash equilibrium payoffs will increase following the merger. In so doing, the return to cooperation falls.

I considered a number of generalizations to my benchmark model. First, I considered multi-market contact, following Bernheim and Whinston (1990). In the case of the network server market, accounting for multi-market contact amplifies the incentives to cooperate compared to the benchmark model. However, as before and according to my model, the HP-Compaq merger actively reduces the incentive to cooperate although coordination remains attractive both before and after the merger.

Second, I considered the effect of a competitive fringe. I modeled the presence of a competitive fringe in two ways. First, I considered a set of rival fringe firms who priced at marginal cost both before and after the merger. Second I considered outcomes when tacit coordination between a partial coalition of firms involved setting the Nash prices that would have resulted had those tacitly coordinating firms undertaken a merger. Intuitively, I found that models incorporating a competitive fringe constrained the coordinating firms' ability to raise prices compared to a fully tacitly coordinating industry. However, I showed that individual players can be either better or worse off when only a subset of firms tacitly coordinate. A smaller number of firms following the objective of maximizing total profits of the cooperative group may be able to improve the rewards paid to individual members of the smaller group since the objective of the group and its members are more closely aligned. On the other hand, I found that with a competitive fringe, individual players may be disproportionately affected by the competitive constraint and the partial coalition may choose to sacrifice that firms' profitability for the overall benefit of the group. When simulating the effect of a partial coalition between HP, Compaq, IBM and Dell, I found that Dell was worse off compared to a coalition of all the firms in the industry while each of the others were actually better off. Clearly, such results rely heavily on the definition of

Finally, I note that in modeling tacit coordination by following Friedman (1971) I continue to make a number of very strong assumptions. In particular, my model is silent about entry and exit by firms, or indeed about the entry or exit of products. In addition, I do not account for uncertainty that, for instance, means it may be difficult to detect whether rival firms are tacitly cooperating or not, following authors such as Green and Porter (1984). As is appropriate for a first piece of empirical work in an area, I recognize that these are strong assumptions and expect that future work simulating the coordinated effects of mergers will attempt to address each of these concerns. I also recognize the importance of performance statistical inference, but this is left for further research – in the current version, the theoretical results grouped in Propositions 1 and 2 provide guidance on the changes in profits, value functions and discount factors one should expect.

TABLE 3.9. Pre- and Post-Merger Value Functions – 0-4 Price Segment

	Pre-Merger			Post-Merger			
Firm	**V.Coll**	**V.Def**	**V.(C-D)**	**V.Coll**	**V.Def**	**V.(C-D)**	**%V.(C-D)**
Panel A: US							
Compaq	106869	100043	6826				
HP	31912	29898	2014	138781	129977	8804	-0.4
NEC	2199	2063	136	2199	2066	133	-2.2
IBM	39790	37295	2495	39790	37350	2439	-2.2
AST	4458	4174	285	4458	4180	278	-2.2
Dell	73654	68946	4708	73654	69050	4603	-2.2
Gateway	11738	11017	722	11738	11033	706	-2.2
Hitachi	120	112	8	120	113	7	-2.2
Micron	2813	2639	173	2813	2643	170	-2.2
SGI	60	56	4	60	56	4	-2.2
Sun	3765	3529	236	3765	3534	231	-2.2
Toshiba	3356	3148	208	3356	3152	204	-2.2
Unisys	1368	1281	87	1368	1283	85	-2.2
VA Linux	8563	8023	540	8563	8035	528	-2.2
Panel B: EU							
Compaq	70533	67937	2597				
HP	42705	41184	1521	113238	109239	3999	-2.9
Fujitsu	24634	23765	870	24634	23890	744	-14.4
NEC	2522	2435	87	2522	2448	74	-14.3
IBM	32362	31235	1127	32362	31393	969	-14.0
Data Gen	2	2	0	2	2	0	-14.6
Dell	27958	26973	985	27958	27114	844	-14.3
Gateway	1473	1423	51	1473	1430	44	-14.4
SGI	434	419	15	434	422	13	-14.6
Toshiba	393	379	14	393	381	12	-14.5
VA Linux	111	108	4	111	108	3	-14.4
Panel C: JP							
Compaq	13137	12896	241				
HP	8242	8092	150	21379	20988	391	-0.3
Fujitsu	15001	14725	276	15001	14735	266	-3.7
NEC	17199	16881	318	17199	16893	307	-3.7
IBM	32689	32035	654	32689	32057	632	-3.3
Dell	25466	24976	491	25466	24993	474	-3.4
Gateway	1779	1748	32	1779	1749	31	-3.7
Hitachi	6514	6397	117	6514	6402	112	-3.8
Mitsubishi	2681	2634	48	2681	2635	46	-3.8
Toshiba	5828	5723	105	5828	5727	101	-3.7

Note: Figures are in thousands 1996 US dollars, rounded to the nearest thousandth. The post-merger values for HP report the percentage change of the post-merger values of the merged entity, HP-Compaq, with respect to the combined pre-merger values of HP and Compaq.

TABLE 3.10. Pre- and Post-Merger Value Functions – 4-10 Price Segment

	Pre-Merger			Post-Merger			
Firm	**V.Coll**	**V.Def**	**V.(C-D)**	**V.Coll**	**V.Def**	**V.(C-D)**	**%V.(C-D)**
Panel A: US							
Compaq	236851	214658	22193				
HP	42057	38555	3502	278908	253537	25371	-1.3
NEC	16	15	1	16	15	1	-10.8
IBM	132809	121300	11508	132809	122126	10683	-7.1
Data Gen	63	59	4	63	59	4	-11.4
Dell	582458	509446	73011	582458	512476	69982	-4.1
NCR	9238	8577	661	9238	8646	592	-10.4
Sun	55843	51442	4402	55843	51827	4016	-8.7
Toshiba	943	865	78	943	871	73	-7.3
Panel B: EU							
Compaq	578699	500550	78149				
HP	117528	105432	12095	696227	610075	86152	-4.5
Fujitsu	97581	88005	9576	97581	90822	6758	-29.4
NEC	20357	18288	2069	20357	18787	1570	-24.1
IBM	199119	178054	21065	199119	183484	15634	-25.7
AST	10661	9569	1092	10661	9814	847	-22.4
Dell	203585	181848	21737	203585	186697	16888	-22.3
Gateway	39	35	4	39	36	3	-21.9
Sun	38224	34538	3686	38224	35643	2582	-29.9
Toshiba	1889	1696	193	1889	1741	148	-23.4
Unisys	771	697	74	771	719	52	-29.7
VA Linux	1560	1401	159	1560	1438	122	-23.3
Panel C: JP							
Compaq	72434	60362	12072				
HP	37760	31538	6222	110194	91919	18275	-0.1
Fujitsu	95836	79348	16489	95836	80032	15805	-4.1
NEC	113564	93563	20001	113564	94364	19200	-4.0
IBM	36953	30871	6082	36953	31138	5815	-4.4
AST	740	624	116	740	628	112	-3.8
Dell	38548	32370	6179	38548	32615	5934	-3.9
Gateway	27	23	4	27	23	4	-4.0
Hitachi	20209	16906	3303	20209	17055	3155	-4.5
Mitsubishi	12536	10520	2017	12536	10603	1934	-4.1
NCR	388	327	61	388	329	59	-3.8
Sun	31720	26493	5227	31720	26729	4991	-4.5
Toshiba	29577	24793	4785	29577	25000	4577	-4.3

Note: Figures are in thousands 1996 US dollars, rounded to the nearest thousandth. The post-merger values for HP report the percentage change of the post-merger values of the merged entity, HP-Compaq, with respect to the combined pre-merger values of HP and Compaq.

TABLE 3.11. Pre- and Post-Merger Critical Discount Factors – 0-4 and 4-10 Price Segments

	0-4 Price Segment			4-10 Price Segment		
Firm	**Pre-**	**Post-**	**%Change**	**Pre-**	**Post**	**%Change**
Panel A: US						
Compaq	0.102			0.549		
HP	0.166	0.080	-52.1	0.697	0.530	-23.9
NEC	0.189	0.193	1.8	0.807	0.821	1.8
IBM	0.158	0.161	1.9	0.644	0.660	2.4
AST	0.196	0.199	1.8	–	–	–
Data Gen	–	–	–	0.817	0.832	1.8
Dell	0.130	0.132	1.9	0.240	0.247	3.2
Gateway	0.179	0.182	1.8	–	–	–
Hitachi	0.192	0.196	1.8	–	–	–
Micron	0.187	0.191	1.8	–	–	–
NCR	–	–	–	0.795	0.810	1.8
SGI	0.198	0.202	1.8	–	–	–
Sun	0.191	0.195	1.8	0.733	0.749	2.1
Toshiba	0.189	0.192	1.8	0.711	0.725	1.9
Unisys	0.197	0.201	1.8	–	–	–
VA Linux	0.189	0.192	1.8	–	–	–
Panel B: EU						
Compaq	0.198			0.303		
HP	0.270	0.114	-57.6	0.632	0.246	-61.1
Fujitsu	0.320	0.354	10.6	0.679	0.744	9.6
NEC	0.364	0.399	9.8	0.650	0.706	8.5
IBM	0.291	0.322	10.8	0.596	0.662	10.9
AST	–	–	–	0.626	0.680	8.5
Data Gen	0.380	0.417	9.7	–	–	–
Dell	0.310	0.343	10.7	0.549	0.608	10.6
Gateway	0.371	0.408	9.7	0.623	0.676	8.4
SGI	0.379	0.415	9.7	–	–	–
Sun	–	–	–	0.710	0.771	8.6
Toshiba	0.375	0.411	9.7	0.646	0.701	8.4
Unisys	–	–	–	0.724	0.782	8.0
VA Linux	0.372	0.408	9.7	0.646	0.700	8.4
Panel C: JP						
Compaq	0.522			0.528		
HP	0.530	0.481	-9.2	0.597	0.482	-19.3
Fujitsu	0.527	0.536	1.7	0.510	0.520	2.0
NEC	0.519	0.528	1.7	0.475	0.484	2.0
IBM	0.408	0.416	1.9	0.602	0.612	1.6
AST	–	–	–	0.578	0.587	1.5
Dell	0.455	0.463	1.8	0.556	0.565	1.7
Gateway	0.581	0.589	1.5	0.607	0.616	1.5
Hitachi	0.573	0.582	1.5	0.632	0.642	1.5
Mitsubishi	0.590	0.599	1.5	0.601	0.610	1.5
NCR	–	–	–	0.586	0.595	1.5
Sun	–	–	–	0.618	0.629	1.6
Toshiba	0.563	0.572	1.5	0.606	0.616	1.6

TABLE 3.12. Multimarket Value Functions

	Pre-Merger			Post-Merger			
Firm	**V.Coll**	**V.Def**	**V.(C-D)**	**V.Coll**	**V.Def**	**V.(C-D)**	**%V.(C-D)**
Panel A: 0-4 Price Segment							
Compaq	190539	180875	9664				
HP	82859	79173	3685	273398	260204	13193	-1.2
Fujitsu	39635	38490	1145	39635	38625	1009	-11.8
NEC	21920	21379	541	21920	21407	514	-5.0
IBM	104840	100565	4276	104840	100800	4040	-5.5
AST	4458	4174	284	4458	4180	278	-2.2
Data Gen	2	2	0	2	2	0	-14.6
Dell	127078	120894	6184	127078	121157	5921	-4.2
Gateway	14991	14187	805	14991	14211	780	-3.0
Hitachi	6634	6510	124	6634	6514	119	-3.7
Micron	2813	2639	173	2813	2643	170	-2.2
Mitsubishi	2681	2634	48	2681	2635	46	-3.8
SGI	494	476	19	494	478	17	-12.1
Sun	3765	3529	236	3765	3534	231	-2.2
Toshiba	9576	9250	327	9576	9260	316	-3.2
Unisys	1368	1281	87	1368	1283	85	-2.2
VA Linux	8674	8130	544	8674	8143	532	-2.3
Panel B: 4-10 Price Segment							
Compaq	887984	775570	112414				
HP	197344	175525	21820	1085328	955530	129798	-3.3
Fujitsu	193417	167352	26065	193417	170854	22563	-13.4
NEC	133937	111866	22072	133937	113166	20771	-5.8
IBM	368880	330225	38655	368880	336748	32132	-16.8
AST	11401	10193	1208	11401	10442	959	-20.6
Data Gen	63	59	4	63	59	4	-11.4
Dell	824591	723664	100927	824591	731787	92803	-8.0
Gateway	66	57	9	66	58	8	-12.6
Hitachi	20209	16906	3303	20209	17055	3155	-4.4
Mitsubishi	12536	10520	2017	12536	10603	1934	-4.1
NCR	9626	8904	722	9626	8975	650	-9.8
Sun	125787	112472	13314	125787	114198	11589	-12.9
Toshiba	32409	27354	5056	32409	27612	4797	-5.1
Unisys	771	697	74	771	719	52	-29.7
VA Linux	1560	1401	159	1560	1438	122	-23.3

Note: Figures are in thousands 1996 US dollars, rounded to the nearest thousandth. The post-merger values for HP report the percentage change of the post-merger values of the merged entity, HP-Compaq, with respect to the combined pre-merger values of HP and Compaq.

TABLE 3.13. Multimarket Critical Discount Factors

Firm	Pre-Merger	Post-Merger	%Change
Panel A: 0-4 Price Segment			
Compaq	0.149		
HP	0.236	0.111	-52.9
Fujitsu	0.386	0.415	7.6
NEC	0.441	0.454	2.8
IBM	0.245	0.255	4.2
AST	0.196	0.199	1.8
Data Gen	0.38	0.417	9.7
Dell	0.202	0.209	3.4
Gateway	0.224	0.230	2.4
Hitachi	0.561	0.570	1.6
Micron	0.187	0.191	1.8
Mitsubishi	0.59	0.599	1.5
SGI	0.349	0.379	8.3
Sun	0.191	0.195	1.8
Toshiba	0.373	0.381	2.0
Unisys	0.197	0.201	1.8
VA Linux	0.19	0.194	1.9
Panel B: 4-10 Price Segment			
Compaq	0.400		
HP	0.636	0.364	-42.8
Fujitsu	0.591	0.623	5.5
NEC	0.499	0.514	2.9
IBM	0.613	0.653	6.5
AST	0.622	0.672	7.9
Data General	0.817	0.832	1.8
Dell	0.364	0.384	5.2
Gateway	0.615	0.645	4.8
Hitachi	0.632	0.642	1.5
Mitsubishi	0.601	0.61	1.5
NCR	0.786	0.801	1.8
Sun	0.691	0.717	3.8
Toshiba	0.61	0.622	1.9
Unisys	0.724	0.782	8.0
VA Linux	0.646	0.700	8.4

TABLE 3.14. Multimarket Value Functions and Critical Discount Factors – Aggregation Across Regions and Price Segments

	Pre-Merger			Post-Merger			
Firm	**V.Coll**	**V.Def**	**V.(C-D)**	**V.Coll**	**V.Def**	**V.(C-D)**	**%V.(C-D)**
US/EU/JP Markets							
Compaq	1078523	956446	122077				
HP	280203	254698	25505	1358726	1215735	142991	-3.1
Fujitsu	233052	205842	27210	233052	209479	23572	-13.3
NEC	155858	133245	22613	155858	134573	21285	-5.8
IBM	473721	430790	42931	473721	437549	36172	-15.7
AST	15859	14366	1493	15859	14622	1237	-17.1
Data Gen	65	61	4	65	61	4	-11.4
Dell	951669	844558	107111	951669	852944	98725	-7.8
Gateway	15057	14244	813	15057	14270	787	-3.1
Hitachi	26843	23416	3427	26843	23569	3274	-4.4
Micron	2813	2639	174	2813	2643	170	-2.2
Mitsubishi	15218	13153	2064	15217	13237	1980	-4.1
NCR	9626	8904	722	9626	8976	650	-9.8
SGI	494	476	18	494	478	17	-12.1
Sun	129551	116001	13551	129551	117732	11820	-12.7
Toshiba	41985	36603	5383	41985	36872	5114	-4.9
Unisys	2138	1977	161	2138	2001	137	-14.9
VA Linux	10235	9531	703	10235	9581	653	-7.1

Critical Discount Factors

	Pre-Merger	Post-Merger	%Change
US/EU/JP Markets			
Compaq	0.386		
HP	0.608	0.347	-42.9
Fujitsu	0.585	0.618	5.5
NEC	0.498	0.512	2.9
IBM	0.594	0.632	6.4
AST	0.582	0.624	7.3
Data Gen	0.816	0.831	1.8
Dell	0.357	0.375	5.1
Gateway	0.233	0.238	2.4
Hitachi	0.629	0.639	1.5
Micron	0.187	0.191	1.8
Mitsubishi	0.601	0.610	1.5
NCR	0.786	0.801	1.8
SGI	0.349	0.378	8.3
Sun	0.687	0.714	3.8
Toshiba	0.601	0.613	1.9
Unisys	0.582	0.619	6.3
VA Linux	0.379	0.396	4.5

Note: Value function figures are in thousands 1996 US dollars, rounded to the nearest thousandth. The post-merger values for HP report the percentage change of the post-merger values of the merged entity, HP-Compaq, with respect to the combined pre-merger values of HP and Compaq.

TABLE 3.15. Competitive Fringe Model Value Functions – 0-4 and 4-10 Price Segments

	Pre-Merger			Post-Merger			
Firm	**V.Coll**	**V.Def**	**V.(C-D)**	**V.Coll**	**V.Def**	**V.(C-D)**	**%V.(C-D)**
Panel A: 0-4 Price Segment							
Region: US							
Compaq	108031	100078	7954				
HP	32263	29908	2355	140294	130022	10272	-0.4
Dell	74469	68970	5500	74469	69074	5395	-1.9
IBM	40222	37307	2915	40222	37363	2859	-1.9
Region: EU							
Compaq	71686	67970	3716				
HP	43414	41204	2210	115100	109295	5805	-2.0
Dell	28429	26986	1443	28429	27127	1302	-9.8
IBM	32882	31250	1632	32882	31408	1474	-9.7
Region: JP							
Compaq	13840	12915	924				
HP	8663	8104	559	22502	21020	1482	-0.1
Dell	26834	25015	1819	26834	25032	1802	-0.9
IBM	34430	32085	2345	34430	32107	2323	-0.9
Panel B: 4-10 Price Segment							
Region: US							
Compaq	241314	214765	26549				
HP	42904	38576	4328	284218	253658	30560	-1.0
Dell	593518	509721	83797	593518	512740	80777	-3.6
IBM	135572	121368	14204	135572	122190	13382	-5.8
Region: EU							
Compaq	603682	501159	102523				
HP	122599	105558	17041	726281	610710	115572	-3.3
Dell	211209	182034	29174	211209	186847	24361	-16.5
IBM	208246	178274	29972	208246	183663	24583	-18.0
Region: JP							
Compaq	92046	60842	31205				
HP	48837	31817	17020	140883	92611	48272	0.1
Dell	47938	32607	15331	47938	32837	15101	-1.5
IBM	47941	31148	16793	47941	31399	16542	-1.5

Note: Figures are in thousands 1996 US dollars, rounded to the nearest thousandth. The post-merger values for HP report the percentage change of the post-merger values of the merged entity, HP-Compaq, with respect to the combined pre-merger values of HP and Compaq.

TABLE 3.16. Competitive Fringe Model Critical Discount Factors – 0-4 and 4-10 Price Segments

Firm	Pre-Merger	Post-Merger	%Change
Panel A: 0-4 Price Segment			
Region: US			
Compaq	0.076		
HP	0.132	0.057	-56.5
Dell	0.100	0.102	1.7
IBM	0.125	0.127	1.6
Region: EU			
Compaq	0.122		
HP	0.177	0.059	-66.5
Dell	0.209	0.226	8.3
IBM	0.196	0.212	8.3
Region: JP			
Compaq	0.157		
HP	0.170	0.131	-23.0
Dell	0.118	0.119	0.8
IBM	0.096	0.097	0.8
Panel B: 4-10 Price Segment			
Region: US			
Compaq	0.495		
HP	0.649	0.473	-27.1
Dell	0.196	0.202	2.9
IBM	0.590	0.604	2.3
Region: EU			
Compaq	0.214		
HP	0.541	0.156	-71.1
Dell	0.461	0.505	9.4
IBM	0.495	0.543	9.6
Region: JP			
Compaq	0.225		
HP	0.294	0.163	-44.4
Dell	0.277	0.280	1.0
IBM	0.297	0.300	1.0

TABLE 3.17. Multimarket Value Functions and Critical Discount Factors in the Competitive Fringe Model – 0-4 and 4-10 Price Segments

	Pre-Merger			Post-Merger			
Firm	**V.Coll**	**V.Def**	**V.(C-D)**	**V.Coll**	**V.Def**	**V.(C-D)**	**%V.(C-D)**
Panel A: 0-4 Price Segment							
Compaq	193557	180963	12594				
HP	84339	79216	5124	277896	260337	17559	-0.9
Dell	129732	120970	8762	129732	121233	8499	-2.9
IBM	107534	100642	6892	107534	100878	6656	-3.4
Panel B: 4-10 Price Segment							
Compaq	937043	776766	160277				
HP	214339	175951	38389	1151382	956979	194404	-2.2
Dell	852664	724363	128302	852664	732425	120239	-6.2
IBM	391760	330790	60970	391760	337252	54507	-10.5

	Critical Discount Factors		
Firm	**Pre-Merger**	**Post-Merger**	**%Change**
Panel C: 0-4 Price Segment			
Compaq	0.097		
HP	0.156	0.065	-2.0
Dell	0.124	0.127	2.6
IBM	0.134	0.138	3.0
Panel D: 4-10 Price Segment			
Compaq	0.284		
HP	0.480	0.232	-51.7
Dell	0.287	0.300	4.6
IBM	0.484	0.511	5.5

Note: Value function figures are in thousands 1996 US dollars, rounded to the nearest thousandth. The post-merger values for HP report the percentage change of the post-merger values of the merged entity, HP-Compaq, with respect to the combined pre-merger values of HP and Compaq.

TABLE 3.18. Multimarket Value Functions and Critical Discount Factors in the Competitive Fringe Model – Aggregated Across Price Segments and Regions

	Pre-Merger			Post-Merger			
Firm	**V.Coll**	**V.Def**	**V.(C-D)**	**V.Coll**	**V.Def**	**V.(C-D)**	**%V.(C-D)**
Compaq	1130600	957729	172871				
HP	298679	255166	43512	1429278	1217316	211962	387.1
Dell	982396	845333	137064	982396	853658	128739	-6.1
IBM	499294	431432	67861	499294	438130	61164	-9.9

Critical Discount Factors			
Firm	**Pre-Merger**	**Post-Merger**	**%Change**
Compaq	0.273		
HP	0.456	0.220	-51.7
Dell	0.278	0.291	4.5
IBM	0.463	0.488	5.4

Note: Value function figures are in thousands 1996 US dollars, rounded to the nearest thousandth. The post-merger values for HP report the percentage change of the post-merger values of the merged entity, HP-Compaq, with respect to the combined pre-merger values of HP and Compaq.

TABLE 3.19. Value Functions for the Compaq-HP-IBM Cartel

	Pre-Merger			Post-Merger			
Firm	**V.Coll**	**V.Def**	**V.(C-D)**	**V.Coll**	**V.Def**	**V.(C-D)**	**%V.(C-D)**
Panel A: 0-4 MarketSegment							
Region: US							
Compaq	99956.5	99790.6	166.0				
HP	29831.1	29817.9	13.2	129787.6	129654.5	133.1	-25.7
IBM	37216.5	37196.3	20.1	37216.5	37252.0	-35.6	
Region: EU							
Compaq	68137.2	67840.8	296.4				
HP	41242.9	41122.3	120.6	109380.0	109095.9	284.1	-31.9
IBM	31271.1	31188.7	82.4	31271.1	31346.7	-75.7	
Region: JP							
Compaq	12887.4	12879.1	8.2				
HP	8085.8	8081.4	4.4	20973.2	20963.3	9.8	-22.1
IBM	32067.8	31999.6	68.3	32067.8	32021.3	46.6	-31.8
US/EU/JP Markets							
Compaq	180981.1	180510.4	470.6				
HP	79159.7	79021.5	138.2	260140.8	259713.7	427.1	-29.9
IBM	100555.4	100384.6	170.8	100555.4	100620.0	-64.6	
Panel B: 4-10 Market Segment							
Region: US							
Compaq	216201.9	213049.6	3152.3				
HP	38398.4	38162.9	235.5	254600.3	251742.8	2857.5	-15.7
IBM	121300.5	120210.1	1090.4	121300.5	121035.5	265.1	-75.7
Region: EU							
Compaq	519135.5	497590.5	21545.0				
HP	105410.7	104515.1	895.6	624546.1	606829.3	17716.8	-21.1
IBM	178429.6	176582.1	1847.5	178429.6	182012.7	-3583.2	
Region: JP							
Compaq	60042.2	59481.8	560.4				
HP	31202.7	30998.9	203.8	91244.9	90697.8	547.1	-28.4
IBM	30533.9	30337.5	196.4	30533.9	30604.8	-70.9	
US/EU/JP Markets							
Compaq	795379.6	770121.9	25257.7				
HP	175011.8	173676.9	1334.9	970391.4	949270.0	21121.4	-20.6
IBM	330264.0	327129.7	3134.3	330264.0	333653.0	-3389.0	
Panel C: Multimarket Contact Across Geographical Markets and Price Segments							
Compaq	976360.7	950632.3	25728.4				
HP	254171.5	252698.4	1473.2	1230532.2	1208983.7	21548.5	-20.8
IBM	430819.4	427514.3	3305.1	430819.4	434273.0	-3453.6	

Note: Figures are in thousands 1996 US dollars, rounded to the nearest thousandth. The post-merger values for HP report the percentage change of the post-merger values of the merged entity, HP-Compaq, with respect to the combined pre-merger values of HP and Compaq.

TABLE 3.20. Value Functions for the Compaq-HP-Dell Cartel

	Pre-Merger			Post-Merger			
Firm	**V.Coll**	**V.Def**	**V.(C-D)**	**V.Coll**	**V.Def**	**V.(C-D)**	**%V.(C-D)**
Panel A: 0-4 Market Segment							
Region: US							
Compaq	100022.2	99794.9	227.3				
HP	29850.4	29819.8	30.7	129872.7	129659.2	213.5	-17.2
Dell	68877.7	68769.0	108.7	68877.7	68873.6	4.1	-96.2
Region: EU							
Compaq	68126.9	67839.9	286.9				
HP	41236.6	41121.7	115.0	109363.5	109095.1	268.5	-33.2
Dell	26990.4	26930.4	60.0	26990.4	27071.6	-81.2	
Region: JP							
Compaq	12890.0	12878.9	11.1				
HP	8087.2	8081.2	6.0	20977.2	20963.0	14.2	-17.2
Dell	24987.9	24946.1	41.8	24987.9	24963.1	24.8	-40.7
US/EU/JP Markets							
Compaq	181039.1	180513.8	525.4				
HP	79174.3	79022.6	151.7	260213.4	259717.2	496.2	-26.7
Dell	120856.0	120645.6	210.5	120856.0	120908.4	-52.4	
Panel B: 4-10 Market Segment							
Region: US							
Compaq	214256.2	213304.4	951.9				
HP	37977.0	38241.8	-264.8	252233.3	252000.9	232.3	-66.2
Dell	526838.0	506984.0	19854.0	526838.0	510013.5	16824.5	-15.3
Region: EU							
Compaq	516128.5	497464.0	18664.6				
HP	104796.7	104476.6	320.1	620925.2	606690.4	14234.7	-25.0
Dell	182301.2	180463.6	1837.6	182301.2	185312.4	-3011.1	
Region: JP							
Compaq	59973.7	59478.9	494.8				
HP	31167.1	30997.1	170.0	91140.9	90693.9	446.9	-32.8
Dell	32074.9	31888.2	186.7	32074.9	32133.0	-58.1	
US/EU/JP Markets							
Compaq	790358.5	770247.2	20111.2				
HP	173940.8	173715.5	225.4	964299.3	949385.3	14914.0	-26.7
Dell	741214.1	719335.8	21878.3	741214.1	727458.9	13755.3	-37.1
Panel C: Multimarket Contact Across Geographical Markets and Price Segments							
Compaq	971397.6	950761.0	20636.6				
HP	253115.1	252738.1	377.0	1224512.7	1209102.5	15410.2	-26.7
Dell	862070.2	839981.4	22088.8	862070.2	848367.3	13702.9	-38.0

Note: Figures are in thousands 1996 US dollars, rounded to the nearest thousandth. The post-merger values for HP report the percentage change of the post-merger values of the merged entity, HP-Compaq, with respect to the combined pre-merger values of HP and Compaq.

TABLE 3.21. Side Payments to Sustain Uniform Critical Discount Factors Across Firms

Firm	0-4 Market Segment Pre-Merger	0-4 Market Segment Post-Merger	4-10 Market Segment Pre-Merger	4-10 Market Segment Post-Merger
Panel A: Compaq-HP-IBM Cartel				
Region: US				
Compaq	-1.9		-25.0	
HP	0.9	-1.9	12.5	-25.0
IBM	0.9	1.9	12.5	25.0
Region: EU				
Compaq	-2.4		-248.6	
HP	1.0	-4.0	108.1	-229.4
IBM	1.3	4.0	140.5	229.4
Region: JP				
Compaq	0.4		-4.1	0.0
HP	0.3	0.4	2.0	-6.3
IBM	-0.8	-0.4	2.1	6.3
US/EU/JP Markets				
Compaq	-3.8		-277.1	
HP	2.3	-5.5	123.0	-259.8
IBM	1.5	5.5	154.1	259.8
Panel B: Compaq-HP-Dell Cartel				
Region: US				
Compaq	-1.8		165.9	
HP	1.3	-2.3	66.1	181.8
Dell	0.6	2.3	-232.0	-181.8
Region: EU				
Compaq	-2.4		-214.8	
HP	0.9	-3.9	108.4	-184.8
Dell	1.5	3.9	106.4	184.8
Region: JP				
Compaq	0.2		-3.5	
HP	0.2	0.1	2.2	-5.2
Dell	-0.4	-0.1	1.4	5.2
US/EU/JP Markets				
Compaq	-4.1		-53.2	
HP	2.4	-6.1	176.3	-10.2
Dell	1.7	6.1	-123.1	10.2

Note: Figures are in thousands 1996 US dollars per quarter.

TABLE 3.22. Rejections of the Null Hypothesis of Equality Between Model-Implied and CAPM Betas Against the Two-Sided Alternative

	0-4 Price Segment		4-10 Price Segment	
	Pre-Merger	**Post-Merger**	**Pre-Merger**	**Post-Merger**
Panel A: Std model				
US	12/12	11/11	7/9 Compaq, Toshiba	6/8 HP, Toshiba
EU	8/9 SGI	6/8 Gateway, SGI	10/10	7/9 Fujitsu, NEC
JP	8/10 Compaq, Dell	7/9 HP, Dell	10/12 Compaq, NCR	9/11 HP, NCR
Panel B: MM model				
All Regions	14/14	12/13 SGI	11/13 Compaq, Dell	11/12 Dell
Panel C: CF model				
US	4/4	3/3	3/4 Compaq	2/3 HP
EU	4/4	3/3	2/4 Dell, IBM	3/3
JP	4/4	3/3	4/4	3/3
Panel D: CF-MM model				
All Regions	4/4	4/3	2/4 HP IBM	2/3 IBM

	Pre-Merger	Post-Merger
All Regions & Price Segments	13/15 Dell, Unisys	11/14 Dell, SGI, , Dell
Panel F: CF-MM model		
Al Regions & Price Segments	2/4 HP, IBM	2/3 IBM

Note: The Table reports the number of rejections at the 1% significance level of the null hypothesis of equality between the model-implied and CAPM betas against the two-sided alternative. For a given combination of rows and columns, the first number is the number of rejections, whereas the second number is the number of firms which operate in the market and for which the CAPM was estimated. Whenever the null was not rejected, the name of the firm is reported.

References

[1] Abreu, D. (1986), "Extremal Equilibria of Oligopolistic Supergames," Journal of Economic Theory, 39(1), 191-225.

[2] Abreu, D. (1988), "On the Theory of Infinitely Repeated Games with Discounting," Econometrica 56 (2), 383-396.

[3] Abreu, D., D. G. Pearce and E. Stacchetti (1986) "Optimal Cartel Equilibria with Imperfect Monitoring", Journal of Economic Theory, 39(1).

[4] Abreu, D., D. G. Pearce and E. Stacchetti (1990) "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring", Econometrica 58, 1041-1063.

[5] Aumann, R. (1986) "Repeated Games." In Issues in Contemporary Microeconomics, ed. G. Feiwel. Macmillan.

[6] Aumann, R. (1989) "Survey of repeated games." In Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern. Manheim: Bibliographisches Institut.

[7] Baker, J.B. and T.F. Bresnahan (1985), "The Gains from Merger or Collusion in Product differentiated Indistries", Journal of Industrial Economics, 33, 427-44.

[8] Berk, J. and P. DeMarzo (2007), "Corporate Finance", Pearson.

[9] Bernheim, D. B., and M. D. Whinston (1990), "Multimarket contact and collusive behavior", RAND Journal of Economics 21 (1990), 1 26.

[10] Berry, S. (1994) "Estimating discrete choice models of product differentiation" RAND Journal of Economics, 25, 2, 242-262

[11] Berry, S., J. Levinsohn and A. Pakes (1995) "Automobile prices in Market Equilibrium" Econometrica 63, 841-890

[12] Berry, S., J. Levinsohn and A. Pakes (2004) "Differentiated Product demand systems from a combination of micro and macro data: The new car market", Journal of Political Economy 112, 68-105.

[13] Bonnet, C. and P. Dubois (2008), "Inference on Vertical Contracts between Manufacturers and Retailers Allowing for Non Linear Pricing and Resale Price Maintenance", working paper.

[14] Brenkers, R. and F Verboven (2006), "Liberalizing a Distribution System: the European Car Market," Journal of the European Economic Association, 4(1), 216-251.

[15] Bresnahan, T. (1982) "The Oligopoly Solution Concept is Identified," Economics Letters, vol. 10, pp. 87-92.

[16] Bresnahan, T. (1986) "Measuring the spillovers from technical advance: Mainframe computers in financial services" American Economic Review, 76, 4, 742-755

[17] Bresnahan, T. (1987) "Competition and Collusion in the American Automobile Market: The 1955 Price War", Journal of Industrial Economics, vol. XXXV (4), 457-482.

[18] Bresnahan, T. and S. Greenstein (1996) "Technical Progress and Co-invention in computing and in the uses of computers" Brooking Papers on Economic Activity: Microeconomics, 1-78

[19] Bresnahan, T., S. Stern and M. Trajtenberg (1997) "Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s" RAND Journal of Economics 28, 0, S17-S44

[20] Brock, W. A., and J. A. Scheinkman (1985), "Price Setting Supergames with Capacity Constraints", Review of Economic Studies 52(3), 371-82.

[21] Carhart, M. 1997, "On Persistence in Mutual Fund Performance". Journal of Finance 45 (5):57-82.

[22] Chamberlin E. (1929) "Duopoly: Value where sellers are few." Quarterly Journal of Economics Vol. 43: pages 63-100.

[23] Chan, L.K., N. Jegadeesh and J. Lakonishok, "Momentum strategies", Journal of Finance, Vol. 51, 1996, pp. 1681-1714.

[24] Compte, O. Jenny, F and Rey, P. (2002) "Capacity Constraints, Mergers, and Collusion" European Economic Review, Vol. 46, 1-29.

[25] Corts, K. S. (1999) "Conduct Parameters and the Measurement of Market Power," Journal of Econometrics 88 (2), 227-250.

[26] Davidson, C. and R.J. Deneckere (1984a), "Horizontal Mergers and Collusive Behavior," International Journal of Industrial Organization, 2:117-132.

[27] Davidson, C., and R. J. Deneckere (1984b) "Excess Capacity and Collusion", International Economic Review, 31(3), pages 521-41.

[28] Davidson, C. and Deneckere, R. J. (1985) "Incentives to form Coalitions with Bertrand Competition" RAND Journal of Economics, 16(4).

[29] Davis, P. J., and P. Sabbatini (2007), "How to Simulate the Coordinated Effects of a Merger", mimeo.

[30] European Commission (2004) Judgment on European Commission vs. Microsoft, Brussels, 24th March 2004.

[31] Fama, E. and K.R. French (1992), "The Cross-Section of Expected Stock Returns", Journal of Finance,Vol. 47, pp.427-465

[32] Fama, E., and K. French (1993), Common Risk Factors in the Returns on Stocks and Bonds, Journal of Financial Economics, Vol. 33, pp. 3-53

[33] Fama, E. and K.R. French (2001), "Multifactor Explanations of Asset Pricing Anomalies", Journal of Finance, Vol. 51, pp.55-84.

[34] Farrell, J. and C. Shapiro (1990), "Horizontal Mergers: An Equilibrium Analysis", American Economic Review, 80, 107-25.

[35] Foncel, J. and Ivaldi, M. (2001) "Operating System Prices in the Home PC market" mimeo Institut D'Economie Industrielle, May.

[36] Friedman J. (1971), "A non-cooperative equilibrium for supergames." Review of Economic Studies, Vol. 38., pages 1-12.

[37] Gartner Group (1999). Gartner Dataquest, February 1999.

[38] Gartner Group (2000). Gartner Dataquest, February 2000.

[39] Gartner Group (2001). Gartner Dataquest, February 2001.

[40] Gartner Group (2002). Gartner Dataquest, February 2002.

[41] Gartner Group (2003). Gartner Dataquest, February 2003.

[42] Gartner Group (2008). Gartner Dataquest, February 2008.

[43] Gollop, F. and M. Roberts (1979). Firm interdependence in oligopolistic markets. Journal of Econometrics 10, 313-331

[44] Gorman, T. (1970) "Two Stage Budgeting", mimeo University of North Carolina reprinted in C. Blackorby and A. F. Shorrocks (eds) Separability and Aggregation Volume I: The Collected Works of Terrance Gorman, Oxford: Oxford University Press

[45] Green E. J and Porter, R. (1984), "Noncooperative Collusion Under Imperfect Price Information", Econometrica, Vol. 52, No. 1, January 1984; 87-100.

[46] Hausman, J., G. Leonard and D. Zona (1994) "Competitive Analysis with Differentiated Products" Annales d'Economie et de Statistique, 34, 159-180

[47] Hausman, J. (1997) "Valuation of New Goods under perfect and imperfect competition" Chapter 5 in Timothy Bresnahan and Robert Gordon (eds) The Economics of New Goods, Chicago: University of Chicago Press

[48] International Data Corporation (1998) "IDC's Server Market Methodology, Database Structure, and Definitions", IDC report #15909R, April 1998

[49] International Data Corporation (2000a) "Server Operating Environments Market Forecast and Analysis, 2000-2004", IDC report #22347, June 2000

[50] International Data Corporation (2000b) "Server Market Forecast and Analysis, 1996-2004", IDC report #22538, July 2000, pp.11-14

[51] International Data Corporation (2002) Delivering value in the Volume Server Market - Sun's V480 Attacks the 4-Way Space.

[52] International Data Corporation (2007) Quad-core Processores Bring higher Performance and Lower Cost to Mainstream Computing, White Paper #206309, April 2007.

[53] Ivaldi, M and S. Lörincz (2005), "A Full Equilibrium Relevant Market Test: Application to Computer Servers", CEPR Discussion Paper No. 4917.

[54] Ivaldi, M. and Verboven, F. (2001) “Quantifying the effects of horizontal mergers in European Competition Policy” Centre for Economic Policy Research Discussion Paper No. 2697

[55] Jegadeesh, N. and S. Titman (1993), “Returns to buying Winners and selling Losers: Implications for stock market efficiency”, Journal of Finance, Vol. 48, pp. 65-91.

[56] Kovacic, W. E., R. C. Marshall, L. M. Marx, and S. P. Schulenberg (2007), "Coordinated Effects in Merger Review: Quantifying the Payoffs from Collusion", mimeo Duke University.

[57] Kühn, K-U. (2004), “The Coordinated Effects of Mergers in Differentiated Procuct Markets” John M Olin Center for Law and Economics, Working Paper Series, paper 34.

[58] Lambson, V. E. (1987), “Optimal Penal Codes in Price-Setting Supergames with Capacity Constraints,” Review of Economic Studies, 54:385-397.

[59] Lambson, V. E. (1994), “Some Results on Optimal Penal Codes in Asymmetric Bertrand Supergames,” Journal of Economic Theory, 62:444-468.

[60] Lambson, V. E. (1996), “Optimal Penal Codes in Nearly Symmetric Bertrand Supergames with Capacity Constraints,” Journal of Mathematical Economics.

[61] Lehrer, E. and A. Pauzner (1999) "Repeated games with differential time preferences", Econometrica, 67, 393-412.

[62] Mertens, J-F. (1987) “Repeated Games.” In Proceedings of the International Congress of Mathematicians.

[63] Motta, M. (2004) "Competition Policy: Theory and Practice", Cambridge University Press.

[64] Nakanishi, M and L. G. Cooper (1974) “Parameter Estimate for Multiplicative Interactive Choice Model: Least Squares Approach “ Journal of Marketing Research, Vol 11, pages 303-311.

[65] Nevo, A. (1998) “Identification of the Oligopoly Solution Concept in a Differentiated Products Industry,”Economics Letters, 59(3), 391-395.

[66] Nevo, A. (2000) “Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry,” The RAND Journal of Economics, 31(3), 395-421.

[67] Nevo A. (2001) “Measuring Market Power in the Ready to Eat Cereal Industry” Econometrica, 69, 307-342.

[68] Porter, R. (1983) "A Study of Cartel Stability: The Joint Executive Committee, 1880-1886", Bell Journal of Economics, Vol. 15, No. 2, Autumn 1983; 301-14.

[69] Roberts, M. (1983). Testing oligopolistic behavior: An application of the variable profit function. International Journal of Industrial Organiz

[70] Rosen, S. (1974) "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition" Journal of Political Economy, 82, 34-55

[71] Rotemberg, J. and G. Saloner (1986) "A Supergame-theoretic Model of Business Cycle and Price Wars during Booms", American Economic Review, 76, pp.390-407

[72] Sabbatini, P. (2004) "How to Simulate the Coordinated Effect of a Merger" Working paper, Italian Antitrust Authority, November 2004.

[73] Salvo, A. (2004) "Inferring Conduct Under the Threat of Entry: The Case of the Brazilian Cement Industry" mimeo, LSE.

[74] Scheffman, D. and M. Coleman (2003) "Quantitative Analysis of Potential Competitive Effects from a Merger" George Mason Univ. Law Review, Winter Antitrust Symposium.

[75] Selten, R. (1965). Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit, (Zeitschrift für die Gesamte Staatswissenschaft 121, 301-24 and 667-89, 1965).

[76] Slade, M. E. (2002). "Market power and joint dominance in UK brewing." Journal of Industrial Economics, Vol. LII, No. 1.

[77] Stavins, J. (1997) "Estimating Demand elasticities in a differentiated product industry: The Personal Computer market" Journal of Economics and Business, 49, 347-367

[78] Stigler, G. J.(1964) "A Theory of Oligopoly." Journal of Political Economy, 72: pages 44-61.

[79] Suslow, V. (1986). "Estimating monopoly behaviour with competitive application to Alcoa." Rand Journal of Economics 17, 389-403

[80] Sybex (2001) Networking Complete, San Francisco: Sybex.

[81] Tirole, J. (1988), The Theory of Industrial Organization, Cambridge: The MIT Press.

[82] United States Courts of Appeal (2001) United States vs. Microsoft Corporation, No. 00-4212

[83] Van Reenen, J. (2006), The Growth of Networking Computing: Quality-adjusted Price changes for Network Servers, The Economic Journal 116, F29-F44.

[84] Villas-Boas, S. B. (2007), Vertical Relationships Between Manufacturers and Retailers: Inference With L imited Data, The Review of Economic Studies, 2007, Vol. 74, 2, pp. 625-652.

[85] Werden, G. and L. Froeb (1994) "The Effects of Mergers in Differentiated Products Industries: Structural Merger Policy and the Logit Model" Journal of Law, Economics, & Organization, 10 (1994) pp. 407-426.

[86] White, H. (1980) "A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedacity" Econometrica, 48, 817-838.