UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

EFFECTS OF THE INTRODUCTION OF SPATIAL AND TEMPORAL COMPLEXITY ON THE OPTIMAL DESIGN, ECONOMIES OF SCALE AND PRICING OF PUBLIC TRANSPORT

TESIS PARA OPTAR AL GRADO DE DOCTOR EN SISTEMAS DE INGENIERÍA

ANDRÉS SALOMÓN FIELBAUM SCHNITZLER

PROFESOR GUÍA:
SERGIO JARA DÍAZ

MIEMBROS DE LA COMISIÓN:
LEONARDO BASSO SOTZ
ANTONIO GSCHWENDER KRAUSE
RALF BORNDÖRFER

SANTIAGO DE CHILE
2019

RESUMEN DE LA TESIS PARA OPTAR AL
GRADO DE: Doctor en Sistemas de Ingeniería
POR: Andrés Salomón Fielbaum Schnitzler
FECHA: 19/06/2019
PROFESOR GUÍA: Sergio Rodolfo Jara Díaz

EFFECTS OF THE INTRODUCTION OF SPATIAL AND TEMPORAL COMPLEXITY ON THE OPTIMAL DESIGN, ECONOMIES OF SCALE AND PRICING OF PUBLIC TRANSPORT

En esta tesis estudiamos modelos microeconómicos para el diseño estratégico de transporte público de buses, incorporando los efectos que implican tanto la composición espacial de la demanda por viajes y la necesidad de representarla en una red, como la heterogeneidad entre la cantidad de viajes realizados en distintos períodos del día. Esto se realiza complejizando espacial y temporalmente los modelos clásicos de una línea estudiados por Jansson (1980) y Jara-Díaz y Gschwender (2009).

Para el análisis espacial, estudiamos el diseño óptimo de estructuras de línea (es decir, el conjunto de rutas de las líneas de transporte público) sobre el modelo urbano propuesto por Fielbaum *et al* (2016, 2017) –basado en la jerarquía entre los centros de la ciudad- y analizamos los resultados del enfoque heurístico, la presencia de economías de escala y sus fuentes, y la densidad espacial de líneas.

Respecto al enfoque heurístico, comparamos las cuatro estructuras básicas propuestas por Fielbaum *et al* (2016) con las resultantes de cuatro heurísticas propuestas previamente en la literatura. Los fenómenos de escala se analizan bajo la definición del concepto de “directness”, que muestra que al aumentar el flujo de pasajeros el sistema prioriza rutas que minimicen los trasbordos, detenciones y los largos de los viajes de los pasajeros, es decir, ésta es una nueva fuente de economías de escala; esto permite estudiar los efectos de este fenómeno en tarifas y subsidios óptimos. Cuando la densidad espacial de líneas se incorpora como variable de diseño, se muestra que ésta crece con el número de pasajeros, manteniendo siempre los costos de acceso iguales a los costos de espera en el sistema, mostrando cierto nivel de sustitución con el nivel de “directness” y constituyendo una nueva fuente de economías de escala.

La heterogeneidad temporal de la demanda se analiza al estudiar los modelos de una línea incluyendo dos períodos: punta y fuera de punta. El sistema se optimiza bajo distintas maneras de operación, como son el considerar una flota única, una flota independiente para cada período y dos flotas que operan de manera conjunta en el período punta (y sólo una de ellas en fuera de punta); el sistema con dos flotas simultáneas es el más eficiente, siendo ligeramente mejor que el de una sola flota. Las soluciones se comparan con aquellas que se obtienen al considerar solamente un período, y los efectos cruzados entre períodos son identificados. Adicionalmente, se estudian estrategias de tipo *second-best*, al comparar la optimización del sistema de acuerdo a las características del período punta, y la utilización de una sub-flota para el período fuera de punta, con la estrategia inversa: como resultado, una regla aproximada es priorizar aquél período en que el número total de pasajeros (en toda su duración) sea mayor.

RESUMEN DE LA TESIS PARA OPTAR AL
GRADO DE: Doctor en Sistemas de Ingeniería
POR: Andrés Salomón Fielbaum Schnitzler
FECHA: 19/06/2019
PROFESOR GUÍA: Sergio Rodolfo Jara Díaz

EFFECTS OF THE INTRODUCTION OF SPATIAL AND TEMPORAL COMPLEXITY ON THE OPTIMAL DESIGN, ECONOMIES OF SCALE AND PRICING OF PUBLIC TRANSPORT

In this thesis we study microeconomic models for the strategic design of buses transit systems, taking into account the distribution of trips in space and its representation over a simplified but meaningful urban network, as well as the heterogeneity of demand across different periods of the day. This is done by means of models that extend in space and time the classic single-line ones and analyzed by Jansson (1980) and Jara-Díaz and Gschwender (2009).

Regarding the spatial analysis, we study the optimal lines structure (i.e. the spatial arrangement of transit routes) design over the urban model proposed by Fielbaum *et al* (2016, 2017) based on the hierarchy of centers, obtaining and analyzing the results from the application of existing heuristics, examining the presence of scale economies and its causes, and introducing the spatial density of transit lines as part of the design.

The heuristic approach is studied by comparing the four basic structures proposed by Fielbaum *et al* (2016) with those that emerge when applying four previously existing heuristics. Scale effects are analyzed defining the concept of “directness”, showing that when the number of passengers increases, the best system evolves such that routes reduce transfers, bus stops and the length of passengers’ routes. Directness is shown to be yet another source of scale economies; optimal subsidies and fares are also studied. When spatial density is considered as a new design variable together with lines structure, frequencies and vehicle sizes, it increases with patronage keeping access and waiting time costs equal, showing some substitution with directness and inducing scale economies as well.

The heterogeneity of demand across different periods is analyzed using single-line models that consider peak and off-peak conditions regarding duration, trip length and traffic conditions. The system is optimized under different operating rules, such as considering a single fleet, considering two fleets that operate independently in each period, or considering two fleets that run together at the peak (only one of them runs at the off-peak); this last system is shown to be the most efficient one, with the single-fleet system just slightly worse. Solutions are compared with those obtained when considering each period in isolation, and crossed-effects among periods are identified. In addition, we study *second-best* strategies: we optimize the peak period in isolation and use a sub-fleet for the off-peak, and we compare the results of this strategy with the opposite one: an approximate rule is to optimize the system according to the conditions of the period that presents the highest total number of passengers (across its whole extension).

AGRADECIMIENTOS

Mis estudios de Doctorado se realizaron en paralelo a momentos muy relevantes de mi vida personal y de mi militancia política. Que todo haya resultado de manera exitosa me lleva a agradecer a las siguientes personas:

A mi profesor guía, Sergio Jara-Díaz, quien constantemente se preocupó de que el Doctorado y su tesis avanzaran de buena manera, de proyectar este trabajo hacia un futuro que permita seguir desarrollando investigación de buen nivel, y que siempre acompañó nuestro trabajo con la mejor disposición y con conversaciones necesarias sobre política, literatura, música y fútbol.

A mi pareja, Carolina Campos, con quien empezamos nuestra relación en los comienzos del Doctorado, y hoy estamos formando nuestra propia familia, con nuestro hijo Rafael que viene en camino, con nuestro Merluza y nuestros planes de viajes próximos. No es posible saber cómo habrían sido estos años de tesis sin su apoyo y compañía, y no hay ninguna razón para querer saberlo, pues ha sido una experiencia de intenso aprendizaje, amor y cariño.

A mi madre, mi padre, mi hermano y el resto de mi familia, porque siempre han estado y estarán. Han sido parte fundamental de lo que soy, y han enseñado la relevancia de contar con relaciones tan sólidas para poder imaginar cualquier tipo de futuro.

Al resto del cuerpo académico del Doctorado, pues los cursos, seminarios y discusiones han sido fundamentales para fortalecer mis conocimientos disciplinares y para fomentar un pensamiento crítico que permita imaginar nuevas preguntas de investigaciones y nuevas líneas de trabajo. Antonio Gschwender, Alejandro Tirachini y Roberto Cominetti han sido de particular colaboración en este sentido. Al ISCI, por su organización de constantes espacios de reflexión académica y por su ayuda económica para realizar este Doctorado.

A mis compañeros y compañeras del Partido Comunes, porque sin la esperanza de un Chile diferente sería muchísimo más difícil cualquier tipo de actividad, y porque el trabajo conjunto ha moldeado mi manera de mirar el mundo.

Por último, al resto de mis amigos y amigas por hacer de estos años mucho más felices, entretenidos e interesantes.

TABLA DE CONTENIDO

Chapter 1. Introduction 1

Chapter 2. Brief review of some necessary previous models: the single-line model, the parametric city and the basic lines structures 5
2.1 The singe-line single-period model 5
2.2 Parametric description of the urban area 6
2.3 The strategic line structures 7
2.4 Main results 8

Chapter 3. The role of heuristics in designing lines structures 11
3.1 Description and application of the heuristics 11
3.1.1 Dubois *et al* (1979) – DBL 11
3.1.2 Ceder and Wilson (1986) – CW 13
3.1.3 Borndörfer *et al* (2007) – BOR 14
3.1.4 Cenek (2010) – CEN 15
3.2 Results and analysis 15
3.2.1 Main results 16
3.2.2 Global indicators 18
3.2.3 Analysis of operators' and users' costs 19
3.2.4 Role of the transfer penalty 20
3.3 Main conclusions 21

Chapter 4. The technical dimensions behind scale economies induced by transit lines structures design. 22
4.1 Introduction: scale economies in public transport 22
4.2. The impact of the discrete nature of lines structure choice on *DSE* 24
4.3. Introducing directness 26
4.3.1 The concept 26
4.3.2 An illustrative model 27
4.4 Analysis over the parametric city 29
4.5 Some results on subsidies and fares 33
4.6 Main conclusions 35

Chapter 5. Introducing lines density in the strategic design of transit networks 37
5.1 Lines density: the parallel lines model revisited 37
5.2 Lines density in the parametric city 41
5.2.1 Lines density as a new design variable 41
5.2.2 Results 43
5.3 Main conclusions 47

Chapter 6. Two periods optimization over a single line 49
6.1 Formulation of the model 49
6.2 Buses full at the peak 50
6.2.1 Comparing optimal and single period frequencies 50
6.2.2 Crossed effects between periods 52

6.2.3 Summary of this case and comparison with previous approaches 52
6.2.4 Numerical analysis 54
6.3 Buses full at the off-peak 57
6.4 Main conclusions 61

Chapter 7. Systems that allow for two fleets 63
7.1 Joint optimization allowing for two fleets 63
7.2. One or two fleets? Comparison of the models 67
7.3 Main conclusions 69

Chapter 8. Second-best strategies 70
8.1 Optimizing one line with fixed capacities 70
8.2 Optimization of one period and adaption of the other 71
8.3 Comparison between the two alternatives 72
8.4 Main conclusions 73

Chapter 8. Synthesis and conclusions 74

Bibliography 79

Appendix 85
Appendix A: Numeric values of the parameters 85
Appendix B: Some details on the application of each heuristic to the city model 85
Appendix C: Analysis of line structures over the isosceles-city 89

Chapter 1. Introduction.

Deciding the adequate public transport provision in space and time presents very particular economic characteristics. The chance of combining diverse routes across different periods has profound implications over design and optimization of the system, and over scale, density and scope economies. All this impacts in a relevant way life quality of the inhabitants of each city, due to the central role that the transit system plays not only in displacing the citizens, but also in prefigure the development of the city itself.

In public transport, each vehicle can carry passengers that travel from different origins to different destinations, which might affect significantly capital, operational and infrastructure costs if transit services are structured to combine efficiently the different trips. The ways these services are organized have effects over a resource that sometimes is ignored in production analysis: users' time. Different designs yield different waiting, access and traveling times. The possible ways to organize a transit system, including their routes, itineraries, stopping strategies or vehicles' sizes require deep economical analyses. Similarly, the fact that the same vehicle is used across different periods induces complex inter-temporal relationships from the point of view of the optimization of the system; crossed effects might emerge (e.g. diversity economies due to savings in capital costs) that interact with the spatial effects mentioned above.

The normative-economic analysis of public transport has been usually based on simple models that, while allowing relevant advances in the area, rely precisely over the omission of spatial and temporal aspects of the design; elements discussed in the previous paragraph cannot be included and require an expanded theoretical framework. This fact, together with the relevance of public transport systems for the cities and the technological changes that are modifying some of their characteristics, make the deepening of transit analysis a priority task. The optimization of transit services, the obtaining and studying of the associated cost functions, or the identification of scale economies are relevant and complex topics, something that is deepened when the analysis is done considering the different Origin-Destination-Period triplets.

The economical analysis of any production process requires understanding the cost function, which is the minimum necessary expenditure that allows generating a certain product, in this case a vector of flows across space and time. The value of the resources consumed (VRC) should consider the time spent by users and the resources spent by operators. To minimize VRC it is indispensable to understand and represent the technical elements that relate the design variables, in order to optimize the system as a function of them. Thereby, the first models of Mohring (1972) and Jansson (1980, 1984) consider exclusively the size of the vehicles and their frequency; by omitting the rest of the variables, they work over a simplified model, consisting in a single line with homogeneous load. Each author introduces other simplifications as well, depending on the specific model, such as the cost function associated to each bus, the chance of skipping some bus stops because vehicles are fully loaded, or the assumption of a constant cycle time, among others.

Although these models are extremely simplified, they reveal some economical conclusions that can be extrapolated to more complex schemes. The most known one is the so-called “Mohring effect”, that shows a notable source of economies of scale in public transport: when the number of passengers increases, optimal frequency increases in response, which induces a decrease in waiting times for all passengers. Similar analyses might be done regarding traveling times and operator costs.

There are several possible directions in which these models could be more realistic and complex, but introducing spatial and temporal complexity is of particular relevance for an economical analysis. A spatial perspective is crucial, since transit systems base their operations in the chance of combining different lines to serve flows between diverse origin-destination (OD) pairs. Incorporating routes design in the optimization problem turns this problem NP-Hard in many of its specifications (as shown by Quak, 2003, Schöbel and Scholl, 2006, or Borndörfer *et al*, 2007); but together with increasing the mathematical complexity, these new design variables could induce new optimization possibilities and new sources of scale economies.

The researchers have faced the impossibility of solving this problem exactly in different ways. There are diverse heuristics (Dubois *et al*, 1979, Ceder and Wilson, 1986, Pattnaik *et al*, 1998, Borndörfer *et al*, 2007, and Cenek *et al*, 2010, propose some examples) that seek for nice solutions; this approach, however, is hardly interpretable in its results. On the other hand, the literature has been solving these problems in networks of an increasing degree of complexity. Chang and Schonfeld (1991), for instance, generalized the one-line problem to several parallel lines, while Jara-Díaz and Gschwender (2003a) compared direct systems with feeder-trunk in a network with the shape of a cross; Tirachini *et al* (2010) studied a radial city, while Daganzo (2010) and Badia *et al* (2014) study a regular-grid and a radial-grid city, respectively. Fielbaum *et al* (2016) propose four basic structures (feeder-trunk, direct, exclusive and hub and spoke) over a parametric city that permits for representations of demand patterns associated to monocentric, polycentric or dispersed cities; Hörcher and Graham (2018) return to a single line, but with heterogeneous loads. Beyond the specific solutions, it is worth mentioning at this point that it is consistently found that the total number of passengers, as well as their internal distribution in the cities, plays key roles when optimizing and comparing different structures.

Although the existence of different periods does not impose interactions as complex as in the spatial analysis, it is very relevant from the operators’ perspective: one of the most relevant sources of scope economies in public transport is the chance of using the same vehicles for different periods. The basic way of facing this problem was proposed by Jansson (1984) in a single-line model, but he only solved it in an intuitive way. Public transport problems considering several periods have also been studied by Chang and Schonfeld (1991) and Medina *et al* (2013) in simple spatial contexts, even though they have not focused their analyses in the impact of the differences between periods. In a different research line, Glaister and Lewis (1978), De Borger *et al.* (1996), Proost and Van Dender (2008), Parry and Small (2009) and Basso and Silva (2014) study some different versions of the problem of optimal fares and subsidies for public and private transport to achieve an optimal modal split considering peak and off-peak periods. Fernández *et al.* (2005) consider various periods in their analysis of operators costs only with fixed bus size.

Walker (2012) proposes a new question related to second-best alternatives: is it better to optimize for the peak and adapt for the off-peak –as it seems to be usually done- or to do it the opposite way? This question reveals that from the optimization process itself, the problem of having different periods adds new topics for a proper transit design.

In this thesis we develop, solve and analyze some models that give us new insights about the most relevant economical aspects of public transport design. Regarding the spatial aspects of the design, we deepen the understanding about which type of lines structures (i.e. which spatial arrangement of transit lines) is more convenient depending on the characteristics of the city served by it, by studying some results of the heuristic approach and comparing them with the results already known over simple networks. We also analyze the effects of considering lines structures in the design when identifying some relevant economic features, such as scale economies or fares and subsidies; and we merge the lines structures analysis with the optimal design of the spatial density of the routes (inspired by Chang and Schonfeld, 1991).

The incorporation of differences across periods needs to start at a more basic point, as there are no prior clear results even for a single line, which is the case that we are going to focus on. We analyze the single-line model considering two periods; we propose different transit systems, seeking for analytical and numerical results, and we compare them to determine under which conditions each of these systems is better. We also tackle the question posed by Walker (2012), i.e., we analyze the second-best solutions explained above.

Eight chapters (after this introduction) compose the thesis. Chapter two synthetizes some previous models that are used often throughout the thesis, namely the single-line model developed by Jansson (1980, 1984) and Jara-Díaz and Gschwender (2003b, 2009); and a parametric urban model that has been proven useful to analyze transport systems, with four basic lines structures proposed over it, which have been developed by Fielbaum *et a*l (2016, 2017).

Chapters three to five are oriented to describe and analyze the spatial aspects of public transport design. Four heuristics are applied over the parametric city model described by Fielbaum *et al* (2017) in chapter three, and the emerging lines structures are compared with the basic ones studied by Fielbaum *et al* (2016). The contents of this chapter correspond to the article by Fielbaum *et al* (2018).

Chapter four is based in the results of Fielbaum *et al* (2019a), focusing on the effect of considering lines structures as a design variable over scale economies. More precisely, it shows that there is a source of scale economies derived from the chance of having more direct routes for the passengers when the total flow of passengers increases. The subsequent conclusions about fares and subsidies are also included.

Chapter five is about the spatial distance between parallel lines, and how does this design variable interact with the whole lines structures design. For this, we first extend Chang and Schonfeld's (1991) model by dropping some unnecessary simplifying assumptions, and we then introduce this additional variable (spatial density of lines) to the parametric city model to observe if there are changes in the comparison between the four basic structures. We study in both schemes the relationship between access time (that cannot be included in the models that do not have lines density as a variable) and the other characteristics of users' trips. These contents are presented in Fielbaum *et al* (2019b).

Chapters six to eight deal with the challenge of transit design considering peak and off-peak periods. In all of them we study the same single-line model, with a demand evenly distributed in space but with a larger flow at the peak. Other exogenous characteristics are also assumed to be different, like congestion and length of the trips.

Chapter six studies the "natural" extension from the single-line single-period model to two periods, which was also studied by Jansson (1984). We optimize the acquisition and operation of a single fleet of vehicles, such that they might (or might not) run full during these periods, and some buses might not be used during the off-peak. Although explicit expressions for the optimal frequencies and bus size cannot always be found, we provide several analytical and numerical results. These are developed in Jara-Díaz *et al* (2017, 2019).

In chapter seven we propose alternative ways to face this same scheme. First we deduce the equations that govern a system composed by two fleets: one that operates alone in the off-peak, and a second one that complements the former during the peak. The vehicles of these two fleets might be of a different size, such that a holding strategy is assumed to avoid having different cycle times at the peak (due to different times at bus stops). This system is compared with the one studied in chapter six and with a system consisting simply in operating each period independently. These results can also be found in Jara-Díaz *et al* (2019).

In chapter eight we analyze the second-best solutions explained above. We study the equations resulting from optimizing one period in isolation, and then using that type of vehicles (as an exogenous variable) when optimizing for the other period. We investigate under which conditions it is better to optimize for the peak and adapt for the off-peak instead of proceeding in the inverse way.

At the end of each chapter a list with the most relevant partial conclusions is provided. In chapter nine, we make a global synthesis of the thesis, explaining the most relevant conclusions and discussing some lines for future research.

Chapter 2. Brief review of some necessary previous models: the single-line model, the parametric city and the basic lines structures.

In this section we provide a brief review of some relevant previous models that are used often throughout the thesis. First, we describe the basic model developed by Jansson (1980, 1984) and Jara-Díaz and Gschwender (2003b, 2009) for the optimal design of frequencies and bus sizes when only one line is considered in a single period. Our models are based over their very useful and simple equations, introducing complexity both in time and space.

Second, we describe a not-so-simple spatial model proposed by Fielbaum *et al* (2016, 2017), in order to analyze the design of line structures. Finding an optimal design is an NP-Hard problem for many cost functions, even if frequencies are not considered; when they are taken into account, a second level of the problem emerges, as these frequencies need to fit users individual route choices. The simplified model explained in this chapter allows for representing different types of cities without increasing spatial complexity too much.

2.1 The singe-line single-period model

The stylized representation of the single-period design problem follows the developments by Mohring (1972), who introduced the "square root formula" in a simple model that optimized the frequency of a circular line (without a beginning or an ending) considering that operators' cost increases with frequency, whereas waiting time decreases with it; the resulting optimal frequency is proportional to the square root of the demand. Many refinements have been made over this basic model. Probably the most important one was the inclusion of the impact of passengers boarding and alighting on both bus cycle and passengers' in-vehicle times in such a way that a closed analytical solution is obtained (Jansson, 1980, 1984). The relation between optimal frequency and demand becomes linear for large patronage volumes (modified square root formula). Jara-Díaz and Gschwender (2003b, 2009) analyzed the impact of a financial constraint on the optimal frequency and vehicle size, expanding Jansson's model by making bus cost dependent on its size. Let us recall this general formulation for the one-period model over a single line, defining at the same time some notation. To do so, we start with the Value of the Resources Consumed, VRC:

$$VRC = B(c_0 + c_1K) + \frac{p_w}{2f}Y + p_v t_c \frac{l}{L} Y \qquad (2.1)$$

The first term corresponds to the operators' costs, while the second and third are those corresponding to the users (their time). Operators' costs is given by the fleet size B times the cost of acquiring and using each bus, which has been shown to have a linear dependency on bus capacity K (Jansson, 1980, 1984); c_0 and c_1 are the corresponding exogenous unit costs. Users are evenly distributed across the line, and their cost depends on total waiting and in-vehicle times, expressed as a function of the number of users that enters the system per hour Y, frequency of the line f, cycle time t_c, and the ratio between trip length l (which is assumed to be equal for all users) and the total route length L, which converts Y into the passengers' flow at every point of the cycle; p_w and p_v are waiting time and in-vehicle time values, respectively. Note that behind c_0 and c_1 there are two types of operators' expenses, operating and capital costs, representing usage and acquisition respectively. As this distinction will become relevant in the analysis of the two-periods case, it is convenient to recall that each of them can be expressed as linear in capacity functions as well (Jansson, 1980. 1984), i.e. capital cost is given by $c_{BC} + c_{KC}K$ and operating costs correspond to $c_{BO} \ + c_{KO}K$.

Equation (2.1) may be solved by expressing all the design terms as a function of the frequency f. Following Jansson (1980, 1984), vehicle cycle time is given by the addition of time in motion T and time at stops, i.e. $t_c = T + tY/f$, where t is boarding-alighting time per passenger. As frequency is total fleet size divided by cycle time, we obtain that $B = fT + tY$. Regarding vehicles capacity, costs minimization induces a value of Knot larger than the minimum necessary to carry all passengers, i.e. $K = \frac{lY}{Lf}$. Replacing B and K in (2.1) yields expression (2.2) where only f is a variable

$$VRC = c_0Tf + \frac{tY^2\frac{l}{L}(c_1+p_v)+p_wY/2}{f} + Tc_1\frac{lY}{L} + tYc_0 + \frac{p_vlYTP}{L} \qquad (2.2)$$

The optimal values for f and K are

$$f^* = \sqrt{\frac{Y\left(\frac{1}{2}p_w+p_vtY\frac{l}{L}\right)+\frac{c_1tY^2l}{L}}{c_0T}} \quad \text{and} \quad K^* = \frac{lY}{Lf^*} \qquad (2.3)$$

According to this result, both optimal frequency and bus size increase with patronage at a decreasing rate but, as shown by Jara-Diaz and Gschwender (2003b, 2009), frequency grows faster than capacity. For large patronage, capacity is asymptotic to a maximum value while frequency increases linearly.

2.2 Parametric description of the urban area.

Let us begin summarizing the parametric representation of an urban area proposed, justified and applied by Fielbaum *et al* (2017). It is a city model based on previous studies regarding topologic (Masucci *et al*, 2009, and Lin and Ban, 2013, are some examples) and economic (from the discussion about monocentricity between Alonso, 1964, and Hamilton and Röell, 1982, to some more recent models like the one by Louail *et al*, 2015) analysis of cities, flexible enough to represent many of the phenomena described in the literature about modern cities and, at the same time, simple enough to admit a precise analysis of public transport lines. Through its parameters, the city model can represent different degrees of monocentricity or polycentricity and its road structure is hierarchical, as observed in most of the cities. It is useful for our purpose because it has a recognizable structure that allows the design of (alternative) strategic line structures.

The city has a CBD and n zones, each one containing a subcenter (SC) and a periphery (P). There are arcs linking the CBD with each subcenter, each subcenter with the periphery of its zone, and neighbouring subcenters. The distance between a subcenter and the CBD (measured in the time needed by a bus) is T_0 and the distance between a subcenter and its periphery is gT_0. The city presents radial symmetry, so the distance between consecutive subcenters is known as well. The CBD is only an attractor and the peripheries are only generators of trips; subcenters generate and attract trips (modeling morning peak). The city and the demand structure are shown in Figure 2.1, where Y is the total patronage and a is the proportion of trips that start from the peripheries, out of which a proportion α goes to the CBD, β to the own subcenter and γ to the other subcenters, such that $\alpha+\beta+\gamma=1$. A proportion $b=1-a$ of trips starts at the subcenters, and they go to the CBD and to other subcenters in proportions $\tilde{\alpha}=\frac{\alpha}{1-\beta}$ and $\tilde{\gamma}=\frac{\gamma}{1-\beta}$ respectively. There are only four flow related independent parameters: α, β, a and Y.

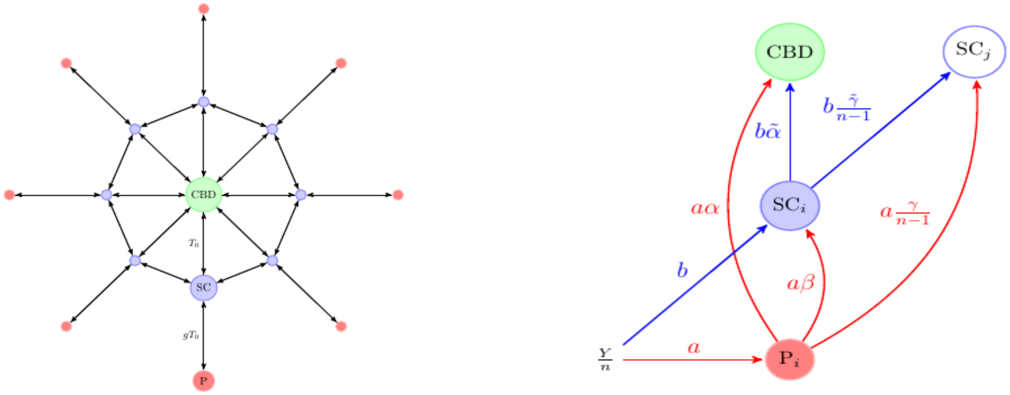


Figure 2.1. Parametric representation of a city (symmetric version) and its demand structure.

The parameters α, β and γ represent the degree of monocentrity, polycentrity or dispersion of the city, respectively: most trips go to the CBD for α →1, most trips go to the own subcenter for β →1, and most trips are distributed towards other subcenters for γ →1. Along with Y they will be varied during the analysis to explore how do they affect the results. Throughout the thesis, this urban model is referred as "the parametric city".

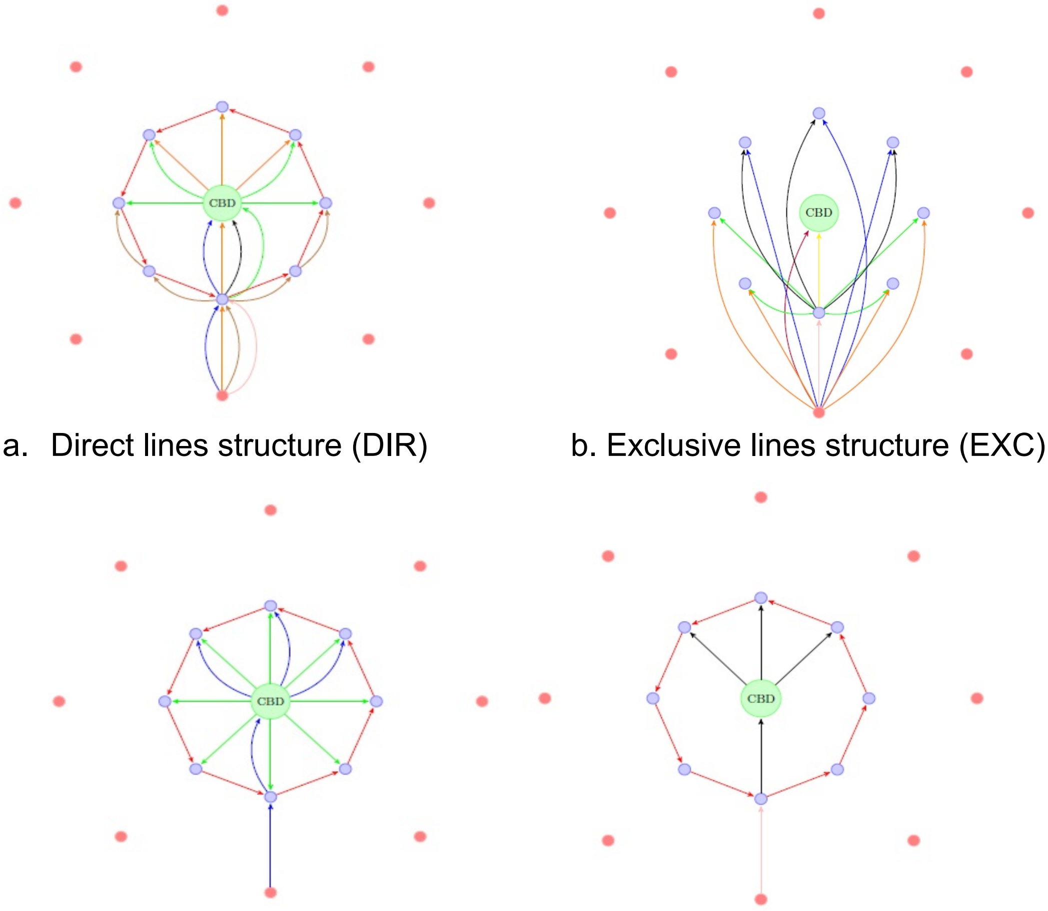
2.3 The strategic line structures

Fielbaum *et al* (2016) analyze four strategic structures over the parametric city model explained in section 2.2. They are represented in Figure 2.2; due to the symmetry of the city, only the lines emerging from the "south" zone, together with the circular lines, are drawn[1]:

- Direct lines structure (DIR): there are lines connecting each OD pair, including short lines for specific pairs; nobody needs to transfer.

[1] In chapters 4 and 5, structure DIR will be called "No transfers" (NT), and EXC will be called "No stops" (NS), to keep the notation used in the associated papers. These changes are useful to prevent being confused with the concept of "directness" introduced in chapter 4.

- Exclusive lines structure (EXC): there is one line for each OD pair without intermediate stops, such that nobody needs to transfer.
- Hub and spoke structure (HS): lines connect opposite zones through the CBD, where a transfer can be made to other zones. A circular line serves the subcenters ring, such that shorter trips are sometimes feasible.
- Feeder trunk (FT): feeder lines connect each periphery with its subcenter. A direct lines structure serves the subsystem composed by the subcenters and the CBD.



a. Direct lines structure (DIR)

b. Exclusive lines structure (EXC)

c. Hub and spoke structure (HS)

d. Feeder-trunk lines structure (FT)

Figure 2.2. Graphic representation of the strategic lines structures

It is worth explaining that the optimization process that will be described in the next section might yield null frequencies for some lines, depending on the parameters, i.e., the actual set of lines might be a subset of the lines shown in Figure 2.2.

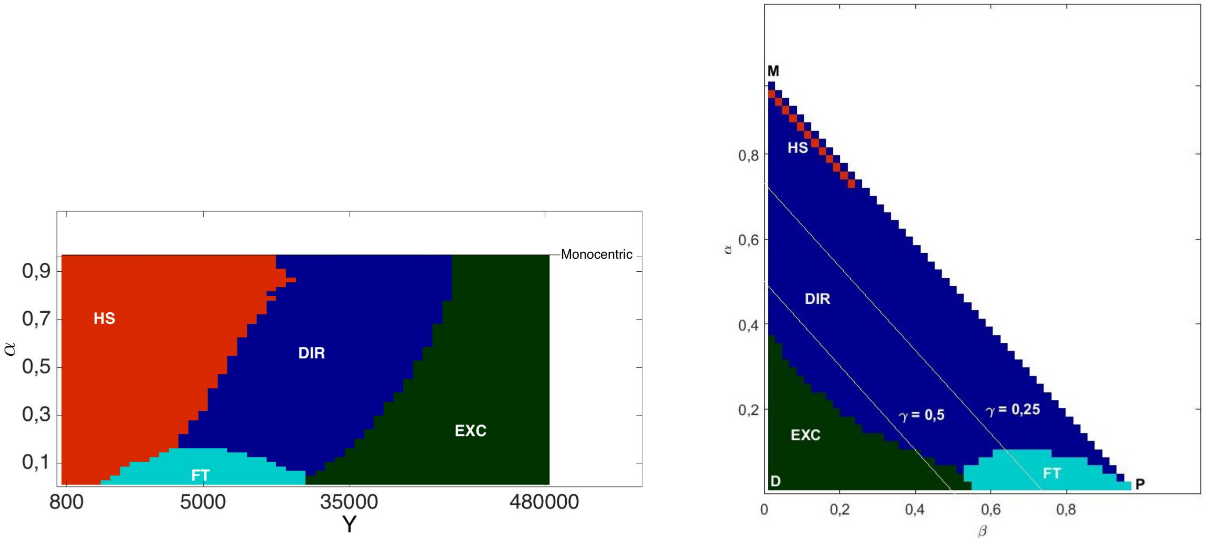
2.4 Main results

The best strategic lines structures were obtained in Fielbaum *et al* (2016) by finding the cost function (minimum value of the resources consumed) for each structure, optimizing frequencies and bus sizes. The value of the resources consumed is a direct extension of the one explained in section 2.1, just summing across all lines for operators, and adding a term that takes total transfers R into account for users:

$$VRC = C_O + C_U = \sum_l \; B_l\,(c_0 + c_1 K_L) + Y(p_v \overline{t_v} + p_v \overline{t_w}) + p_R R \qquad (2.4)$$

B_l is the total fleet of line l and K_l the capacity of its vehicles. Regarding users, $\overline{t_v}$ and $\overline{t_w}$ are the average in-vehicle and waiting times respectively. c_0, c_1, p_v, p_w and p_R are exogenous price related parameters.

Capacities, travel times and waiting times can be written as a function of the frequencies of each line on each structure, such that the frequencies vector becomes the main decision variable that can be optimized -as explained in detailed in Fielbaum *et al* (2016)- in order to find the minimum VRC. Transfers depend on the lines structure only[2]. Passengers assignment sometimes needs iterations, because more than one route is possible. The overall minimum yields the best line structure for different combinations of α,β and Y, which we present in Figure 2.3 (from Fielbaum *et al*, 2016), using colors for each structure.



a)(α, Y) space, $\beta = \gamma$ b) (α, β) space, $Y = 24000$

Figure 2.3. Optimal strategic structure for different combination of the parameters

In Figure 2.3a we analyze the effect of α (between 0 and 1) and Y (between 800 and 480,000, logarithmic scale), keeping $\beta = \gamma$, i.e peripheral passengers whose destinations are subcenters split in half between the own subcenter and the foreign subcenters. In figure 2.3b patronage is fixed at $Y = 24{,}000$ in order to analyze the effect of α and β. The rest of the parameters are shown in Appendix[3] A. In both figures each color represents one of the structures presented in Figure 2.2, such that the conditions (parameters values) under which each one dominates (i.e. that presents the lowest total

[2] The number of transfers could depend on frequencies because they might turn a route that includes transfers more attractive than a direct alternative. Nevertheless, this theoretical possibility does not happen when optimizing the system.

[3] Throughout the thesis, different parameters are going to be used over this same model. We vary the parameters to make the conclusions more robust, and because there are some analyses that require different emphases. All these parameters are based in different representations of Santiago, Chile.

cost) emerges clearly. To facilitate the interpretation, the extreme monocentric (*M*), polycentric (*P*) and dispersed (*D*) cases are shown when possible.

Figure 2.3a shows that Y has a very clear role: as patronage increases the structures with no transfers (DIR and EXC) dominate. HS and FT are convenient up to a moderate demand range, with FT advantageous only for very low proportion of trips to the CBD. For a mid-range patronage (Figure 2.3b), DIR dominates for most cases where α+ β is larger than γ (from mono to polycentric cities), except when α is small. When γ is larger, EXC becomes the best. As evident, every structure can become optimal under certain conditions.

For synthesis, DIR does not work well for dispersed cities (no ability to collect trips) but is optimal when most of the trips are radial; routes are not always the shortest ones, inducing the largest in-vehicle times. EXC becomes best only for large patronage; it requires a large fleet of relatively small buses inducing large waiting times but the smallest in-vehicle times (no intermediate stops). HS collects trips and induces large frequencies, reducing in-vehicle times; it dominates for low levels of patronage using a small fleet of large vehicles. Finally, the virtues of FT (little idle capacity) show up only if the city is dispersed.

Chapter 3. The role of heuristics in designing lines structures.

In this chapter we are going to apply some heuristics over the parametric city, to analyze the line structures that emerge from them and to compare their results with those from the four basic ones. Out of the many heuristics available in the literature we selected four for this analysis: Dubois *et al* (1979), Ceder and Wilson (1986), Borndörfer *et al* (2007) and Cenek (2010). The selection is based on diversity - date of publication and methodological approach - and feasibility, i.e. applicable to our scheme[4]. Parameters are shown in Appendix A.

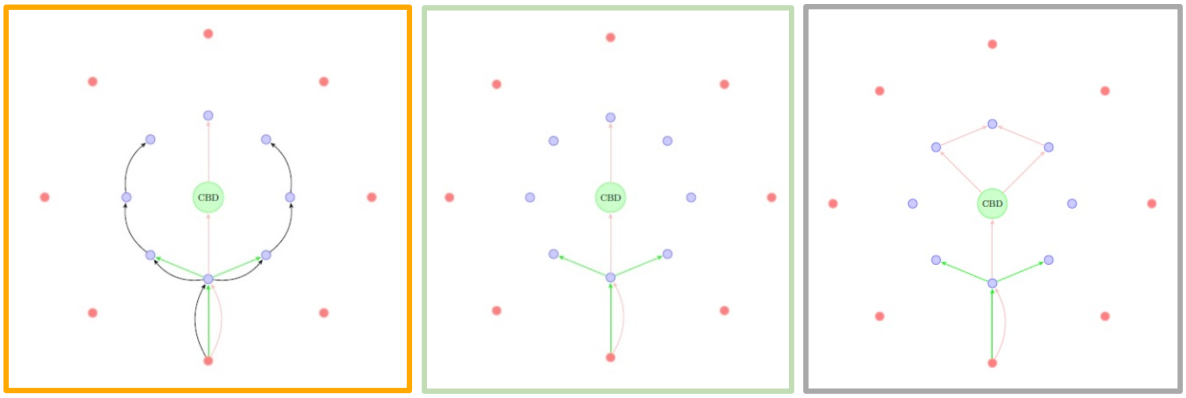
For each heuristic, we solved for one zone and replicated that solution for the others, following the same procedure used to design and analyze the basic strategic structures, which takes advantage of the symmetry of the city; this obviates some unilateral choices when building the routes. In the following sub-sections we describe and apply each of the heuristics, providing five elements: the main idea, a brief description of the procedure, comments and modifications (if any), results and analysis. As frequencies are going to be calculated afterwards using the same procedure for all - i.e. minimizing VRC as in Fielbaum *et al* (2016) - we focus on those parts of the heuristics conceived to construct and select the routes (for instance, in Dubois *et al*, 1979, we omit the selection of an optimal subset of streets and the search for optimal frequencies). Specific details of the procedures to apply each heuristic to our city model are given in Appendix B.

3.1 Description and application of the heuristics

3.1.1 Dubois *et al* (1979) - DBL.

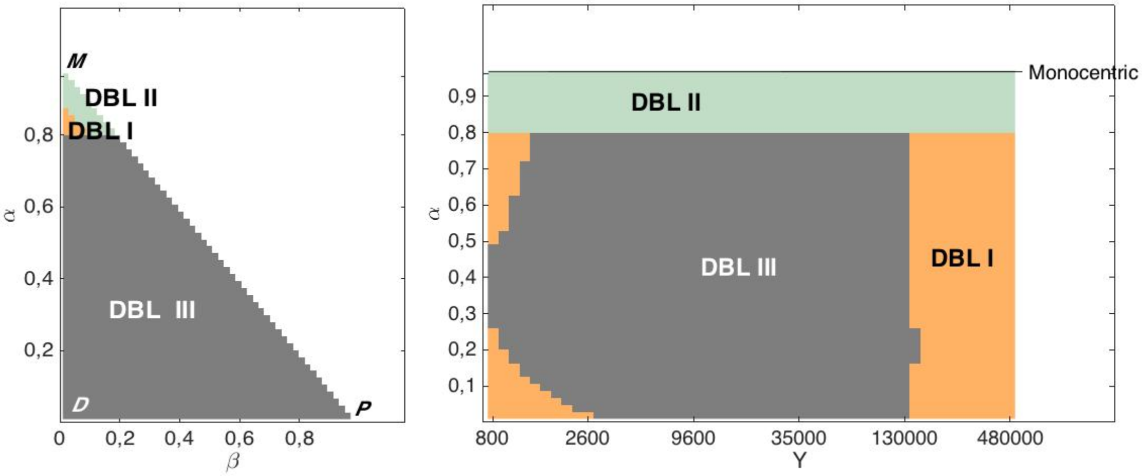
- Main idea: "Lines must be rather straight and at the same time a sufficient number of passengers must be picked up" (p. 801).
- Brief description of the heuristic: it begins by identifying the set of shortest routes that connect the most distant O-D pairs. Then it admits deviations for all of these routes, but keeping their length variation within a range $1+\sigma$, where σ is imposed *a priori*. With these routes as candidates, it builds the structure adding them in some order until the graph is fully connected. Finally, if the total number of transfers is too large or if the total travel time increases too much relative to the shortest paths, the routes that diminish either in a more effective way are added.
- Comments: σ represents the trade-off between having more direct routes (hence diminishing travel times for users and cycle times for operators) and collecting more passengers (allowing lower waiting times and smaller fleets). This parameter is supposed to be fixed, but we solved for each possibility up to 0.5. Also, we built the structure adding the routes from the shortest to the longest.
- Results: As shown in Figure 3.1, two structures emerge for $\sigma < 0.328$ ($\approx \frac{1}{3}$) and only one for $\sigma > \frac{1}{3}$. Figure 3.2 shows the structure that presents the minimum VRC for each combination of α, β, γ and Y.

[4] For example, van Nes *et al* (1988) solve the problem just minimizing user costs - represented by the number of transfers - within a given maximum budget for operators; there is no clear way to adapt that idea to this scheme.



DBL-I.$\sigma < \frac{1}{3}$, γ > 10%. DBL-II. $\sigma < \frac{1}{3}$, γ < 10%. DBL-III. $\sigma > \frac{1}{3}$, all γ.

Figure 3.1. Lines structures obtained with the DBL heuristic.



a) (α, β) space, $Y = 24000$ b) (α, Y) space, $\beta = \gamma$

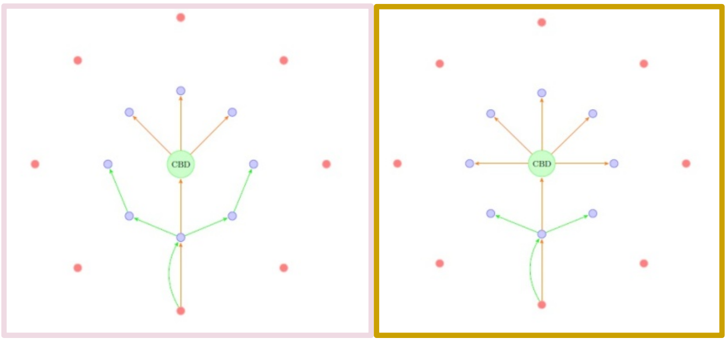
Figure 3.2. Best DBL structure

- Analysis: all resulting structures in Figure 3.1 are mostly direct[5], and they emerge nearly independently from the demand parameters; in fact, they play a role because we turned σ into a varying parameter. If $\sigma > \frac{1}{3}$ we always obtain DBL-III. If $\sigma < \frac{1}{3}$ only very small values of γ make one line vanish (the black one in DBL-I). Regarding the search for the optimal σ, Figure 3.2 shows that structure III - with a larger allowed deviation from the shortest routes - dominates in most cases, because it is efficient collecting passengers and reducing waiting times; Figure 3.2b shows that this happens for $0.1 < \gamma < 0.5$ and Figure 3.2a suggests that this extends to $\gamma = 1$. This virtue weakens when patronage is large because large frequencies diminish the relevance of waiting times, favoring structures I and II with small or no deviations from the shortest-path. In highly monocentric cities structure II is optimal because it is based on trips to the CBD.

[5] In all cases transfers might occur, but by a small fraction of the total patronage. In our application with eight zones, structure II imposes transfers towards four foreign subcenters out of nine destinations, but involving less than 5% of the total flow. Although structure III imposes transfers only towards two foreign subcenters, the corresponding flow could be much larger in a dispersed city (large γ); for example, if γ was 70% the flow requiring transfers could reach up to 20%.

3.1.2 Ceder and Wilson (1986) – CW

- Main idea: instead of searching for complete bus routes, they are built node to node, controlling for excessive length of the passengers' routes.
- Brief description of the heuristic: a set of nodes is defined as "terminals". Beginning with an arbitrary terminal, routes are built starting from that node passing through unconnected new nodes but not exceeding by more than σ the length of the shortest route from the terminal to each node (similar to Dubois *et al*, 1979), until no new routes are possible. The procedure is then applied from a new terminal, obtaining a tree[6] from each terminal.
- Comments: general objective is similar to the previous heuristic, but in this case routes are built instead of explored, which may be relevant in terms of computational time for big networks. As σ plays the same role as in DBL, we treated it similarly exploring all values up to 0.5.
- Results: we obtain two possible structures shown in Figure 3.3, depending on whether $\sigma < 0.2517$ $(\approx \frac{1}{4})$; they are compared in Figure 3.4 only in the (α, Y) space for $\beta = \gamma$, as in the (α, β) space with $Y = 24{,}000$, structure CW-I dominates everywhere**.**



CW-I. $\sigma < \frac{1}{4}$ CW-II. $\sigma > \frac{1}{4}$

Figure 3.3 Lines structures obtained with the CW heuristic.

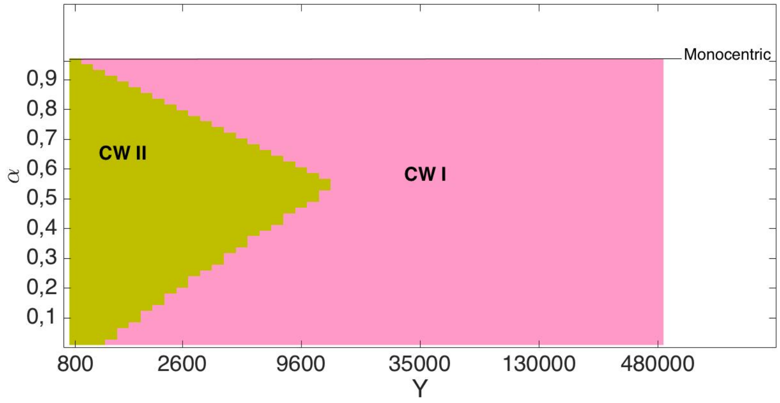


Figure 3.4 Best CW structure.

- Analysis: again emerging structures are based on direct trips and, again, they are only sensitive to the demand pattern through σ. Note that for small σ routes are shorter and

[6] A graph is called a *tree* if it is connected and it has no cycles.

passengers spend less time on board, while a larger σ permits to pick up more passengers with the same line (generating longer routes). Collecting more passengers is useful when patronage is small because passengers perceive higher frequencies with a smaller fleet, but when Y increases this advantage vanishes. This makes structure II more competitive when the destinations are distributed more evenly across the city.

3.1.3 Borndörfer *et al* (2007) – BOR

- Main idea: to solve a Linear Programming Problem (LPP) that calculates frequencies and assigns passengers, by means of minimizing travel times and operator costs.
- Brief description of the heuristic: it starts with all possible lines, each with a given bus size; then it defines an optimization problem that minimizes total costs, i.e. the costs related to travel times for the passengers plus the operators fixed cost per bus. Then it minimizes over the frequencies of each line and the passenger assignment, with restrictions that ensure physical feasibility. The heuristic component comes from the possible huge number of variables; so approximated methods are needed to solve the optimization problem.
- Comments: because of the simplicity of the graph in our case, considering only lines that are shortest-path between their origins and destinations generates a manageable number of variables such that the solution may be calculated exactly. The emerging structure is composed by those lines whose frequency is not null. As passengers' assignment is optimized to obtain minimum total cost, they are assumed to use the system in a way that may differ from what would be their actual individual choices. Waiting times and number of transfers play no role.
- Result: the set of lines with positive frequency yields only one structure for all combinations of parameters and bus sizes, shown in Figure 3.5; no comparison is needed.

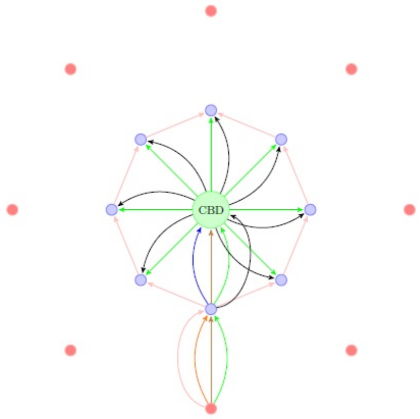
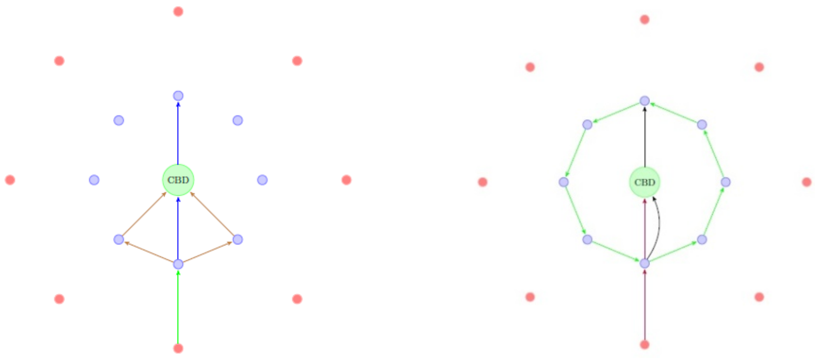


Figure 3.5. Lines structure obtained with the BOR heuristic.

- Analysis: although this heuristic also yields a direct lines structure, it is quite different from the previous ones. This structure is absolutely insensitive to the demand pattern; it represents the set of lines that – as in all other cases –receives the same treatment described in section 2.4 in order to obtain frequencies that minimize VRC.

3.1.4 Cenek (2010) – CEN

- Main idea: to build line routes that connect centers and that provide short alternatives to the largest number of passengers.
- Brief description of the heuristic: some nodes are pre-defined as “centers”. Every arc has a starting weight, calculated as the number of passengers that would cross it if everyone chose shortest routes. Then, each route is built starting at a center, continuing through the arcs with the highest weights, and ending when either another center or a border-like node is reached. After a line is created, the weight of its arcs is reduced by the minimum of its components.
- Comments: although it is an interesting heuristic for our parametrically described city, as it is based on the centers structure, it would create only one-arc lines if applied literally. Therefore, we relaxed the criteria of ending the line when another center is reached.
- Results: the results depend on whether a is larger than $a(\alpha+\frac{3}{7}\gamma)+b(\tilde{\alpha}+\frac{3}{7}\tilde{\gamma})$, i.e. the construction of routes does depend on the demand pattern. The two possible structures are represented in Figure 3.6.



CEN-I. $a > a(\alpha+\frac{3}{7}\gamma)+b(\tilde{\alpha}+\frac{3}{7}\tilde{\gamma})$ CEN-II. $a < a(\alpha+\frac{3}{7}\gamma)+b(\tilde{\alpha}+\frac{3}{7}\tilde{\gamma})$

Figure 3.6. Lines structures obtained with the CEN heuristic.

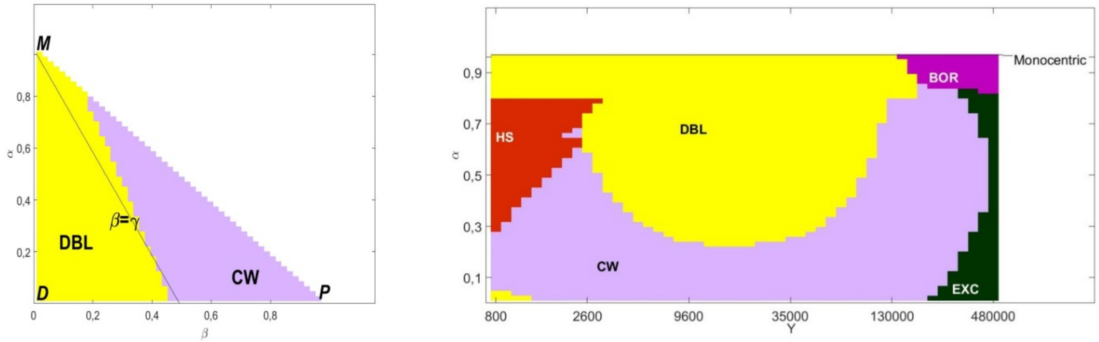
Analysis: The first structure makes all peripheral passengers that go to the CBD choose between a transfer at the own subcenter and a deviation to a neighboring subcenter, which seems disadvantageous for the users without diminishing operators’ costs. The second case is very similar to the original HS structure presented in Section 2. Grossly speaking, structure I emerges when $a \to 1$ (i.e most trips start at the peripheries) and structure II emerges when $\alpha \to 1$ (i.e. the city is monocentric). CEN-I and CEN-II do not compete, as only one of them emerges for a given combination of the demand parameters; no comparison is needed.

3.2 Results and analysis

In this section we present the overall results considering all the structures presented in the two previous sections. The idea is to show which structures are dominant (i.e. those that exhibit the minimum total cost) for the different values of the urban parameters. This requires the calculation of the minimum VRC for each and every one of the structures. As VRC is a function of frequencies and vehicle sizes of all lines belonging to a given structure, the optimal values for these variables were found using the approach described at the beginning of section 2.3, where VRC is written as an explicit function of the frequency vector and optimized (following Fielbaum *et al*, 2016).

3.2.1 Main results

Results are described in the two spaces previously used: the (α,β) space with $Y = 24{,}000$ (Figure 3.7a), and the (α, Y) space with $\beta = \gamma$ (Figure 3.7b). In general, the strategic structures studied in Fielbaum *et al* (2016) happen to be inferior to those generated by applying DBL, CW or BOR (shown in Figures 3.1, 3.3 and 3.5 respectively) in most regions of the spaces that described the city, with some exceptions for very low and very high patronage. As the structures obtained with heuristics are generally (improved) extended versions of the DIR structure, our results show that heuristics reinforce the strength of DIR by improving its performance by means of non-trivial modifications, such that they dominate nearly everywhere; out of these, a few transfers might occur only in the DBL structure.



a) (α,β) space, $Y = 24000$ b) (α, Y) space, $\beta = \gamma$

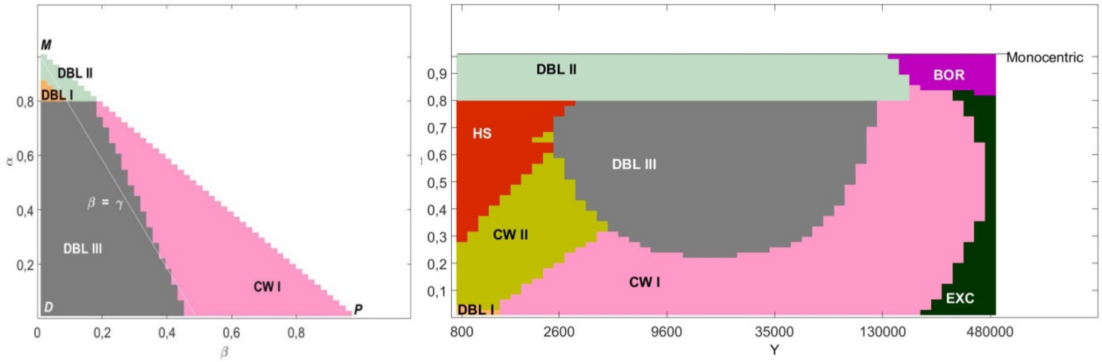
Figure 3.7 Dominant structures

Let us examine these results further. The original strategic structures dominate only for very small or very large patronage. EXC dominates when patronage is very high, as its main problem (large waiting times) diminishes importance when frequencies have to be high; this is consistent with previous results (Fielbaum *et al*, 2016; Gschwender *et al*, 2016) and its analysis is deepened in chapter 4. An interesting remark here is that the heuristics are not able to produce exclusive services, as they do not allow a route to skip stops. HS dominates for small cities with an important CBD but not fully monocentric; its structure - that exploits the relevance of the CBD - is useful in these cases. FT disappears; the zones where it dominated in the base case are now dominated by CW-I, which is a direct-type structure but that has a line very similar to the feeder lines in FT.

The heuristics generate DIR-like structures by admitting deviations from the shortest possible lines in order to collect passengers along the routes, constraining the increase in travel times and sometimes the number of transfers, which is exactly the declared objective of Dubois *et al* (1979), as quoted above. For example, DIR has one line for each foreign zone reached through the CBD (what we called the H set in Fielbaum *et al*, 2016), while CW covers all zones in H with only two lines from each periphery; by reducing the number of lines a good combination of a smaller fleet size with high frequencies is achieved. This same idea lies behind DBL-III (quite similar in shape to CW-I). The resulting structures based on DBL and CW improve on the strategic structures generally by less than 20% of total cost; for highly monocentric cities and intermediate demand levels, a 50% improvement can be achieved. A general analysis of the relative performance of all structures is presented in section 3.2.2.

Figure 3.7a shows that as polycentrism grows (larger β) the CW heuristic works better; when $\beta < \gamma$, DBL structure dominates, in a zone where previously DIR or EXC did; and when $\beta > \gamma$ CW is generally better, replacing DIR and FT. From Figure 3.2b, if $Y \in [5{,}000; 100{,}000]$ then DBL is always dominant if α >0.3, improving over DIR; moreover, it also dominates for most of the demand range if α>0.8, beating mainly HS and DIR. For a wide Y range and $\beta = \gamma$, increasing monocentrism makes DBL more effective but CW dominates for $\alpha < 0.25$. BOR is optimal only in the extreme scenario of a very large monocentric city (when $\beta = \gamma$) and CEN never appears.

Conceptually these results can be summarized by saying that the more dispersed the city the better is the DBL structure (usually DBL-III), while polycentrism is better served with the CW-I structure, as shown in Figure 3.8 where we detail which type of σ related structure (shown in Figures 3.1 and 3.3) is the winner when either DBL or CW are dominant. The intuition behind this is that the dominant structures are nearly identical but, as explained above, DBL has a line that specifically serves various external subcenters instead of many lines, which is quite good for collecting passengers. This is convenient when there is a high degree of dispersion; otherwise large fleets would be required or large waiting times would be obtained. Note that this is achieved because, in spite of the similarities, the DBL heuristic extends a path as much as possible while CW searches all possible forward movements from a node.



a) (α, β) space, $Y = 24000$ b) (α, Y) space, $\beta = \gamma$

Figure 3.8. Dominant lines structures in detail.

The flexibility of DBL and CW to obtain different line structures seems to be a relevant advantage to adapt well to different kinds of cities, but this arises because we treated the exogenously fixed tolerance (σ) parametrically, implicitly searching for a good σ. Figure 3.8 shows that up to $\alpha = 0.8$ when DBL dominates it is almost always with structure III. In the case of CW, structure I is the dominant one but structure II is better for low patronage.

Let us recall that, as shown in Chapter 2, the analysis involving only the four basic strategic structures indicated that each one dominates for different urban schemes. Although the analysis including the heuristics shows that the direct-type structures increase their range of dominance to all non-extreme cases, it is relevant to point out that both CW and DBL generate DIR-type because they are conceived as such. One might wonder whether heuristics based on EXC, HS or FT-type structures could dominate in the zones where the corresponding basic structures originally did.

In summary, it is interesting to realize that, depending on the demand pattern, the solution may vary widely. Actually, five structures out of eight are dominant under certain conditions. This means that to find a good structure for a specific city, it is necessary to start observing the global conditions to determine first the type of strategic structure to be implemented. In our example, two original strategic designs - HS and EXC - dominate for very low or very high patronage respectively, and the selected heuristics are not able to generate superior schemes because they are conceived with the idea of improved direct lines. In these cases heuristics built upon the HS or EXC concepts could do a better job. For a specific city it seems convenient to use a heuristic search after a careful analysis of the advantages and disadvantages of what we have called strategic designs, such that the detailed design could be constructed with the help of a heuristic adequately conceived for that structure.

3.2.2 Global indicators

Our analysis so far shows which structure dominates for each combination of the parameters. But some unresolved questions remain. On the one hand, it would be useful to have some global indicators (independent of the parameters that describe the city) for the comparison among structures, and in particular to decide within a set of heuristics which one is the best. On the other hand, the analysis so far gives no clue about how good (or bad) are the structures that do not dominate in a specific point that represents a city. For short, it is relevant to ask how far from the dominant structures are the non-dominant ones.

To explore this, the natural idea is to compare the value of the resources consumed VRC, the function that let us decide which structure is dominant. So for each value of α, β and Y, we calculated the percentage difference between the VRC of each structure and the VRC of the dominant structure for the same parameters, i.e., how bad did each non-dominant structure did. In Table 3.1 we synthetized these calculations, showing for each structure the maximum and the mean difference along the different values of the parameters. First row is in the (α, Y) space, second row is in the (α, β) space and the last row shows the global figures.

Max/mean	**DIR**	**EXC**	**HS**	**FT**	**DBL**	**CW**	**BOR**	**CEN**
$(\boldsymbol{\alpha}, \boldsymbol{Y})$	20.3/ 10.1	59.6/ 21.2	45.7/ 14.8	61.9/ 27.8	40.7/ 4	21.2/ 2.7	32.6/ 14.8	141.3/ 58.4
$(\boldsymbol{\alpha}, \boldsymbol{\beta})$	64.7/ 15.6	52.1/ 20.2	88.4/ 25.2	71.8/ 28.5	46.2/ 8.7	79.1/ 7.7	89.7/ 22	246.6/ 90.2
Global	64.7/ 12.9	59.6/ 20.7	88.4/ 20	71.8/ 28.2	46.2/ 6.4	79.1/ 5.2	89.7/ 18.4	246.6/ 74.3

Table 3.1: Relative difference between the value of the resources consumed by each structure and the lowest value (maximum/average)

Results in Table 3.1 show that, globally, DBL exhibits the lowest difference with the minimum VRC, i.e. it produces the most *reliable* structures, those that are never too far from the minimum. However, structures from CW are the *best in average* (about 5%

from the dominant) but sometimes very far; DBL comes a close second under this measure, and DIR comes third. On the other hand, CEN structures are the worst for each of the indicators. This again points to the conclusion that a single heuristic for each kind of city is not the best way to face this problem, as shown in Table 3.1 by the first numbers in the last row (maximum global deviation from the optimum). These figures indicate that all structures may be spending at least 46.2% additional resources with respect to the dominant one for some demand pattern.

Note that, with the only exception of EXC (whose relation with Y has already been explained) the maximum relative difference is always achieved when covering the (α, β) space. This indicates that the total number of passengers may be less relevant than how these passengers are spread throughout the city.

3.2.3 Analysis of operators' and users' costs

Which are the elements that make each line structure win under different urban structures represented by the parameters combinations? To explore this, let us look at the two main components behind VCR - i.e. users' and operators' costs – for each of the different dominant structures. Note that we are not looking for the optimal structures under the specific interest of users or operators, i.e. we do not optimize operators' or users' costs by their own. We simply want to capture how each overall dominant structure is perceived by each type of actor by using the optimal frequencies and capacities obtained in order to evaluate separately C_O and C_U.

In Figure 3.9 we show the variation of C_O (solid lines) and C_U (dotted lines) for each structure as total patronage increases for $\alpha = 0.5$, keeping $\beta = \gamma$. CEN was suppressed as it systematically showed the largest costs for both operators and users in the whole range analyzed. Only in this figure Y is not represented in a logarithmic scale in order to emphasize the resulting linear shape. $\alpha = 0.5$ was chosen because - as shown in Figure 3.8 - five structures become dominant as Y increases.

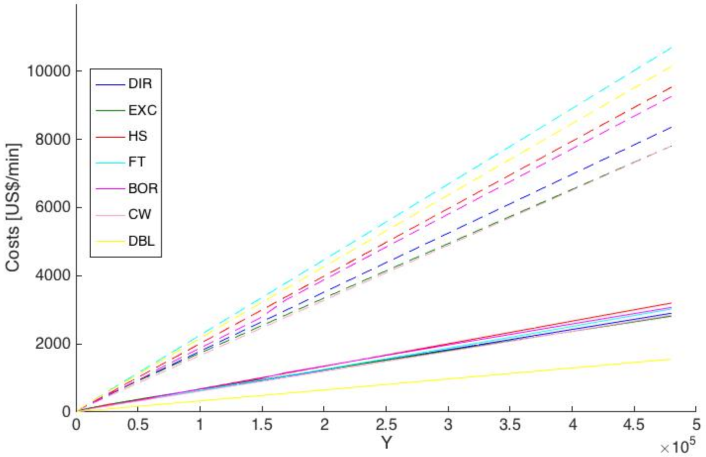


Figure 3.9. Operators (solid) and Users (dotted) Costs functions; $\alpha = 0.5$, $\beta = \gamma$.

Figure 3.9 shows that - given the exogenous parameters used in our study - users costs weight more (0 to 10,000) than operators' (0 to 3,000) and both increase with patronage for all structures. DBL (in yellow) presents, by far, the lowest operator costs, but is relatively costly in terms of users' resources. Recall that this structure is DIR-type but with a stronger tendency to collect passengers. This characteristic is good for operators, but users need to tolerate longer trips. So when DBL wins, it does because of operators' costs.

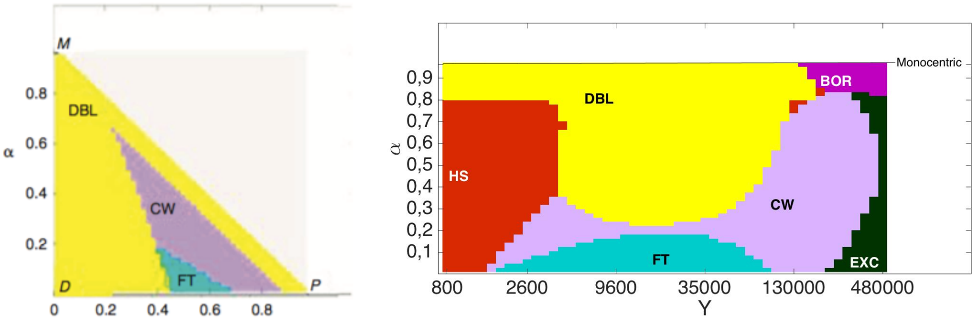
Operators' costs for all structures but DBL are quite similar, which makes users' costs particularly relevant when excluding DBL. Users costs present a larger variance, at least for large values of the patronage. FT and DBL present the worst user costs, while EXC and CW curves are quite close to each other, exhibiting the lowest users' costs.

Finally, in an additional analysis, we looked at the behavior of both operators and users costs for $\beta \in (0{,}0.5)$, $\alpha = 0.5$ and $Y = 24{,}000$ for all structures, as this permits a further look at the role of the subcenters. As expected, the main result is that increasing polycentricity (β) while diminishing dispersion (γ) makes all costs decrease steadily because trips are shorter.

3.2.4 Role of the transfer penalty

Now we examine the role played by the transfer penalty p_R, which has been shown in previous analyses to be relevant in determining the optimal line structure (Fielbaum *et al.*, 2016; Gschwender *et al.*, 2016). It is worth looking at the results that would be obtained if we assume an extreme case in which the only cost associated to a transfer is the additional waiting time, i.e. $p_R = 0$.

The results are shown in Figure 3.10, where the most interesting novelty is the re-emergence of FT for low levels of monocentrism, particularly for intermediate values of β and a wide intermediate range of patronage. The other structure that relies on transfers, HS, significantly increases its dominance area. Interestingly, DBL now dominates for small values of γ (i.e. those close to the hypotenuse in Figure 3.10a). Note that the best DBL structure in the area $\gamma < 0.1$ is DBL II, different from Figure 3.2a (with p_R equal to 24 in-vehicle minutes), which now dominates over CW in that area. This happens because DBL II presents mandatory transfers that are now less penalized. This confirms that p_R is a key parameter indeed; given the large variability reported in the literature, further research on this is badly required (see Currie, 2005, Raveau *et al.*, 2014 or Garcia-Martinez *et al*, 2018, for some results).



a) (α, β) space, $Y = 24{,}000$ b) (α, Y) space, $\beta = \gamma$

Figure 3.10 Dominant structures for $p_R = 0$.

3.3 Main conclusions

- Heuristics create mostly direct-type lines structures, which are seldom sensitive to the OD pattern.
- Lines structures that emerge after applying the heuristics are competitive. In particular, for most combinations of the parameters, DBL or CW are dominant.
- Making routes that do not follow shortest paths in order to collect more passengers improve the lines structures. The maximum deviation from shortest paths is a parameter that should be optimized.
- If the number of passengers is very small, HS is dominant; if it is very high, EXC is dominant.
- CW presents the lowest average cost, and DBL presents the lowest maximum difference with respect to the cost of the dominant lines structure across all the possible value of the parameters.
- All structures can be up to 45% more costly than the dominant one. From this point of view, spatial distribution of the trips is more relevant than the total number of trips.
- The numeric value of the pure transfer penalty plays a key role when comparing lines structures, including those created by the heuristics.
- It seems better to start by defining a global strategy (such as HS, FT, DIR or EXC) and then applying a specific heuristic, rather than trying to create a global heuristic that is useful for every city.

Chapter 4. The technical dimensions behind scale economies induced by transit lines structures design.

4.1 Introduction: scale economies in public transport.

Cost functions and economies of scale are economic concepts that are quite relevant for the normative analysis within production theory, where the technical process is represented by a production or transformation function that summarizes the conversion of inputs into outputs. Although production processes usually involve many inputs and outputs, the engineering technology represented by the production function is usually formulated using aggregates, something that indeed applies to the study of transport activities, where product was described using single scalar measures as ton or passenger-miles until mid-eighties, and by means of a vector of a very small dimension thereafter, including flows related variables, service quality variables and network description variables. The compact description of output prompted two definitions in the literature around the analysis of scale economies, both referring to proportional expansions of output: returns to density (called RTD) and returns to scale with variable network size (called RTS). The former considered a proportional expansion of outputs keeping network size fixed, while the latter considered a simultaneous expansion of both flows and the network by the same proportion (Caves *et al.*, 1984; Keaton, 1990). However, using aggregate output descriptions blurs the technical relations with inputs and has some unpleasant consequences in the analysis of economies of scale in transport activities.

Behind any compact description of transport output lays the true output of any transport firm: a vector of origin-destination (OD) flows of different things during different periods (Jara-Díaz, 1982a). In very simple transport systems the analytical derivation of the technical relations between inputs and flows - the production function – can be done, such that the corresponding cost functions can be obtained analytically as well[7]. This approach proved very useful to show that the use of aggregates introduced ambiguity in the economic analysis in transport because, for example, the same amount of passenger-miles could require very different types and amounts of inputs depending on how these passenger-miles are distributed in space. Most importantly, scale economies should be studied holding the origin-destination system constant, as introducing new OD pairs means introducing new products, which would require the analysis of economies of scope; this means that "economies of scale with variable network size" is actually an ill-defined concept, as shown by Basso and Jara-Díaz (2006a) while "economies of density" is better suited to the definition of economies of scale.[8]

A corollary from this story is that more attention has to be paid to the transport production process itself in order to fully understand scale economies. Besides the preceding difficulties, economic efficiency analysis in transport has to include a usually forgotten input, which is users' time, an input whose level is affected by the specific

[7] See for example the analysis of the backhaul transport system involving two flows only (Jara-Díaz, 1982b) or the three-nodes system studied by Jara-Díaz and Basso (2003).

[8] Sometimes RTD has been defined adding the condition that lines structure is unchanged after an increase in flows (Basso and Jara-Díaz, 2006b).

combination of vehicles, terminals and routes, chosen at the design level. This is indeed the case in public transport analysis.[9]

The provision of public transport services exhibits various technical characteristics that have been shown to affect its degree of scale economies (DSE), which is the quotient between average and marginal costs. First of all, the so-called Mohring effect described in chapter 2, where an increase in patronage makes optimal frequency larger and waiting times lower. In addition to this waiting time effect, as demand increases the system can also be adapted by incorporating more lines in space, thus reducing another component of users' cost, namely the walking time (this is further studied in chapter 5). Lines density has been modeled for a bus feeder system by Hurdle (1973), in a rectangular area by Kocur and Hendrickson (1982) for a single period, by Chang and Schonfeld (1991) for multiple periods, and by Small (2004), who analyzed the impact of road pricing on public transport. All of them obtain a cube root formula for both the optimal frequency of each line and for the optimal number of routes.

A third variable that can be adapted according to the demand level is the size of the vehicle, which also increases with patronage. As operators' cost per passenger diminish with vehicle size (due to fixed costs per vehicle), this is also a source of scale economies. However, when vehicles size increases the time spent at each stop also increases because more passengers board to, and alight from, each single vehicle, thereby increasing cycle time - which affects operators' cost as a larger fleet is needed - and users' in-vehicle time. Both effects reduce the degree of scale economies. Including these effects in his model of an isolated public transport line, Jansson (1980) obtained the modified square root formula explained in chapter 2 for optimal frequency, such that frequency is proportional to the square root of the demand if waiting times dominate (low demand and frequencies), or directly proportional to the demand if in-vehicle times dominate (larger demand and frequencies). In all models the adjustment of frequency and vehicle size generates scale economies that, nevertheless, diminish as flow increases.

In this chapter we introduce in the analysis of scale economies in public transport an important element of design that responds in a discrete way to increases in flow: lines structure, i.e. the way in which vehicles serve a number of routes in order to move a given set of flows (product). As flows grow these arrangements evolve in a way that should be studied specifically; understanding the evolution of design including lines structure and analyzing its impact on total costs and scale economies is the main objective of the chapter[10].

[9] There are studies where passengers' costs is not included. Farsi *et al*. (2007) and Viton (1992), for instance, study the operators' production function with an emphasis on multi-modal industrial organization; they consider aggregate outputs, and recognize the presence of scale economies. On the other hand, Fernández *et al*. (2005) find some sources of diseconomies of scale when studying operators' costs in bus corridors.

[10] Scale economies in public transport have also been reported in other dimensions. Tirachini *et al*. (2010a), for example, show that when crowding discomfort is considered diseconomies of scale are found for high levels of patronage, a result that vanishes when more than one line is considered (Tirachini *et al*., 2010b). Tirachini and Hensher (2011) and Jara-Díaz and Tirachini (2013) have studied the impact of the boarding-alighting-paying methods, finding yet another source of economies of scale. Considering different modes also impacts the analysis, as shown by Tirachini and Hensher (2012) or Basso and Jara-Díaz (2010, 2012).

4.2. The impact of the discrete nature of lines structure choice on *DSE*.

As discussed in chapters 2 and 3, how adequate a lines structure is depends on the flow pattern, such that the set of transit routes becomes a design variable that should be optimized together with frequencies and vehicle sizes. Taking this into account, the question is whether adjusting line structures contribute to scale economies in transport networks.

Considering operators' costs only, Basso and Jara-Díaz (2006b) study the difference in the analysis of scale economies when lines structures are fixed or a variable to be optimized. Kraus (2008) formulates the problem over a general network including users and operators' costs. He finds an expression for optimal frequency for every possible line in a network ("where each path of the network is a potential bus line", pp.175); by plugging the result back into the cost formulation (envelope theorem) he argues that in a cost-minimizing network (i.e. "one for which the sum of user and capacity costs for the network's outputs is at a minimum", pp. 171), the DSE is not affected by a consideration of lines structures. In public transport, however, passengers choose their routes aiming at the minimum individual cost, which yields a pattern that differs from the total costs minimum because of the presence of externalities. As we will see, the emergence of a new line (i.e. a change in lines structure) that occurs at some point of the continuous growth in flow, happens with its frequency jumping discretely from zero to a value that makes the new line attractive. This discrete change in frequency prevents the envelope theorem to be applied. In addition, the problem of finding an optimal lines structure has been shown to be NP-Hard, such that in big-size networks - as is the case in any real city - it is unfeasible to conceive "every possible line".

This motivates the proposition expressed below, where we show that the degree of scale economies increases when the lines structures is changed (unless common lines exist everywhere[11]), taking into account that frequencies $f_L$on every line are a function of flows q due to some underlying optimization procedure, i.e. $f_L = f_L(q)$. The proof is based precisely on the discrete jump made by frequencies from zero when passengers choose their routes minimizing their individual costs. An example is offered and analyzed in detail in the section 4.3.

Let us define a vector of OD flows q as a *threshold point* if there exists at least one line L such that frequency $f_L(q) = 0$ and $f_L(q(1+\varepsilon)) > 0 \ \forall \ \varepsilon > 0$ with no alternative lines for some of its passengers. This means that after a ray increase in q at least a new line appears, such that a *threshold point* implies a new lines structure.

Proposition 4.1: Consider a network served by a public transport system. Then at every threshold point the DSE increases discretely, i.e. $\lim_{\varepsilon \to o^+} DSE(q_\varepsilon) > \lim_{\varepsilon \to o^-} DSE(q_\varepsilon)$, with $q_\varepsilon = q \cdot (1+\varepsilon)$.

[11] In the literature the case known as "common lines" appears when for some portions of the route, the passenger is indifferent to choose within a certain set of lines because they all make almost the same trip. See for instance Chriqui and Robillard (1975) or Cominetti and Correa (2001).

Proof: the proof consists in two parts.

1) First, let us show that f_L increases in a discontinuous way from zero. Define f_0 such that if f_L< f_0, then the waiting cost for passengers using line L is larger than the total cost of any other route. In that case no passenger chooses L and the optimal frequency is zero. Then if $f_L > 0$, it must be larger than f_0.

Hence, the choice of an optimal line structure is essentially discrete, i.e. $q = F(X_{1,}\ldots, X_N, Z)$ where $X_{1,\ldots,}X_N$ represent continuous variables, such as frequencies or capacities, and Z represents a discrete variable: lines structure. The second part of the proof applies to any kind of production function that contains a discrete decision. It is developed for a single product Y, which is equivalent to a ray analysis for a multi-product scheme. In the rest of the chapter, Y is actually the total sum of the flows $Y = \sum q_i$.

2) Consider $Y = F(X_{1,}\ldots, X_N, Z)$. If Y is such that any increase leads to a change from Z_1 to Z_2 when choosing $(X_1, \ldots, X_N, Z)$ to minimize the production costs, then $\lim_{\varepsilon \to o^+} DSE(Y+\varepsilon) > \lim_{\varepsilon \to o^-} DSE(Y+\varepsilon)$. To prove this, consider the cost functions C_1 and C_2 associated to Z_1 and Z_2 respectively, i.e. $C_j(Y)$ represents the minimum expenditure to produce Y given Z_j. Let us look at the average and marginal costs for C_1 and C_2 at Y. As C_1 and C_2 are continuous functions, then $C_1(Y) = C_2(Y)$. Regarding the marginal costs, the derivative of the cost with respect to q verifies $\frac{\partial C_2}{\partial Y} < \frac{\partial C_1}{\partial Y}$, because C_2 becomes lower than C_1 when q grows. As average costs are equal and marginal costs are lower for C_2, it is direct to conclude that the ratio between average and marginal cost, i.e DSE, increases.

$$\lim_{\varepsilon \to o^+} DSE(Y+\varepsilon) = \lim_{\varepsilon \to o^+} DSE_2(Y+\varepsilon) = DSE_2(Y) > DSE_1(Y) = \lim_{\varepsilon \to o^-} DSE_1(Y+\varepsilon)$$
$$= \lim_{\varepsilon \to o^-} DSE(Y+\varepsilon)$$

Q.E.D.

This is represented in Figures 4.1, where average costs for C_1 and C_2 are shown. In the exact point where these two cost functions coincide (i.e. where the optimization process induces a change from Z_1 to Z_2), a black arrow shows that a) the marginal cost is lower for Z_2 and b) the global DSE increases discreetly.

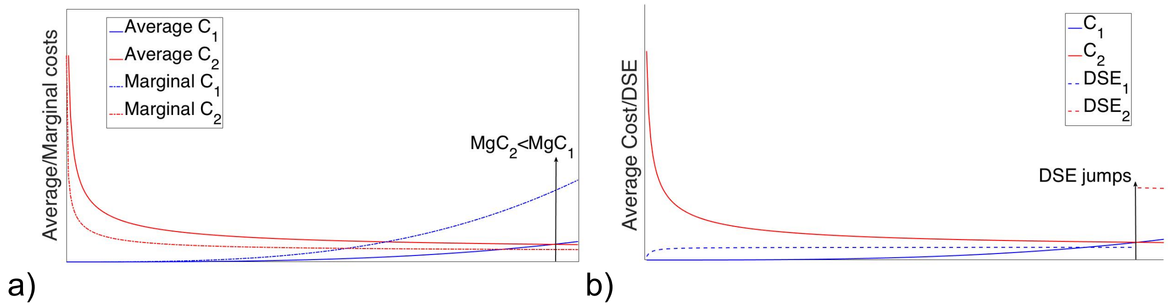


Figure 4.1. Change in *DSE* due to (discrete) change in lines structure

As a conclusion, the structural design of public transport systems involves variables as frequency f (and the associated fleet B), density D and vehicle capacity K, that can be treated as continuous, and lines structure, which has a discrete nature and introduces technical novelties that are worth studying. In the following section we introduce a multidimensional concept that helps analyzing the relation between lines structure and scale economies.

4.3. Introducing directness.

4.3.1 The concept.

We have proved that changes in lines structures always lead to a discrete (local) increase in scale economies. This general result, however, says nothing about what exactly are the transport-related technical elements that help understanding what lies behind this. We do know from the literature that increasing demand induces higher frequencies, larger vehicles and an increase in the density of lines. As a result, waiting, access and egress times diminish (scale economies) while in-vehicle and cycle times increase (scale diseconomies). What is the equivalent technical effect that links overall demand with lines structure and scale economies? And how do scale economies behave once a change in lines structure has occurred?

This is a quite complex question, as variables such as frequency, vehicle size or lines density can be represented by a single, well-defined continuous variable. A lines structure, however, can be conceptually described with some precision by a generic description, e.g. feeder-trunk or hub-and-spoke, but cannot be represented by a single variable. Further, changes in line structures are not continuous but discrete, occurring at some specific levels of total patronage. Both elements not only increase the mathematical complexity of the associated optimization problem, but also add new challenges to scale economies analyses.

Generally speaking the literature on lines structures in the last fifteen years shows that, for low levels of overall demand distributed in space, those structures involving transfers tend to be appropriate, e.g., hub-and-spoke or feeder-trunk systems[12]. As patronage increases, lines get organized along the idea of routes that follow more closely the origin-destination pattern avoiding transfers, increasing what can be called “directness”, such that each new passenger generates positive externalities on the rest of the passengers because a) transfers diminish, b) distances travelled diminish, and c) number of stops diminish[13]. Element a) has a clear positive impact on users, b) diminishes in-vehicle-time for all, and c) diminishes in-vehicle-time for users and cycle

[12] Gschwender *et al*. (2016), for example, study a Y-shaped city. They show that as the patronage increases, the optimal structure changes in one of the following ways (depending on trip distribution): from No transfers to No stops, from Feeder-trunk to No stops, or - the only odd case - from No transfers to Feeder-trunk. Daganzo (2010) studies a grid city served with direct lines within an internal region and with hub and spoke from the external region, optimizing the size of the internal region; he shows that the larger the patronage, the larger the zone served with direct lines (internal region). Badia *et al*. (2014) extend the paper by Daganzo (2010) and this conclusion remains valid; also, the set of lines becomes denser when the number of passengers increases.

[13] This is an extension of the concept of OD-directness originally defined by Laporte *et al* (2011) on the lines network as the fraction of the OD-pairs that can be joined without transfers.

time for operators. So these elements seem to contribute to increase the DSE through the reduction of average users' costs, but all effects should be analyzed. In order to represent directness in a more precise way, we propose the following three (continuous) indices: average transfers required per trip, average stops required per trip (including the extremes) and the average across all passengers of the ratio between their traveled distance and the length of the shortest path that link their origin and destination. Note that these flow-related indices can also be defined as averages across OD pairs, such that these new "network indices" can be calculated irrespective of the assignment of flows.

The concept of directness has an extreme case in non-stop services (which have been called exclusive in previous chapters[14]), where each OD pair is served by one line only, providing a service similar to a private car but with lower operating costs per passenger and larger waiting times. From this viewpoint, as directness increases the number of passengers with different origins and destinations sharing the same vehicle diminishes. It is worth noting that a connection between patronage and directness has emerged in the transit network design literature. For example, the parameter σ studied in chapter 3 for DBL and CW heuristics -that controls the maximum admissible deviations from the shortest paths- represents exactly the trade-off between more directness ($\sigma = 0$) and bus-sharing; σ is inversely related with directness. Recall that it was found systematically that small values for this parameter were optimal when patronage was large, i.e. increasing directness was the optimal response to demand increases.

4.3.2 An illustrative model

In order to illustrate in a simple way what has been discussed above, let us consider the network and flow characteristics as represented in Figure 4.2, where two destinations are located at the same distance L_0 from a single origin, forming an isosceles triangle; the distance between the destinations is Q (Figure 4.2a). The total number of passengers in the system is Y – half on each OD pair as represented in Figure 4.2b – and the question is whether it is better to have only one line carrying all the passengers (full bus-sharing, Figure 4.2c), or two lines, one for each destination (full directness, Figure 4.2d); λ represents the load of the lines on each directed arc. The directness indices are shown in Table 4.1 (note that in this case the flow indices and the network indices coincide, as there is only one flow assignment option).

Under this setting the characteristics of each passenger trip are known. Modeling the operators and users costs as in Jara-Díaz and Gschwender (2003b), a simple analysis (detailed in Appendix C) yields the results shown in Table 4.2, where capacities, fleets, cycle times, waiting times and in-vehicle times for each of these systems are expressed as functions of the corresponding frequency f, vehicle speed V, boarding-alighting time t, plus Y, Q and L_0. Note that in the two-lines case lines are symmetric and exhibit the same frequency.

[14] As anticipated in chapter 2, we change the nomenclature in this chapter because the "exclusive" structure presents a higher directness than the direct one.

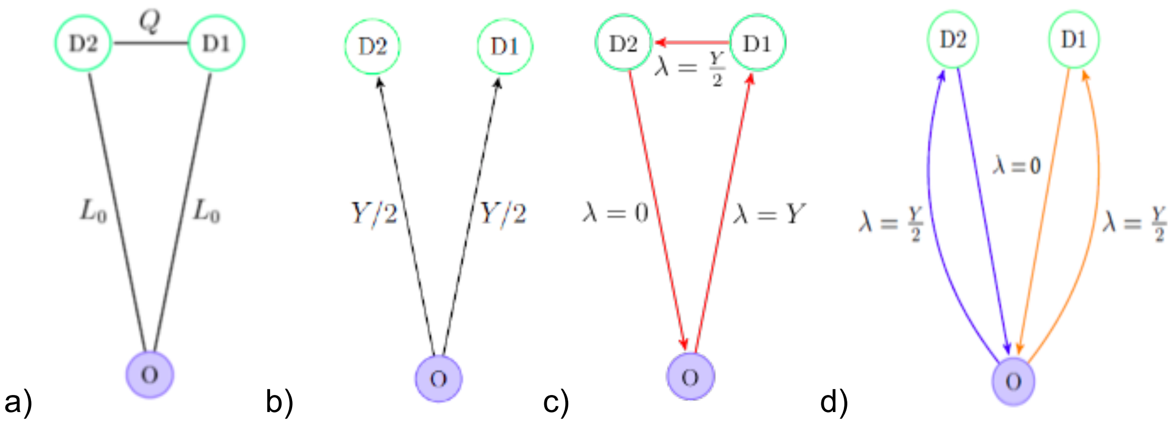


Figure 4.2. Network (a), transport demand (b) and alternative service structures: one shared line (c) and two direct lines (d).

SERVICE STRUCTURE DIRECTNESS INDICES	Bus-Sharing	Direct
Number of transfers	0	0
Number of stops	2.5	2
Distance traveled/Minimum distance	$1+\frac{Q}{2L_0}$	1

Table 4.1. Indices of directness for the alternative service structures.

	One line (Bus-sharing)	Two lines (Direct)
Bus capacity K	Y/f	$Y/2f$
Cycle time	$\frac{2L_0+Q}{V}+2t\frac{Y}{f}$	$\frac{2L_0}{V}+t\frac{Y}{f}$
Fleet B	$\frac{f(2L_0+Q)}{V}+2tY$	$\frac{4fL_0}{V}+2tY$
Waiting time t_w	$Y/2f$	$Y/2f$
In-vehicle time t_v	$\frac{1}{2}\left(\frac{L_0}{V}+\frac{1}{4f}tY\right)+\frac{1}{2}\left(\frac{L_0+Q}{V}+\frac{3}{4f}tY\right)$	$\frac{L_0}{V}+\frac{1}{4f}tY$

Table 4.2. Elements of the alternative service structures as a function of frequency.

Using the technical relations from Table 4.2 the VRC can be written as a function of frequency and optimized as shown in Appendix C. Optimal frequencies and capacities are shown to increase with Y (as in Jansson, 1984), such that the scale effects (explained in section 4.1) are preserved. Optimal frequencies can be substituted in VRC in order to obtain the cost function C_i for each system:

$$C_1 = 2\sqrt{\frac{c_0(2L_0+Q)}{V}Y(2c_1tY + \frac{p_w+p_vtY}{2})} + 2c_0tY + \frac{c_1Y(2L_0+Q)}{V} + \frac{p_vY}{2}\frac{(2L_0+Q)}{V} \quad (4.1)$$

$$C_2 = 2\sqrt{\frac{c_0 4L_0}{V}Y(c_1tY + \frac{p_w+p_vtY/2}{2})} + 2c_0tY + \frac{2c_1YL_0}{V} + p_vY\frac{L_0}{V} \quad (4.2)$$

Note that C_1 and C_2 can be written as $C_i(Y) = \sqrt{\alpha_iY^2 + \beta_iY} + \varepsilon_iY$, with $\alpha_1 > \alpha_2,\ \varepsilon_1 > \varepsilon_2$ and $\beta_1 < \beta_2$. For high values of patronage, α and ε dominate, so the 2-lines structure (full directness) is better, because the shorter routes become more relevant for both users (through p_v) and operators (through c_1). On the other hand, when Y is small, β dominates, such that the system with only one line (full bus sharing) is better due to the lower waiting times (through p_w). The average costs resulting from C_1 and C_2 are shown in Figure 4.3a using the parameters shown in Appendix A. DSE is represented in Figure 4.3b for each system, with the solid lines representing DSE for the optimal structure.

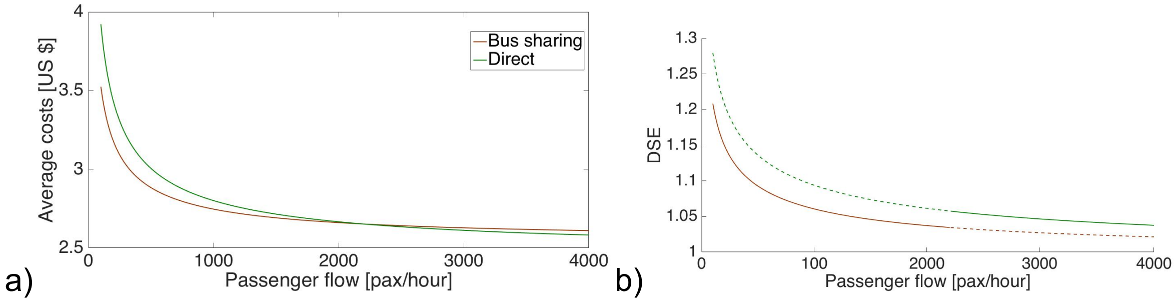


Figure 4.3. Average costs (a) and *DSE* (b) for Bus-sharing and Direct services.

The general property advanced in section 4.2 and Figure 4.1 emerges very clear: the DSE "jumps" when Y reaches a certain volume that makes the direct lines superior, which is explained because of more direct routes and fewer stops. What about DSE after the lines structure changes? Using the short notation introduced above DSE can be expressed as

$$DSE_i = 1 + \frac{\beta_i}{2\alpha_iY + \beta_i + 2\varepsilon_iY\sqrt{\alpha_i + \beta_i/Y}} \quad (4.3)$$

This expression shows that economies of scale are always present, but $\lim_{Y\to\infty} DSE = 1$, suggesting that the positive externalities induced by each of the elements that constitutes "directness" in this model get exhausted in spite of the upward jump in DSE induced by the change in lines structure: eventually everybody travels along the shortest possible route and with no intermediate stops.

4.4. Analysis over the parametric city.

What would happen if the underlying spatial setting was better represented such that lines could be structured following many possible arrangements? In order to examine the role of the evolution of total costs as patronage increases with an impact on the transit lines structure under a more general setting, we use the parametric city and the four basic lines structures described in chapter 2.

The trips paths followed by the passengers are not known a priori because they depend on optimal frequencies (some of which could be zero) that in turn depend on Y. In order to characterize the structures in terms of directness independent of Y, Table 4.3 shows the network indices of the four structures calculated as averages across OD pairs – instead of passenger trips – in a city with eight zones (n=8, 136 OD pairs)[15]. Directness increases from FT to HS, then to NT and finally to NS.

Structure	FT	HS	NT	NS
Number of transfers	0.47	0.35	0	0
Number of stops	3.06	3.06	3.06	2
Distance traveled/Minimum distance	1	1	1	1

Table 4.3. Network indices describing directness for each lines structure.

Figure 4.4a shows the results of Fielbaum *et al* (2016) regarding the average cost of each line structure; as Y increases the optimal structure changes from hub and spoke, to no transfers and finally to no stops, i.e., directness increases (and feeder-trunk is never optimal). In Figure 4.4b this evolution is shown by means of the corresponding DSE of the optimal structure for each level of the total flow: scale economies indeed increase after each change (including the emergence of the circular line in HS, which is a change in lines structure rigorously speaking), and decrease thereafter. For synthesis, *the possibility of deciding the line structure introduces directness as a new source of economies of scale which are finally exhausted after full directness is achieved.*

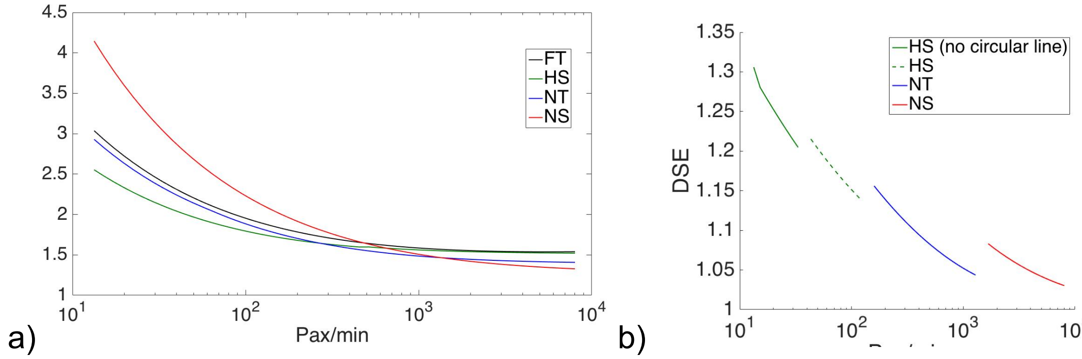


Figure 4.4. Average costs and overall *DSE* as directness increases.

Which design elements lie behind these results regarding scale economies? Having found the superior structures, an analysis of directness can be made taking into account the passengers' trips. Figures 4.5 show the evolution of each of the three flow indices

[15] Note that whenever some lines vanish as a result of the optimization process (zero frequency) the flow directness indices may increase.

that define directness as a function of the number of passengers whose growth induces lines structure changes from HS to NT to NS.

As represented in Figure 4.5a, transfers occur only for low values of Y where the HS dominates; the emergence of the circular line generates a reduction in the number of transfers and also in the number of stops and distance traveled, which shows up in Figures 4.5b and 4.5c. The average stops per trip decreases down to 2 when the no-stops structure dominates for high values of Y (Figure 4.5b). The ratio between the distance traveled and the minimum distance possibly required (called “detour” in Figure 4.5c) generally decreases except when changing from HS to NT, as some passengers experience longer trips because some short lines disappear in favor of longer ones that collect more passengers; note that this is counterbalanced by the reduction in transfers, showing that sometimes there is a trade-off between the different components of directness.

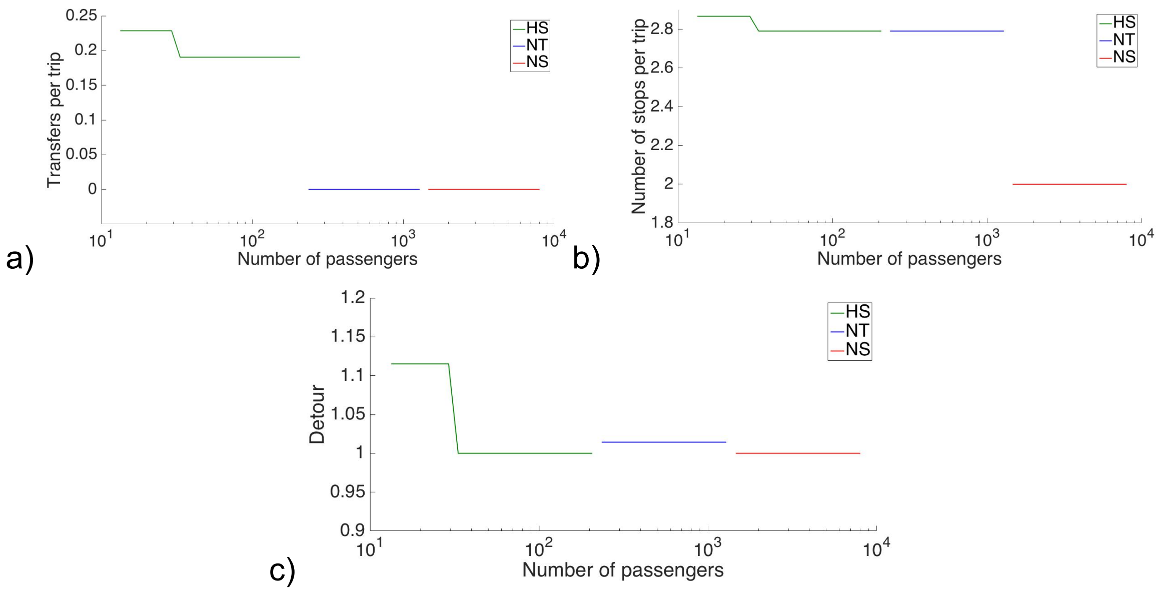


Figure 4.5. The three flow indices of directness as a function of patronage.

The physical measures of directness translate into users’ time and users’ costs, which are shown in Figures 4.6. Figure 4.6a summarizes the “equivalent time” associated to each of the directness indices: length of the routes translates into time-in-motion, the number of stops (together with vehicle load) translates into time at stops, and each transfer is valuated as 24 minutes in motion (as in Fielbaum *et al*, 2016). Their sum is the total equivalent time (TET) presented at the top of Figure 4.6a, and it synthesizes the total effect of directness on users; the fact that TET diminishes when lines structure changes clearly shows that increasing directness as patronage increases, contributes to scale economies. The slight increase of TET within each structure is caused by the larger time at stops induced by larger vehicles, an effect that is almost irrelevant when compared with the rest including the reduction in the number of stops each time the structure changes. Note that the more than 10 minutes reduction of TET is mostly explained by the reduction in time-in-motion and transfers (some 4 minutes each) against the 2 minutes reduction in time at stops.

Figure 4.6b shows the average costs per passenger due to in-vehicle time, waiting time and transfers, which are the three components of the users' cost function. Looking at the points where lines structure changes, it becomes evident that increasing directness makes in-vehicle time and transfer cost decrease, but there is a local increase in waiting time because directness diminishes bus-sharing and each passenger now has less lines to choose from. This local increase in waiting times, however, is more than compensated by the frequency growth as patronage increases within each structure (Mohring effect).

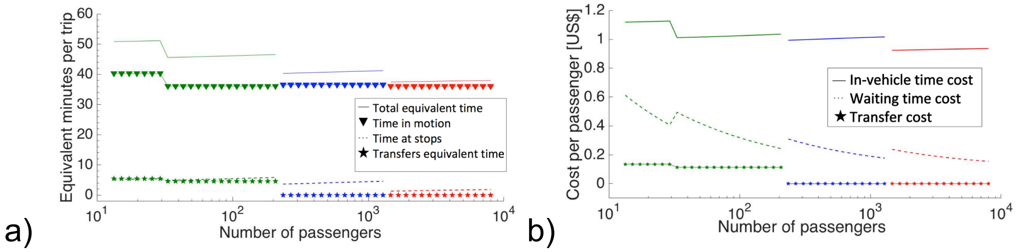


Figure 4.6. Effects of directness on equivalent users' times and users' costs.

So far, we have interpreted scale economies in terms of users' costs; what about operators' costs? Which are the effects of directness? To tackle these questions, let us recall that total operators' costs are given by $c_0 B + c_1 \Sigma$ where B is total fleet and Σ is total number of seats. Let us analyze both variables.

In Figures 4.7 we show (a) number of seats per passenger and (b) number of buses per passenger as a function of patronage. Seats per passenger drop significantly when lines structure changes. This effect occurs because bus-sharing diminishes (when directness increase) reducing the idle capacity of buses as we now explain in detail. The size of the buses for a line is given by its most loaded segment, such that idle capacity is present in the rest of the arcs used by the line; only in the NS structure buses are always full. Within a given structure increasing Y increases cycle time through boarding-alighting time; this makes Σ/Y an increasing function of Y

Figure 4.7b reveals that the number of vehicles per passenger decreases nearly in a continuous way, which shows that the effect of the change in lines structure over total fleet as Y grows is less important than the increase in bus size. In other words, when Y increases, optimal frequencies and vehicle capacities increase as well, but frequency grows at a decreasing rate precisely because the capacity grows making fleet per capita decrease.

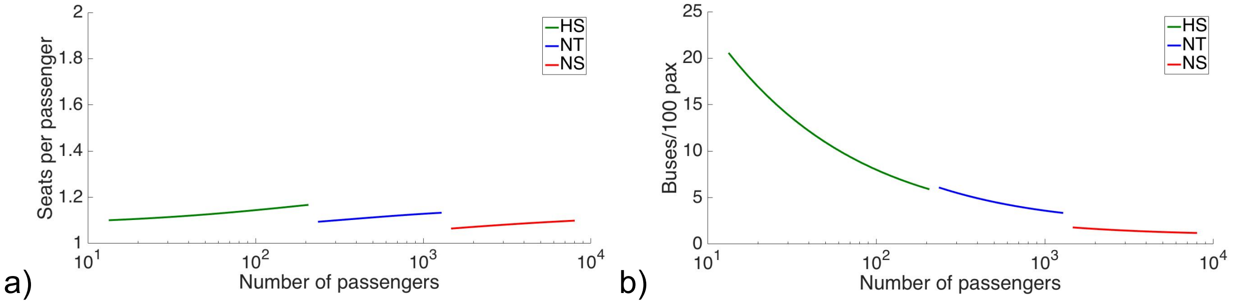


Figure 4.7. Effect of directness on the components of operators' costs.

In summary, including lines structures as part of the (optimal) design of public transport services in an urban space introduces yet another source of scale economies which has been defined here as directness, a concept that encompasses many elements summarized by three indices that capture transfers, routes length, and stops; as directness increases the total equivalent time for users decreases, approaching the (time related) characteristics of a private car trip.

4.5 Some results on subsidies and fares

Scale economies are strongly related with the calculation of fares and subsidies. When scale economies are present, it is efficient to have a larger production than what is reached in an unregulated equilibrium, and the simplest way to achieve this is using subsidies (which reduce fares). The usual expression for the subsidy is the difference between the average cost AC and the marginal cost MC:

$$S = AC - MC \tag{4.4}$$

Regarding the fare, a distinction needs to be done between monetary and non-monetary costs. Only operators' costs are monetary, and they are partially covered by the subsidy, such that the difference must be fulfilled with the fare:

$$F = AC_O - S \tag{4.5}$$

Using that total cost is the sum of operators' and users', (4.5) can be also written as:

$$F = AC - AC_U - S = MC - AC_U \tag{4.6}$$

Expression (4.6) shows that users pay the marginal cost of the production (as usual) minus what they are already spending by means of their times. With (4.5) and (4.6), fares and subsidies can be calculated for the different models studied so far. Let us begin with the single-line model studied in chapter 2:

$$F = c_0 t + \frac{c_0 T t l (p_v + c_1)}{L\sqrt{c_0 T\left(\frac{p_w}{2Y} + \frac{tl}{L}(p_v + c_1)\right)}} \quad \text{and} \quad S = \frac{c_0 T p_w}{2\sqrt{c_0 T Y\left(\frac{p_w}{2} + tY\frac{l}{L}(p_v + c_1)\right)}} \tag{4.7}$$

Note that, as the number of passenger grows, subsidy per passenger decreases at a decreasing rate, asymptotically approaching 0; this occurs because scale economies get exhausted. The fare, on the other hand, increases at a decreasing rate and converges to

$$\lim_{Y \to \infty} F^* = c_0 t + \sqrt{c_0 T t \frac{l}{L}(P_v + c_1)} \tag{4.8}$$

As shown above, average and marginal costs are enough to calculate and represent optimal price and subsidy for all levels of Y. These are shown in Figure 4.8, where it is represented that increasing Y diminishes the subsidy S per passenger.

It is interesting to note that although the subsidy per passenger decreases steadily with patronage, the total subsidy YS shown in equation (4.9) increases with Y up to a limit given by equation (4.10).

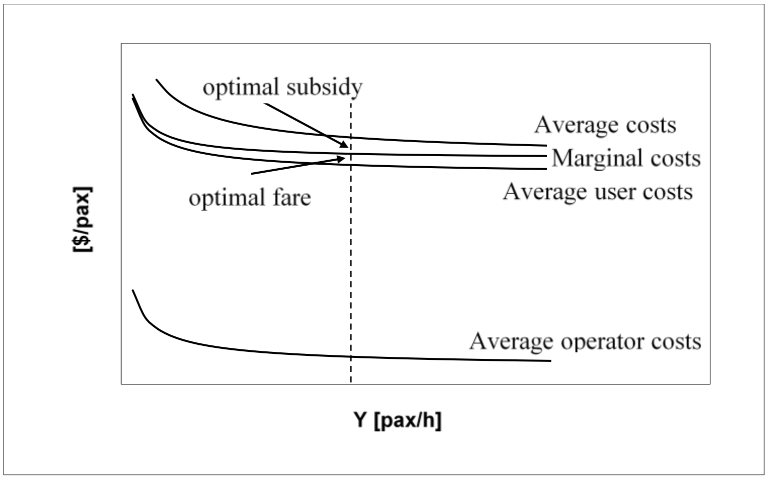


Figure 4.8. Fares and subsidies in the single-line model.

$$YS^{*} = \frac{c_0 T P_W}{2\sqrt{c_0 T\left[\frac{P_W}{2Y}+t\frac{l}{L}(P_v+c_1)\right]}} \tag{4.9}$$

$$lim_{Y\rightarrow\infty}\, YS^{*} = \frac{P_W\sqrt{c_0 T}}{2\sqrt{t\frac{l}{L}(P_v+c_1)}} \tag{4.10}$$

These general effects are preserved as the system gets more complex. In the illustrative model (section 4.3) using that the subsidy is given by:

$$S = AC - MC = (DSE - 1)MC \tag{4.11}$$

And recalling that the degree of scale economies converges to zero (equation 4.3), we conclude that subsidy per passenger also converges to zero. Further, from equations (4.1) and (4.2) marginal costs are also convergent, while (4.3) shows that the same happens with $Y \cdot DSE$, such that total subsidy YS^{*} increases up to a limit again.

What happens at the threshold points where lines structure changes? As average costs are the same but marginal costs are lower in the emerging structure, from (4.4) it is apparent that subsidy should increase. Nevertheless, an analogous conclusion for fares cannot be obtained because it depends on the comparison between the respective operators' and users' costs. Figure 4.9 shows subsidy and fare for the system from section 4.3, with their respective "jumps"; it can be observed that fare decreases (discretely) when the change in lines structure occurs.

Regarding the parametric city (section 4.4), the only analytical result is obtained regarding convergence: as the no-stops structure is the dominant for every large value of Y, and because this structure is just a combination of single lines that do not interact, the result from the single-line model is preserved and total subsidy is an increasing but convergent function of Y.

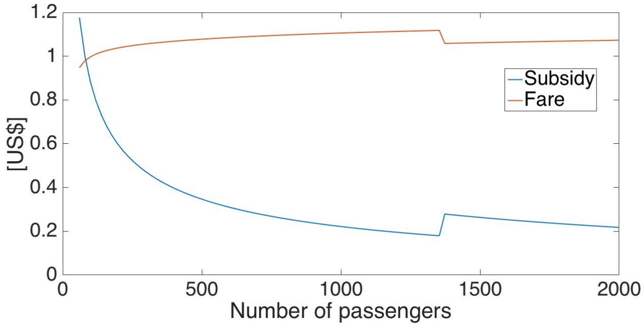


Figure 4.9. Subsidies and fares in the illustrative model.

Lower levels of the patronage can only be studied numerically. Figure 4.10 shows that subsidy decreases and fare increases with the number of passengers, but the contrary occurs (discretely) each time the lines structure is changed, which is exactly the same effect observed in Figure 4.9 for the illustrative model.

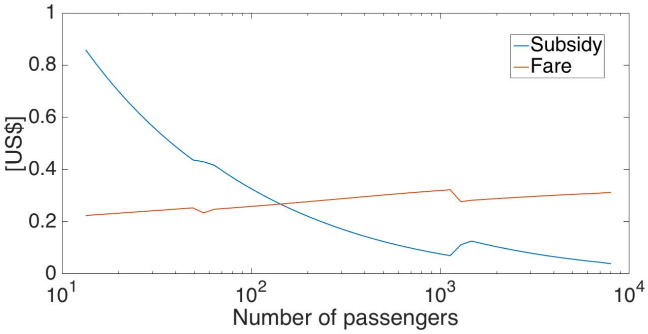


Figure 4.10. Subsidies and fares in the parametric city.

4.6 Main conclusions

- When the number of passengers increases, there are specific points in which the lines structure changes. At these points, while the average cost remains continuous, marginal cost decreases discreetly, inducing a discrete jump on the degree of scale economies.
- The engineering aspects behind this relationship between lines structure and scale economies can be explained by a novel concept: the “directness”, which encompasses average number of transfers, average number of stops and average length of the routes.
- All these indices are shown to improve (decrease) each time the lines structure changes in a very simple network as well as in the parametric city described in chapter 2, with some few exceptions that show that, ocassionally, there is a trade-off between these 3 components of directness.

- Each time lines structure changes, waiting time increases and idle capacity decreases.
- While the lines structure remains unchanged, scale effects previously identified for single-line models remain valid.
- Scale effects related to directness get exhausted when the number of passengers is too large, such that the degree of scale economies tends to 1.
- In every system, optimal subsidy per passenger decreases to zero continously, but with discrete jumps each time lines structure changes. Total subsidy converges to un upper bound.

Chapter 5. Introducing lines density in the strategic design of transit networks.

The parametric city that we have used in chapters 3 and 4 has a relevant limitation that is troublesome: streets are inevitably aggregated because of the simplicity of the network representation, i.e., arcs usually represent many arcs. This leads to models that - by construction - ignore the access to the transit network as a variable and treat every arc as a collector of many lines that actually run along different parallel streets. This is not a minor thing, as it also has an evident impact on the calculation of waiting times which are underestimated because the modeled frequencies are the sum over possibly many lines that run in parallel (which, in turn, affects all other variables). The different strategic line structures are affected in different ways by this omission; for example, structures that involve lower frequencies may be favored. The impact of density on waiting times and transfers in addition to the introduction of access time as a new element has also an effect on the analysis of scale economies in transit networks. The challenge we address in this chapter is to incorporate spatial density of transit lines as a design variable without going into a highly detailed description of its network.

5.1 Lines density: the parallel lines model revisited.

Chang and Schonfeld (1991) incorporated in the single-line analyses a first idea of the spatial density of transit lines, considering the distance R between parallel lines (spacing) such that the time spent walking to the bus stops (access time) becomes part of users' costs[16]. Implicitly, they considered a very fine grid street pattern and assume that all passengers Y – homogeneously distributed in space - need to travel to a faraway point where all lines converge, as represented in Figure 5.1.

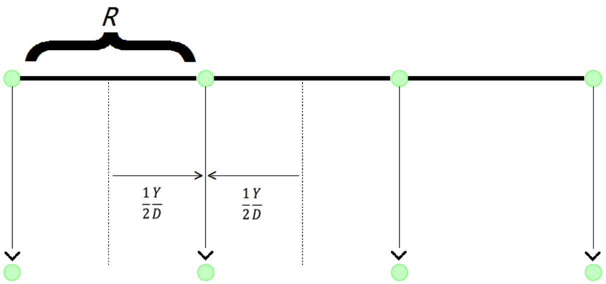


Figure 5.1. Chang and Schonfeld's transit design problem.

The decision (design) variables are the spatial separation of the lines and their frequencies, considering that each passenger uses the closest line. The model is built assuming two strong simplifications, namely that vehicles' cycle time does not depend on the number of users, dismissing boarding and alighting time, and that operators' costs do not depend on vehicle size (in other words, they assume that $t = c_1 = 0$). Under Chang and Schonfeld's assumptions, the optimal frequency follows a cube root rule and the average cost happens to be the sum of a constant plus a term that is inversely proportional to the cube root of Y, such that economies of scale are present. It

[16] Kocur and Hendrickson (1982) considered a similar model, but user costs are not shown explicitly.

is relevant to point out that this decreasing term is related to the waiting time, to the operators' costs[17] and to the access time, as increasing patronage induces larger frequencies and smaller spacing, diminishing access time (two positive externalities). This last effect is a novelty regarding Mohring's and Jansson's approaches, so it is worth analyzing it further. What happens if we drop the two simplifications of the model?

Let us keep the same spatial representation and formulate operators' and users' costs including the two omitted effects: patronage influences cycle time through boarding and alighting, and vehicle size influences operators' costs. As transit lines run in parallel, we can work either with spacing or with its inverse, the spatial density D (number of lines per unit width). To calculate the value of the resources consumed per unit of time and width, it is necessary to express the different components of VRC as a function of the decision variables (frequency and spacing), noting that cycle time t_c is given by $t_c = T + t\frac{Y}{fD}$. Then $B = t_c fD$, $K = \frac{Yl}{fDL}$, $t_v = T\frac{l}{L} + t\frac{l}{L}\frac{Y}{fD}$, $t_w = \frac{1}{2f}$ and $t_a = \frac{1}{4Dv_a}$ with v_a the walking speed. Passengers are now divided into fD buses.

Then the value of the resources consumed per unit of time and width can be easily shown to be

$$VRC = c_0 TfD + c_0 tY + c_1 Y\frac{l}{L}T + \frac{c_1 tlY^2}{fDL} + p_a\frac{Y}{4Dv_a} + p_w\frac{Y}{2f} + p_v Y\left(T\frac{l}{L} + \frac{Ylt}{fDL}\right) \qquad (5.1)$$

Equation (5.1) shows the effect of each of the two decision variables very clearly. It is evident that analytically D acts very similar to f, as they always appear as a product with the exception of the two terms dealing with their direct specific effects on access time (D) and waiting time (f). Therefore, D is yet another source of scale economies, just as the so-called Mohring effect through f. Note that the single-line model explained in chapter 2 can be looked at as a particular case with $D = 1$ such that access time plays no role. Several results are going to be obtained analyzing (5.1) and based in this first general conclusion:

Proposition 5.1: at each of the parallel lines of this model, the relation between frequency and number of passengers is the same as in the single-line model.

Proof: Deriving (5.1) with respect to f and multiplying times f^2D:

$$0 = c_0 Tf^2D^2 - c_1 t\frac{l}{L}Y^2 - p_w\frac{Y}{2}D - p_v t\frac{l}{L}Y^2 \qquad (5.2)$$

Deriving (5.1) with respect to D and multiplying times fD^2:

$$0 = c_0 Tf^2D^2 - c_1 t\frac{l}{L}Y^2 - p_a\frac{Y}{4v_a}f - p_v t\frac{l}{L}Y^2 \qquad (5.3)$$

The right hand sides of equations (5.2) and (5.3) have three identical terms, such that the remaining ones have to be equal, which means that at the optimum

$$f = uD \text{ with } u = 2\frac{p_w v_a}{p_a} \qquad (5.4)$$

Using equation (5.4), equation (5.2) can be re-written as

[17] When patronage increases, both optimal frequencies and optimal bus size increase. As vehicle size is not considered in the cost function for this model, this conclusion is not particularly relevant.

$$0 = c_0 T \frac{1}{u^2} f^4 - c_1 t \frac{l}{L} Y^2 - p_w \frac{Y}{2}\frac{f}{u} - p_v t \frac{l}{L} Y^2 \qquad (5.5)$$

Dividing it by f^4 it becomes evident that $\frac{df}{dY} > 0$, and then $\frac{dD}{dY} > 0$. Equation (5.5) has f to the power of 4 and the analytical solution is quite complex. However, noting that the equation is quadratic in Y, it can be manipulated to obtain

$$\frac{Y}{D} = \frac{1}{2t\left(\frac{l}{L}\right)(c_1+p_v)}\left[\frac{-p_w}{2} + \sqrt{\frac{{p_w}^2}{4} + 4c_0 T t \left(\frac{l}{L}\right)^2 (c_1+p_v) f^2}\right] \qquad (5.6)$$

From equation (5.6) one obtains expression (5.7) for f which shows a remarkable novel result: without the limitations imposed by Chang and Schonfeld's assumptions, each of the lines (that carries Y/D passengers) replicates the same relation between frequency and patronage obtained in the one-line model, such that frequency (and vehicle capacity) increases with patronage per line:

$$f^* = \sqrt{\frac{\frac{Y}{D}}{Tc_0}\left(\frac{p_w}{2} + t\frac{Y}{D}\frac{l}{L}(p_v+c_1)\right)} \qquad (5.7)$$

Q.E.D.

Corollary 5.1.

- The number of passengers per line $y = \frac{Y}{D}$ increases with Y, because $\frac{dy}{dY} = \frac{dy}{df}\frac{df}{dY}$; the first factor is positive by equation (5.6) and the second was already shown to be positive.
- Bus size increases with Y because $\frac{dK}{dY} = \frac{dK}{dy}\frac{dy}{dY}$ and both terms are positive.

The relationship between Chang and Schonfeld's and the single-line models shown by equation (5.7) suggests some conclusions regarding scale economies. First, those sources of scale economies presented for the one-line model remains valid for each of the lines in this model as well: increasing patronage induces a larger frequency and vehicle size, diminishing waiting time, and also induces larger times at stops increasing both in-vehicle time and operators' costs. Second, introducing space adds a new dimension: the separation between lines R which, as shown earlier, diminishes with patronage inducing a reduction in access time, working in favor of scale economies. Note that the effect of Y on the optimal frequency in equation (5.7) is "softened" by lines density D (the inverse of R), which also increases with patronage such that the Mohring effect is mitigated[18]; as passengers are divided into many lines, buses become smaller when compared to Jansson's approach, which diminishes the diseconomies of scale provoked by the time at the stops.

[18] In the original (simpler) model by Chang and Schonfeld (1991) this reduced effect is captured by their resulting cube root formula for *f** that can be recovered here by making *t*=0 in equation (5.7).

Although in our improved model the very complex system of equations does not permit to obtain closed analytical expressions for the decision variables, we expect the degree of scale economies DSE to be larger than in the single line case as a third dimension that works in favor of scale economies has been added; we now show that this is in fact the case. In order to compare both models one should add access cost to the transit line in the single line model. As access time does not depend on the design, and therefore is constant per passenger, the total access cost is linearly increasing with Y, which means that in the total cost function it has the form qY, such that both average and marginal costs increase by q. As the single-line DSE is larger than 1 when access is not considered (i.e. when $q = 0$), it is easy to see that adding the same constant term to the average cost in the numerator and to the marginal cost in the denominator makes DSE smaller, such that DSE decreases with q. This means that DSE has an upper bound for $q = 0$. This upper bound is represented in Figure 5.2 (blue line) together with the DSE of the new model (red line). The new DSE is always larger than the upper bound of the single-line DSE, which shows that DSE increases when lines density is included. Note that scale economies get exhausted eventually. Simulation parameters are shown in the Appendix A; a sensitivity analysis on these parameters maintains these conclusions.

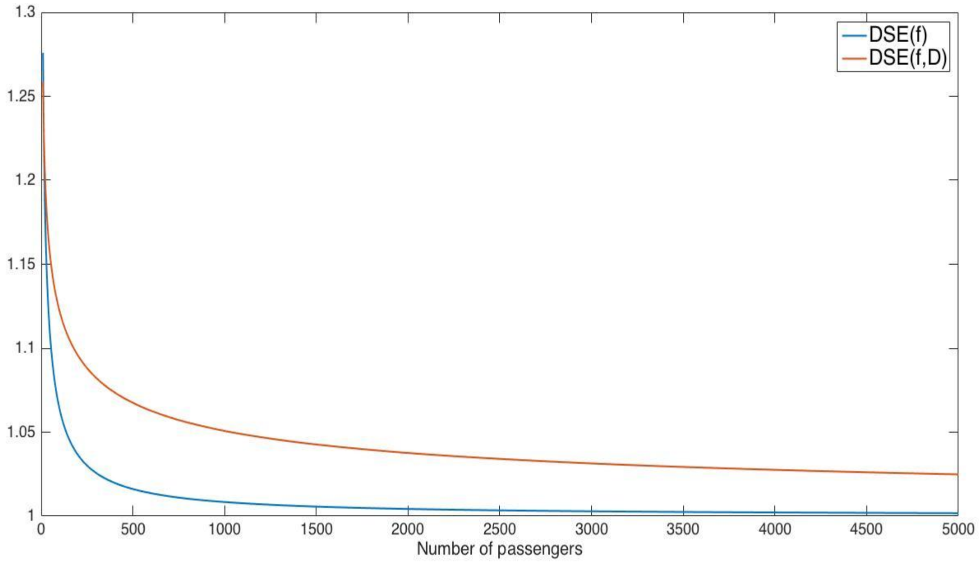


Figure 5.2. The effect of lines spacing on the Degree of Scale Economies

Let us now take a closer look at equation (5.4) that implies $\frac{f}{D} = 2v_a \frac{p_w}{p_a}$. An analogous property was also found by Chang and Schonfeld and other authors before them (Hurdle, 1973, Schonfeld, 1981 and Kocur and Hendrickson, 1991); the nice thing is that it remains valid even if the quite strong assumptions regarding costs depending on bus size and the absence of boarding-alighting time are dropped. This property states that frequency and lines density grow at the same rate, irrespective of the number of passengers, the boarding-alighting times, the distances traveled by buses or passengers, etc. The intuition behind these is quite attractive: the optimal fleet of vehicles of an optimal size could be distributed in a large number of lines with a small frequency or *vice versa*; what the result says is that this trade-off between D and f is resolved by making the average waiting time value $\frac{p_w}{2f}$ equal to the average access time value $\frac{p_a}{4Dv_a}$, which proves Corollary 5.2.

Corollary 5.2: Average waiting cost is equal to average access cost in this model.

5.2. Lines density in the parametric city

In order to analyze if previous conclusions are still valid in a more complex scheme, let us add the design variable D to the parametric city studied in previous chapters.

5.2.1 Lines density as a new design variable

How can we adapt the parametric city in order to consider also lines density? To do so, we will consider that each former line l is now a "super-line" containing D parallel lines per unit width, each one with the same frequency f_l. This design variable D represents the spatial density of lines. The frequency of the super-line F_l is then given by

$$F_l = f_l D \tag{5.8}$$

The introduction of D has two effects on the users' cost that have to be taken into account: each user has to walk to the closest line inducing an access time, and buses distribute on more lines diminishing perceived frequency, increasing waiting time. If P is the total width represented by each arc, passengers walk in average $\frac{P}{4D}$ to access the nearest line. We assume that whenever transfers are required the bus stops coincide in space such that walking is negligible. The design exercise was made using the four strategic lines structures explained in chapter 2, including D as a design variable in addition to frequency and bus size for every line. The value of the resources consumed now includes also access times and its value $p_a\bar{t_a}$, as shown in (5.9).

$$VRC = \Sigma_L B_L(c_0 + c_1 K_L) + Y(p_v\bar{t_v} + p_w\bar{t_w} + p_a\bar{t_a} + p_R\bar{R}) \tag{5.9}$$

Now we show that VRC can be expressed as a function of frequencies and D in a very compact way. First note that fleets, (optimal) bus sizes, in-vehicle times and transfers can be expressed as functions of frequencies only. In this case, they depend on the frequencies of the super-lines, i.e.

$$B_l = B_l(F_1, \dots, F_L), K_l = K_l(F_1, \dots, F_L), \bar{t_v} = \bar{t_v}(F_1, \dots, F_L), \bar{R} = \bar{R}(F_1, \dots, F_L) \tag{5.10}$$

The intuition behind these relations is straightforward. Fleets depend on the total number of buses running per unit time; buses should be large enough to carry the maximum load on each line, which again depends on the total number of buses per hour; in-vehicle time depends on total time in-motion (which is constant) and on time spent at bus stops, which depends on the number of passengers that board a specific bus, which in turn depends on the total number of buses; and finally, the number of transfers usually depends on the lines structure only, unless passengers assignment plays a role, but if that is the case D will play no role because access time is the same for every route.

Considering relations (5.10), we can rewrite VRC as a function of D and frequencies. If no common lines exist, then

$$VRC = G(f_1 D, \dots, f_L D) + \frac{\theta_a}{D} + \sum_l \frac{\theta_l}{f_l} \tag{5.11}$$

G is a differentiable function that encompasses all terms in equation (5.10) but waiting and access users' cost. $\theta_l$contains all the information related to waiting costs (like the number of passengers and their assignment, among others) for each line, respectively, and $\theta_a = \frac{p_a}{v_a}\frac{YP}{4D}$.

If common lines are present, the third term in equation (5.11) becomes more complex, as some passengers' routes can use more than one line. In this case, one needs to sum across the OD-pairs w rather than the lines:

$$VRC = G(f_1 D, \dots, f_L D) + \frac{\theta_a}{D} + \sum_{w \in OD} \frac{\theta_w}{f_1 \varepsilon_{1w} + \dots + f_L \varepsilon_{Lw}} \tag{5.12}$$

Where ε_{lw} is a binary variable whose variable is 1 if passengers of the OD-pair w use line l, and 0 otherwise[19]. Note that equation (5.11) can be written as (5.12), with only one $\varepsilon_{lw} = 1$ for each w, and

$$\theta_l = \sum_w \theta_w \varepsilon_{lw} \tag{5.13}$$

Now we can prove the following Proposition.

Proposition 5.2: total access costs equal total waiting costs in this scheme.

Proof:
Making the derivative with respect to f_l in (5.12) yields:

$$D\partial_l G - \sum_{w \in OD} \frac{\varepsilon_{lw}\theta_w}{(f_1 \varepsilon_{1w} + \dots + f_L \varepsilon_{Lw})^2} = 0 \Rightarrow D f_l \partial_l G = \sum_{w \in OD} \frac{\varepsilon_{lw}\theta_w f_l}{(f_1 \varepsilon_{1w} + \dots + f_L \varepsilon_{Lw})^2} \tag{5.14}$$

Here we are using ∂_l to represent the partial derivative with respect to the l-th variable in G, and we are omitting the arguments of the function G to simplify notation. Making the derivative with respect to D yields:

$$\sum_l f_l \partial_l G - \frac{\theta_a}{D^2} = 0 \Rightarrow \sum_l D f_l \partial_l G = \frac{\theta_a}{D} \tag{5.15}$$

Introducing (5.14) into the second equality in (5.15):

[19] If some OD pairs need to make some transfer, the θ corresponding to the intermediate stages of their trips take also those passengers into account.

$$\frac{\theta_a}{D} = \sum_l \sum_{w \in OD} \frac{\varepsilon_{lw}\theta_w f_l}{(f_1\varepsilon_{1w}+\cdots+f_L\varepsilon_{Lw})^2} = \sum_{w\in OD} \frac{\theta_w}{(f_1\varepsilon_{1w}+\cdots+f_L\varepsilon_{Lw})^2} \sum_l \varepsilon_{lw} f_l = \sum_{w\in OD} \frac{\theta_w}{f_1\varepsilon_{1w}+\ldots.+f_L\varepsilon_{Lw}} \tag{5.16}$$

Q.E.D.

The proposition proves that the property obtained in our generalized simple parallel lines model in Section 5.1 remains valid: at the optimum design, average waiting and access costs are equal.

The intuition behind this result is interesting, as it is not the usual microeconomic equality between some marginal costs. For any given fleet, optimizing D means deciding into how many parallel lines are we distributing the buses. If access cost is larger than waiting cost, then splitting the buses into more lines will induce savings in access costs that will outbalance the losses in waiting costs (analogously in the opposite situation); this happens due to the particular functional forms of both access and waiting costs, that depends on $\frac{1}{D}$ and $\frac{1}{f_l}$ respectively.

5.2.2. Results

Numerical analysis was done using Y as the variable. The rest of the parameters are found in the appendix A. The procedure to find the best lines structure for each Y has two steps (which are analogous to the steps followed in previous chapters): first, for a given structure, the optimal (social cost minimizing) frequencies are found for each line together with the optimal D. Second, the best lines structure is found as the one that exhibits the minimum VRC across structures for each Y. The results are shown for a wide range of passenger volumes, which makes the logarithmic scale preferable in order to facilitate the analysis for lower values of Y.

The optimal D for each of the four structures as a function of patronage (result of step one) is shown in Figure 5.3. Optimal lines density increases with Y within each line structure, which fits intuition and is consistent with results in section 5.1. Note that structures that are less direct (as defined in chapter 4), present, in general, larger D: their ability to collect passengers is useful to split each line into many.

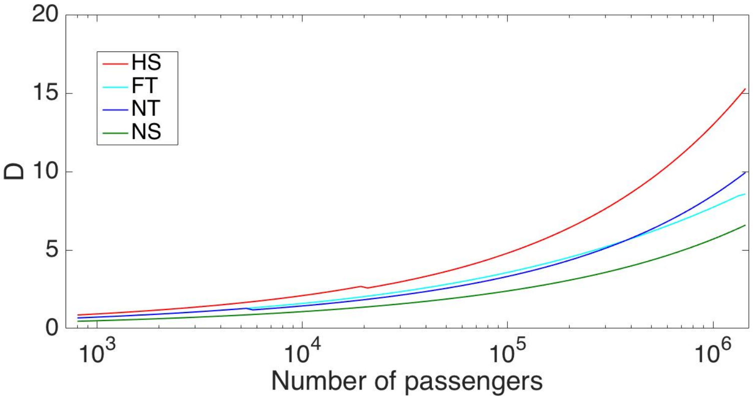


Figure 5.3. Optimal D per lines structure.

As the number of passengers increases, the best lines structure (i.e. the one that minimizes VRC) evolves from HS to FT, then NT and finally to NS. The D corresponding to the optimal structure is shown in Figure 5.4. Within each structure the lines density increases with the number of passengers, but it decreases locally when the line structure changes; this happens because “directness” increases, and it is necessary to compensate the fact that less passengers are being collected. Overall, however, D increases with Y.

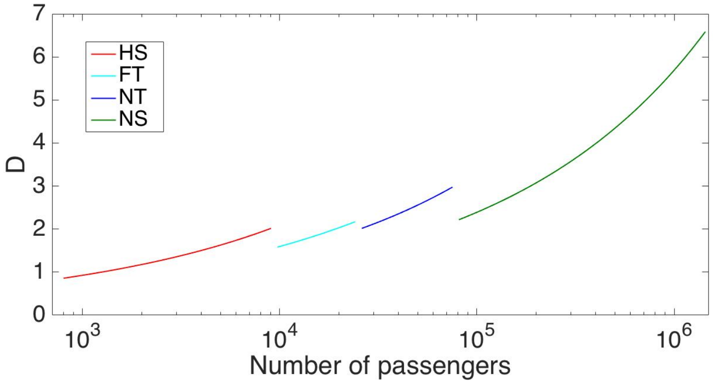


Figure 5.4. Optimal D.

In Figure 5.5 we show the impact of including D as a new design variable. The dotted lines represent the average cost curves of the best structures when $D = 1$ (fixed), while the solid lines show the result when D is optimized (i.e. the ones that correspond to Figure 5.4). In both cases the evolution is towards those structures that are more direct. The novelty here is that introducing D not only reduces VRC but also postpones the emergence of NT and NS (the most direct lines structures) as Y increases. This can be graphically seen by noting the lower levels of patronage at which a change in lines structure occurs along the dotted lines when compared with the solid lines in Figure 5.5. Note that the difference in average cost between the best structures considering D or not increases with Y from nearly zero to nearly 24%.

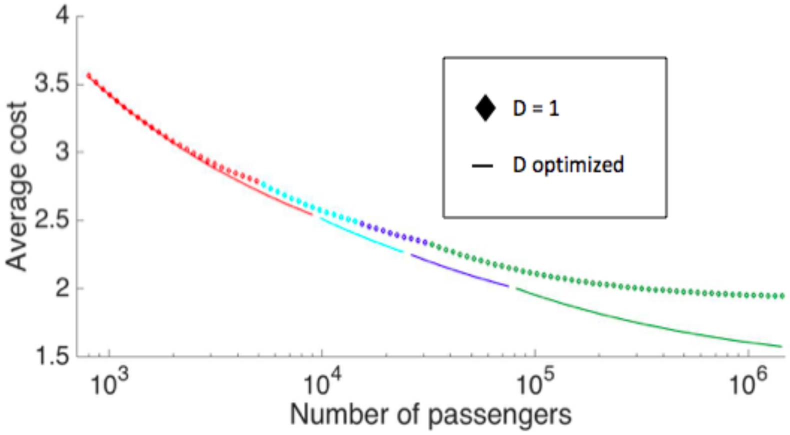


Figure 5.5. Comparison of average costs when lines density is fixed or optimized.

We have shown that under an optimal design average access and waiting costs are going to be equal. Their (joint) evolution as Y increase is shown in Figure 5.6, where one can see that for any given lines structure these costs decrease (the Mohring effect operates). But when the lines structure changes, this curve jumps upwards (just as D decreases locally). This is consistent with the results of chapter 4 when studying the impact of lines structure changes on waiting times.

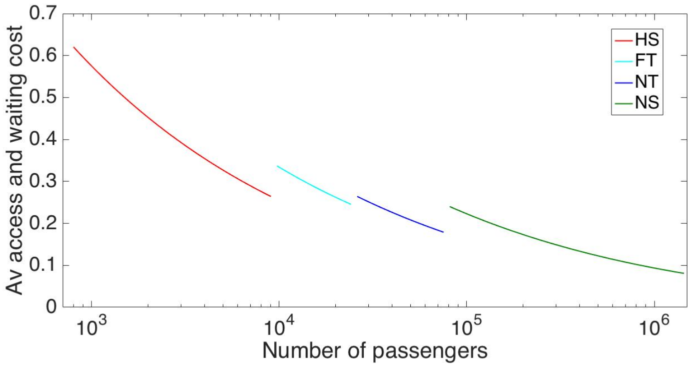


Figure 5.6. Average access (and waiting) costs.

In section 5.1 we studied the effect of considering lines density over scale economies. The same analysis can be replied here, by comparing the DSE obtained in this model and the DSE if access time is considered but with $D = 1$ fixed (Figure 5.7). When the same lines structure is being used in both models, DSE is larger in the model that optimizes D (solid lines); the only exception occurs at the beginning of the graph. This exception coincides with the zone in which the optimal D is lower than 1: frequencies are large in this zone, making the Mohring effect less relevant, which explains the lower DSE. Besides, at the specific points in which lines structure changes, the DSE jumps discreetly (as shown in chapter 4); as this happens earlier when D is not optimized, there are some points in which these two models do not present the same lines structure, and that show a larger DSE for the $D = 1$ model (specifically, when it turns from HS to FT). Nevertheless, in the big picture, the DSE is larger for the model that optimizes lines density, i.e., this is indeed a source of scale economies.

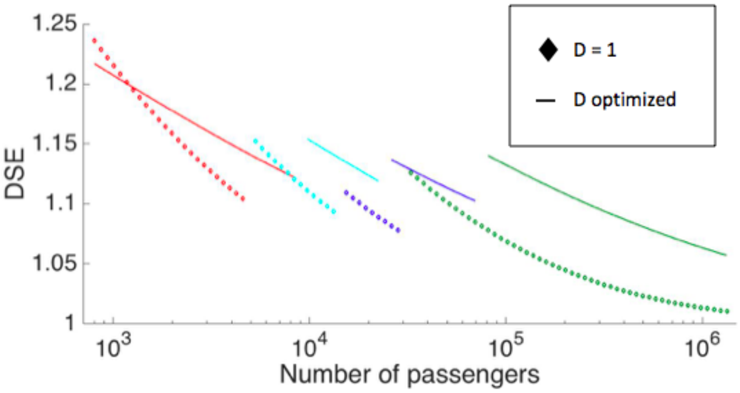


Figure 5.7. The effect of optimizing lines density on *DSE.*

As many previous studies have pointed out (see, for instance, Daganzo, 2010, Gschwender *et al*, 2016, or Fielbaum *et al*, 2016), the internal distribution of the trips has a strong impact over public transport design. This is why it is worth studying how does D respond to changes in demand parameters α, β and γ (recall that they represent the degrees of monocentrism, polycentrism and dispersion, respectively). Figure 5.8 shows the optimal D when $\alpha \in (0.1, 0.9)$ and $\beta = \gamma$, i.e., when the city evolves from an irrelevant CBD to almost full monocentrism. For this analysis, we are using $Y = 24{,}000$ passengers per hour, and optimal lines structure evolves from FT to HS; as the city becomes more monocentric (hence, trips are more concentrated such that it becomes easier to collect the users) lines density increases. It is worth explaining that the sudden decrease within HS occurs due to a new line emerging (i.e., its frequency departs from zero at that point).

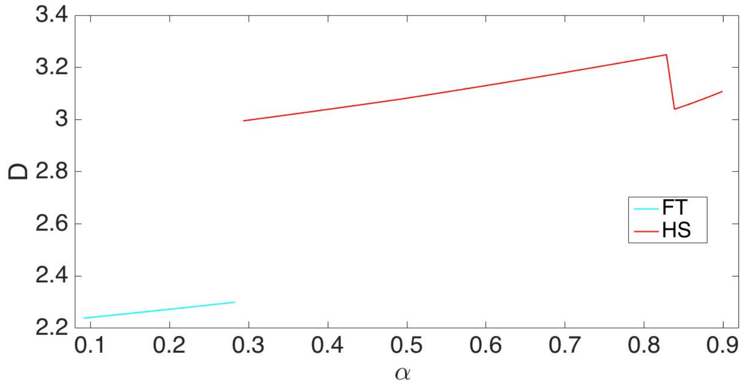


Figure 5.8. Optimal lines density when internal distribution of the trips changes.

Here we have treated D as a continuous variable: the number of lines per width unit. Nevertheless, as streets are exogenous, flexibility to adjust this variable is somewhat limited. This is why it is worth analyzing if results previously found are still valid if D is modeled as discrete. A graphical analysis (Figures 5.9) reveals that they are valid, but some approximations are needed. Figure 5.9a shows the optimal discrete D as a function of the number of passengers (same parameters as in Figure 5.4), Figure 5.9b shows the differences in average costs (similar to Figure 5.5), and Figure 5.9c reveals the ratio between access and waiting costs. This last Figure is quite interesting, as these costs are not equal anymore, because the discreteness of D prevents full adjustment; nevertheless, as Y grows, this ratio approaches 1 oscillating around it.

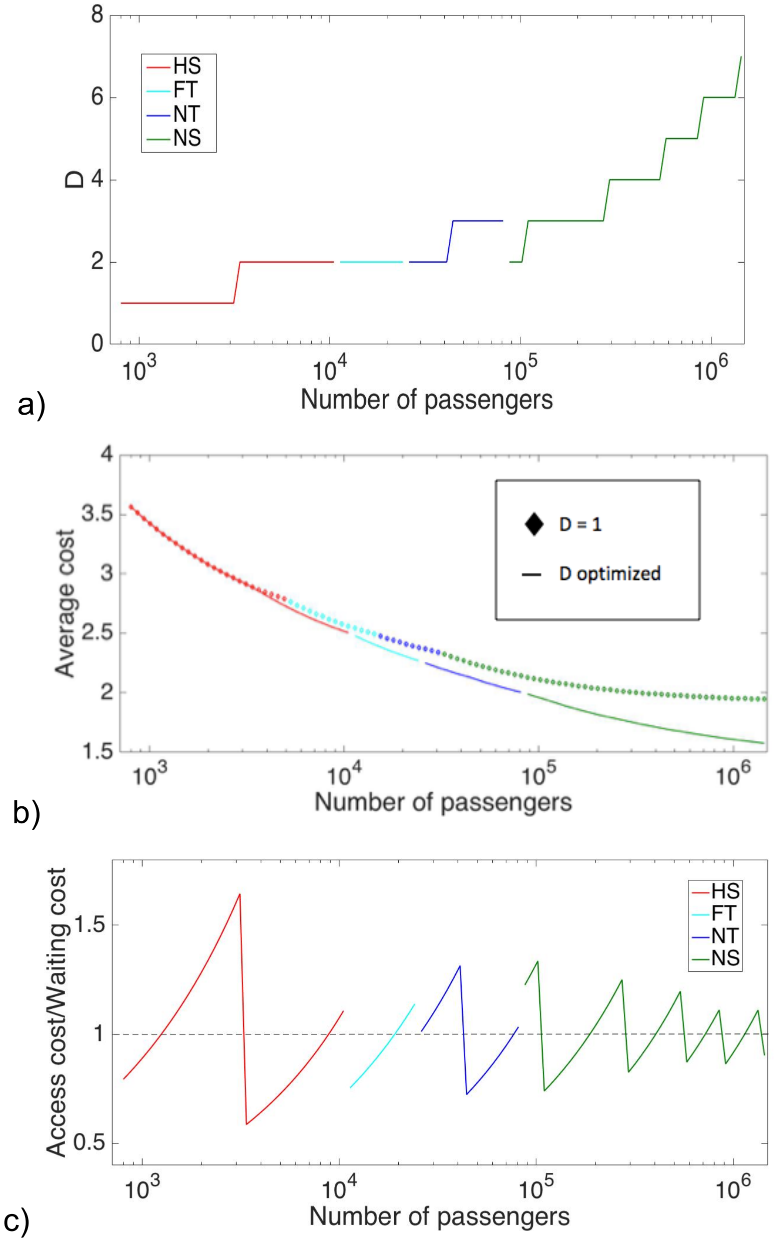


Figure 5.9. Results when D is discrete.

5.3 Main conclusions

- When the single-line model is extended to admit parallel lines, the relationship between each of these lines' frequency and number of passengers is the same as in the original single-line mode.
- In this same context, lines density induces a scale effect equivalent to the Mohring effect. The degree of scale economies increases due to the inclusion of this new design variable.

- In this context, as well as in the parametric city described in chapter 2, it is proved that lines density is adjusted to make average waiting cost equal to average access cost.
- Lines density increases with the number of passengers, but it presents a discrete decrease each time lines structure changes.
- Lines density increases if trips are more concentrated (as is the case for monocentric cities).
- Including lines density delays the emergence of more direct lines.
- All these conclusions remain (approximately) valid if lines density is modeled as a discrete variable.

Chapter 6. Two periods optimization over a single line

In this chapter, we return to the single-line model explained in chapter 2, but we consider that there are a peak and an off-peak period, such that the optimization process minimizes total daily costs, represented by the sum of these two periods' costs. It is assumed that a single fleet of buses is acquired for both periods; what we need to optimize is the size of this fleet, the size of its vehicles and how many of them are used during each period.

6.1 Formulation of the model

Recall that in the single-period model, explicit solutions can be found by deriving the $VRC.$ Before moving into the two periods formulation, it is useful to note that defining in equation (2.1) $A = Tc_0$ and $G = \frac{Y^2 lt}{L}(c_1 + p_v) + \frac{p_w Y}{2}$, the VRC can be re-written as $VRC = Af + \frac{G}{f} + W$, where W collects all terms that are independent of f and plays no role in the optimization problem. With this notation, the result of the optimizations process yields $f^* = \sqrt{\frac{G}{A}}$, an expression that will prove useful in the analysis that follows.

Let us extend this previous model to the case of optimizing simultaneously peak (denoted by *P)* and off-peak (denoted by *N*) periods. The general scheme is preserved, with a single fleet composed by buses of the same size K, but offering frequencies f_P and f_N in the peak and off-peak periods, respectively. The number of passengers per hour, the length of their trips, and the time needed to tour the whole circuit (related to the velocity of the buses) are also dependent on the period, so the following parameters are needed: $Y_P, Y_N, l_P, l_N, T_P$and T_N. The durations of each period are denoted by $E_P$and E_N.

Operators' cost must be divided into two types: capital costs (buying the buses and terminals) that are the same for all buses, and operational costs (like maintenance or fuel) that depend on how many hours the bus is used. Each of these components is well described by a function that is linear in vehicle size, as advanced in chapter 2. The capital costs are caused by the largest fleet needed considering both periods (usually the peak period); in the other period not necessarily all buses will be used. Operating costs will depend on the number of buses used during each period of length E_i. Therefore, the Value of the Resources Consumed in a day is now described by:

$$VRC_2 = max(B_P, B_N)(c_{BC} + c_{KC}K) + (B_P E_P + B_N E_N)(c_{BO} + c_{KO}K) + \frac{p_w}{2}(\frac{Y_P E_P}{f_P} + \frac{Y_N E_N}{f_N}) + \frac{p_v}{L}(l_P t_{cP} Y_P E_P + l_N t_{cN} Y_N E_N) \quad (6.1)$$

Two capacity restrictions must be fulfilled to be able to carry all the passengers at both periods. Optimality conditions assure that at least one of them must be active.

a) $K \geq \frac{Y_P l_p}{f_P L}$ and b) $K \geq \frac{Y_P l_p}{f_P L}$ (6.2)

The maximum fleet is associated with the largest demand flow which defines the peak period, making $max(B_P, B_N) = B_P$.

6.2 Buses full at the peak

Let us assume first that buses run full during the peak, as is commonly observed, making $K = \frac{Y_P l_p}{f_P L}$ (constraint 6.2a active). Replacing these and the values of B_i and $t_{ci}\,(i = P, N)$ with the equations for B and t_c described in chapter 2, the following expression is obtained:

$$VRC_2 = (f_P T_p + tY_P)(c_{BC} + c_{KC}\frac{Y_P l_P}{f_P L}) + [(f_P T_P + t\ \ Y_P)E_P + (f_N T_N + t\ \ Y_N)E_N][c_{BO} + c_{KO}\frac{Y_P l_P}{f_P L}\ \ ] + \frac{p_w}{2}(\frac{Y_P E_P}{f_P} + \frac{Y_N E_N}{f_N}) + \frac{p_v}{L}(l_P Y_P E_P[t\frac{Y_p}{f_P} + T_P] + l_N Y_N E_N[t\frac{Y_N}{f_N} + T_N]) \tag{6.3}$$

Expanding on the terms introduced in the single period case we define A_i and $G_i\,(i = P, N)$ and δ as

$$A_P = T_P c_{BC} + T_P E_P c_{BO} \tag{6.4}$$
$$G_P = \frac{tY_P{}^2 l_P(c_{KC} + c_{KO}E_P)}{L} + tY_N E_N c_{KO}\frac{Y_P l_P}{L} + \frac{p_w}{2}Y_P E_P + \frac{p_v}{L}l_p E_P Y_P{}^2 t \tag{6.5}$$
$$A_N = T_N E_N c_{B0} \tag{6.6}$$
$$G_N = \frac{p_w}{2}Y_N E_N + \frac{p_v}{L}l_N E_N Y_N{}^2 t \tag{6.7}$$
$$\delta = \frac{T_N E_N c_{KO} Y_P l_P}{L} \tag{6.8}$$

The value of the resources consumed can now be written as:

$$VRC_2 = A_P f_P + \frac{G_P}{f_P} + A_N f_N + \frac{G_N}{f_N} + \delta\frac{f_N}{f_P} \tag{6.9}$$

The derivatives with respect to frequencies in compact form are

$$\frac{\partial VRC_2}{\partial f_N} = (A_N + \delta\frac{1}{f_P}) - \frac{G_N}{f_N{}^2} \tag{6.10}$$

$$\frac{\partial VRC_2}{\partial f_P} = A_P - \frac{G_P + \delta f_N}{f_P{}^2} \tag{6.11}$$

Making these expressions equal to zero yields equations of order 5 for the optimal frequencies, such that no analytical solution can be found as established by the Abel-Ruffini theorem (see, for example, Alekseev, 2004). Although they will be solved and analyzed numerically in Section 6.2.4, some interesting properties can be deduced by inspection.

6.2.1 Comparing optimal and single period frequencies

Let us begin analyzing the differences between the optimal frequencies when jointly optimizing both periods and the sub-optimal solutions considering each period by itself.

We will start with an inspection of frequency in the peak period. From equation (6.11) the solution for $f_P$fulfills:

$$f_P{}^* = \sqrt{\frac{G_P + \delta f_N{}^*}{A_P}} \tag{6.12}$$

This expression can be compared with the solution for the one-period problem

$$f_P{}^1 = \sqrt{\frac{G}{A}} \tag{6.13}$$

Looking at the denominators, A and A_P are equal, as the latter is just the sum of the different components within the former. Regarding the numerators, $G_P > G$ due to the presence of the term that multiplies Y_N (involving c_{KO}); moreover, there is an additional positive term $\delta f_N{}^*$ in (6.12) that also involves c_{KO}. This yields

$$f_P{}^* > f_P{}^1 \tag{6.14}$$

i.e., the optimal frequency in the peak period is higher if the buses are also going to be considered for the off-peak period. The interpretation of inequality (6.14) is quite interesting. Looking at equations (6.12) and (6.13) the differences between the numerators in the squared roots are related with the optimal value of bus capacity (through c_{KO}), that has to be chosen taking into account that buses in this process are also going to run during the off-peak period; note that the effect through $G_P$is related with the off-peak time at stops while passengers board and alight, while the effect through δ is due to the time in-motion. For short, bus size becomes more relevant as increasing K is now more costly than in a single peak-period analysis, making it better to have smaller (cheaper) buses and higher peak frequencies. Note that another implicit result is that the optimal vehicle capacity is smaller than what a single peak-period analysis would yield; this fits the intuition that bus sizes should be somewhere in between peak and off-peak optimal sizes when considered in isolation. This result was overlooked by Jansson (1980, 1984) because in his first two-periods model bus size is costless, and he argues that $f_P = f_N$ would be optimal in most cases; in his second model (where bus cost depends on size) he solves the problem analytically assuming equality between frequencies, which is something examined below.

The analysis for $f_N{}^*$is similar but the result is inconclusive. To see this, note that from equation (6.10)

$$f_N{}^* = \sqrt{\frac{G_N}{A_N + \frac{\delta}{f_P^*}}}. \tag{6.15}$$

A_N does not include c_{BC} because fleet size depends on peak frequency, implying that the daily fixed capital cost for a bus is not affected by off-peak operations; therefore, A_N is smaller than A. However, there is a positive term involving c_{KO} that adds on A_N, induced by the cost associated to bus sizes that are higher than the size calculated when optimizing this period by itself. Therefore, the denominator may be larger or smaller than A depending on the relative values of c_{BC} and c_{KO} . Regarding the numerator, $G_N < G$ because all the terms involving bus size related costs disappear. Hence, the relation between ${f_N}^*$and ${f_N}^1$is unclear because there are elements that operate in opposing directions: on the one hand, an extra bus means no extra capital cost because total fleet is decided with respect to the peak-period (as in Jansson, 1980, 1984); this pushes ${f_N}^*$upwards. On the other hand, increasing ${f_N}^*$ does not induce a reduction in bus size (given by the peak) and, therefore, there are no operators' savings because of this; in addition, the larger buses increase operating costs, which pushes ${f_N}^*$ downwards. Which of these phenomena is going to dominate depends on whether daily fixed costs prevail over operating costs or vice versa. The answer is non-trivial, as some of the main components (namely drivers wages) may be considered differently depending, for example, if it is possible to hire drivers just for some hours of the day. If daily fixed costs prevail, off-peak frequencies are going to be higher in the two-periods scheme.

6.2.2 Crossed effects between periods.

We now study how ${f_N}^*$and ${f_P}^*$react to changes in the patronage of the other period. Let us first explain it intuitively. If Y_P increases, then both ${f_P}^*$ and K will increase. From the point of view of off-peak operations, this means that each bus on the street is more expensive than before, so optimal frequency will diminish, i.e., $\frac{\partial {f_N}^*}{\partial Y_p} < 0$. If the off-peak patronage increases, off-peak frequency will also increase. This means that the size of the buses become now more relevant, because more buses will be running the whole day. So capacities must go down, meaning the peak frequencies will increase, that is $\frac{\partial {f_P}^*}{\partial Y_N} > 0$.

To see the first phenomenon analytically, recall that when deriving VRC_2 with respect to f_N, we obtain equation (6.15), where the only element that depends on Y_P is the ratio Y_P/f_P (implicit in δ/f_P). Frequency ${f_P}^*$increases with Y_P at a decreasing rate (Jara-Diaz and Gschwender, 2003b) because the system responds not only through frequency but also increasing bus size; therefore the ratio Y_P/f_P increases when Y_P increases, and so does K. So ${f_N}^*$ will indeed decrease as Y_P increases.

In the case of ${f_P}^*$given by equation (6.12), both G_P and ${f_N}^*$ increase with Y_N, such that the optimal peak frequency increases. This happens because when off-peak flow increases, both off-peak frequency and cycle time increase and, therefore, off-peak fleet (and operating costs) increase; as vehicle size weights more, this effect that can be softened by reducing vehicle size. This, however, requires peak frequency to increase.

6.2.3 Summary of this case and comparison with previous approaches.

The preceding analytical findings and intuitive explanations are summarized in Table 6.1. The two-periods model presented and analyzed here can be used to investigate the conditions imposed by Jansson (1980, 1984), summarized in the first section. In his first model, Jansson assumes that each bus has the same cost, independently of its size. In our scheme, this is equivalent to put $c_{KC} = c_{KO} = 0$. From equations (6.4)-(6.8) it is direct to observe that, under these conditions: $\delta = 0$; A_P, G_P and G_N become equal to the value they would have in the single period case; and $A_N$only considers operating costs. From the analysis in sub-section 6.3.2 (summarized in Table 6.1, top three rows), the only element that still matters is that buses are already acquired when deciding off-peak frequencies; all the other effects depend on the role of bus size, which does not affect costs under Jansson's assumption. The result is quite direct, both analytically (from equations 6.12 and 6.15) and conceptually: for the peak period, all the conditions are exactly equal to the single period problem, such that optimal frequencies coincide; for the off-peak, as $A_N < A$, optimal frequencies are going to be higher than those obtained by solving this period in isolation. Therefore, even if size had no effect on buses costs, optimal frequencies are going to differ across periods, although off-peak frequency would be larger than in the single period analysis, getting closer to the peak period frequency (which seems to be what lies behind Jansson's intuition).

Fact	**Considerations**	**Effects**
Peak bus sizes affect off-peak period	1. Bus size has to be chosen considering off-peak operating cost, that increases with K $(G_P + \delta f_N > G)$. 2. Off-peak bus size must be sufficient to carry peak flows (Presence of δ / f_Pin the denominator of f_N).	1. Optimal K is lower than in the single peak-period case; peak frequency is pushed upwards. 2. Optimal K is larger than single off-peak period case; off-peak frequency is pushed downwards (buses might not be full).
Peak fleet is larger, so enough buses for off-peak are available.	There is no capital cost associated to off-peak $(A_N < A)$.	Off-peak frequency is pushed upwards
The size of the buses is the same in both periods	Increasing off-peak frequency cannot be fully compensated by a reduction in vehicle size in order to reduce operators' costs $(G_N < G)$.	Off-peak frequency is pushed downwards. Idle capacity appears.
Increasing off-peak passengers increases off-peak frequency.	Bus operating costs become more important; smaller buses are better.	Peak frequency increases. $\frac{\partial {f_P}^*}{\partial Y_N} > 0$
Increasing peak passengers increases the size of the buses.	Each bus running during the off-peak period becomes more costly.	Off-peak frequency decreases.$\frac{\partial {f_N}^*}{\partial Y_P} < 0$

Table 6.1. Qualitative impacts of the two periods relations.

It is interesting to analyze the expressions that represent ${f_P}^*$and ${f_N}^*$under Jansson's strong assumption, i.e. when $c_{KC} = c_{KO} = 0$. These expressions are

$${f_P}^* = \frac{Y_P}{T_P}\left[\frac{p_w/2 + Y_P p_v l_p t/L}{c_{BC}/E_P + C_{BO}}\right] \text{ and } {f_N}^* = \frac{Y_N}{T_N}\left[\frac{p_w/2 + Y_N p_v l_N t/L}{C_{BO}}\right] \qquad (6.16)$$

There are four period-specific elements that are worth looking at in this comparison: flow, trip length, time in motion and duration of the peak period. By definition the peak flow is larger than the off-peak flow and trip length is usually longer during the peak as well. Both elements contribute to make the peak frequency larger than the off-peak one. However, time in motion could be larger in the peak due to congestion, which contributes to reduce the difference. A short peak period also works in that direction. This means that Jansson's claim that frequencies are equal under his strong assumption on costs, might hold under very specific circumstances, although observed relative values of these parameters make us expect a systematic positive difference between ${f_P}^*$and ${f_N}^*$even if this assumption holds, something that can be explored numerically as well.

6.2.4. Numerical analysis

As explained in section 6.2.2, closed analytical solutions for the frequencies are impossible to obtain. In this Section we show the optimal values for frequencies, fleet and buses size obtained numerically using the parameters presented in Appendix A.

In Figure 6.1 we show the behavior of total fleet (yellow) and the size of buses (blue) as patronage increases, keeping constant the ratio between peak and off-peak flows (10/3). Just as in the single period case, and as expected, both fleet and bus size increase with patronage. Let us analyze now period-specific frequencies and usage.

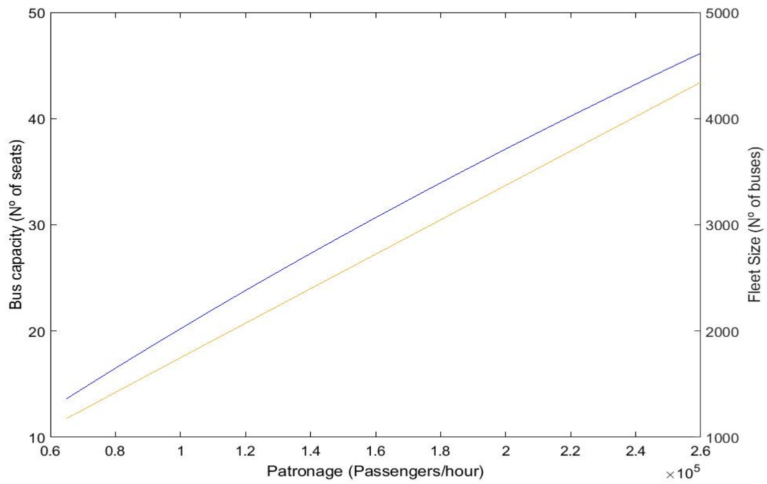


Figure 6.1. Optimal fleet size and buses capacity as total patronage grows proportionally.

As explained, relations between peak and off-peak frequencies are strongly determined by the relationship between the different components of the operators' costs. So we studied the resulting frequencies when varying the ratio between total capital costs and total operating costs (Figure 6.2), and the ratio between size-independent and size dependent bus costs (Figure 6.3). In both cases the operators' cost of a standard bus (100 seats running 18 hours on the streets) was kept constant. This is done in order to alter minimally the ratio between operators' and users' costs. In both figures, the scale varies from one half to double the original value of these ratios.

In these Figures we show the optimal frequencies considering two periods (solid lines) and those that would be obtained considering each period in isolation (dotted lines). Peak-period frequencies are drawn in red and off-peak in green. As expected, single peak-period frequencies are larger than those considering off-peak parameters in isolation, but the optimal peak frequency considering two periods are even larger (as predicted theoretically) because of the convenience of smaller buses due to the effect on vehicle size of the off-peak operating costs. The numerical novelty is that frequencies during the off-peak period are smaller than those obtained for a single period analysis. As a result, the difference between optimal frequencies considering two periods is larger than the difference between frequencies when periods are considered in isolation. As a consequence, the (single) optimal vehicle size is smaller than the single peak period case and larger than the single off-peak case.

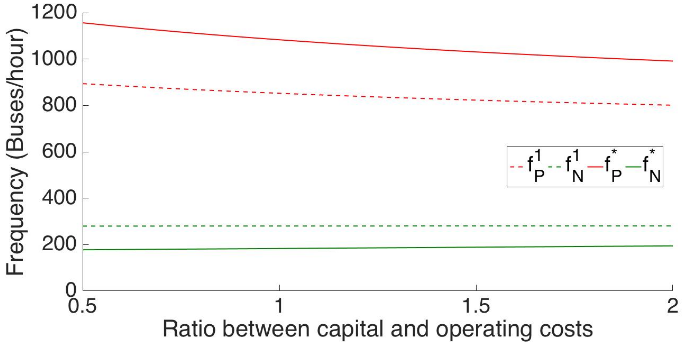


Figure 6.2. Optimal frequencies for varying total capital costs over total operating costs ratio.

Comparing our results with Jansson's (1980, 1984), he assumed that frequencies are equal, while we show that differences are even larger than when solving peak and off-peak separately, as predicted theoretically in the previous section. Following Figure 6.2, the difference between the optimal frequencies (two periods) diminishes as capital costs become more important, which fits intuition: once a bus is bought, it is better to use it.

Figure 6.3 confirms the same general conclusions. Besides, peak frequency drops significantly which means that bus size increases as costs related to size turn less relevant; this pushes optimal (two-periods) frequencies towards those from a single period analysis. Extending Figure 6.3 to the right (where size dependent costs vanish, as in Jansson's model) the difference between peak and off-peak frequencies diminishes but remains significantly positive.

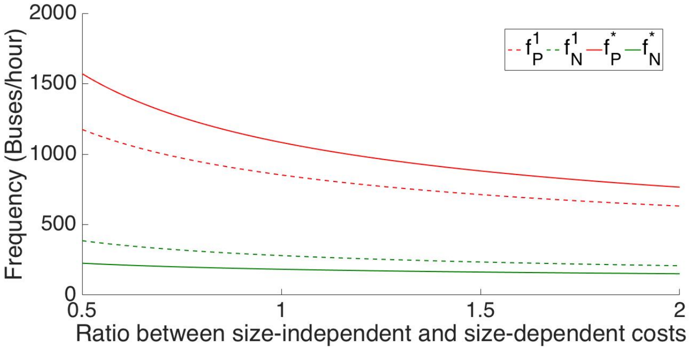


Figure 6.3. Optimal and single period frequencies for varying size-independent over size dependent bus costs ratio.

Next, in Figures 6.4 and 6.5 we analyze the impact of Y_P and Y_N on frequencies (right y-axis) and capacity of buses (left y-axis). The optimal capacity (number of seats of each bus) is presented in blue; the load of each bus in the off-peak period is also shown (diamonds line), in order to verify that it is always lower than the optimal (assumed to be commanded by the peak load).

Figure 6.4 shows responses to changes in peak-period passengers flow. As expected, both peak-frequency and bus size increase when Y_P does, because buses run full in this period. As predicted analytically, off-peak frequency decreases (although very slowly) due to the rise in bus size. Accordingly, off-peak bus load increases but capacity is always enough.

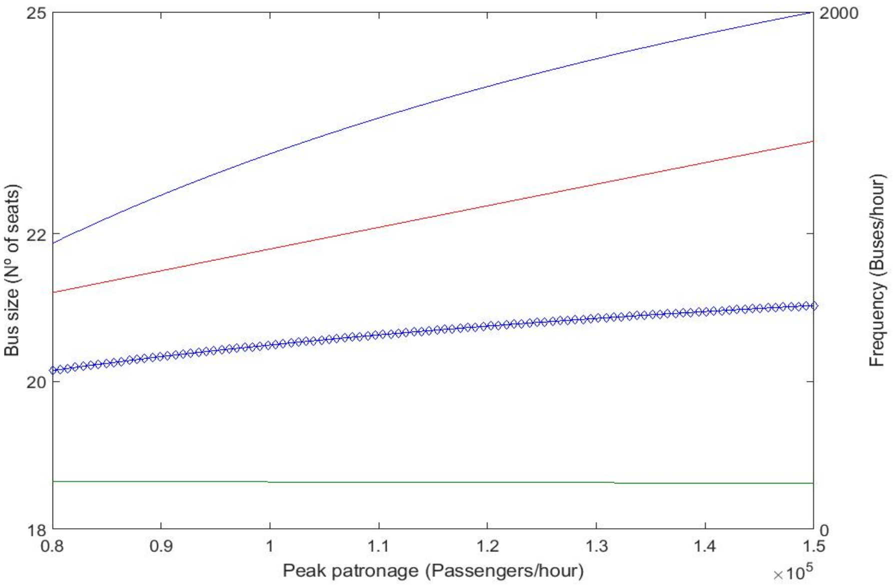


Figure 6.4. Optimal frequencies and bus size for varying peak passengers flow.

Figure 6.5 shows responses to changes in the off-peak period passengers flow. As predicted analytically, buses become smaller as $Y_N$increases, because each large bus becomes more costly as f_N grows in response to new passengers. These smaller sizes make peak frequencies increase as well. Note that off-peak load approaches capacity as patronage increases, so there will be a point where size begins to be determined by off-peak conditions, which is analyzed in section 6.3.

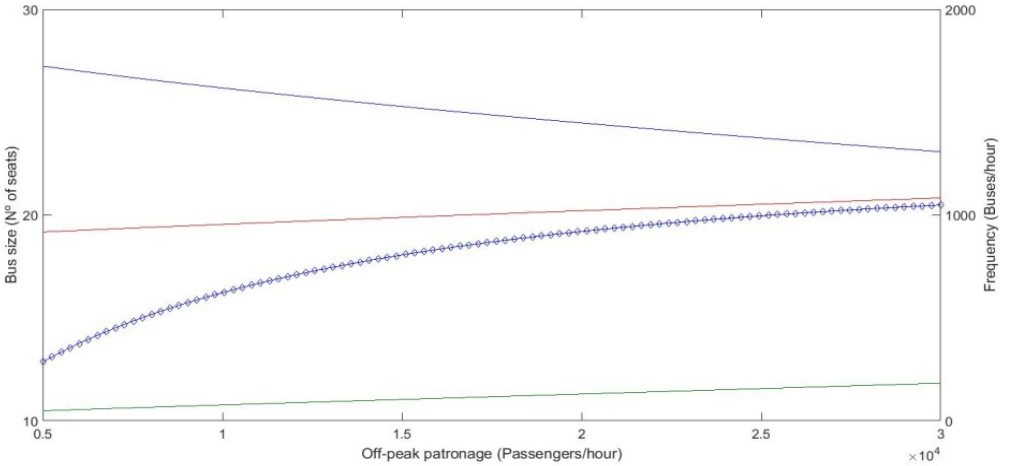


Figure 6.5. Optimal frequencies and bus size for varying off- peak passengers flow.

6.3 Buses full at the off-peak

Figure 6.5 revealed that buses could run full at the off peak; intuitively, this can happen because optimal frequency at the off-peak is lower than at the peak such that buses could run at capacity under some conditions. In this section, we will analyze the case in which the off-peak constraint (6.2b) is active (buses full at the off-peak), i.e.

$$K = \frac{Y_N l_N}{f_N L} \tag{6.17}$$

It should be noted that fleet size is still given by the peak, because both cycle time and frequency are larger than at the off-peak (larger flow, trip length and in-motion time). Then VRC_2 can be written as a function of frequencies using equation (6.17) and the expressions that link fleet size and cycle time with the corresponding frequencies (in chapter 2). This yields:

$$VRC_2 = (f_P T_p + tY_P)(c_{BC} + c_{KC}\frac{Y_N l_N}{f_N L}) + [(f_P T_P + t \quad Y_P)E_P + (f_N T_N + t \quad Y_N)E_N][c_{BO} + c_{KO}\frac{Y_N l_N}{f_N L} \quad] + \frac{p_w}{2}(\frac{Y_P E_P}{f_P} + \frac{Y_N E_N}{f_N}) + \frac{p_v}{L}(l_P Y_P E_P[t\frac{Y_p}{f_P} + T_P] + l_N Y_N E_N[t\frac{Y_N}{f_N} + T_N]) \tag{6.18}$$

Like in the previous section, this function can be re-written as:

$$VRC_2 \quad = A_P f_P + G_P/f_P + A_N f_N + G_N/f_N + \varepsilon f_P/f_N + W \tag{6.19}$$

where

$$A_P = T_P c_{BC} + T_P E_P c_{BO} \tag{6.20}$$
$$A_N = T_N E_N c_{BO} \tag{6.21}$$
$$G_P = \frac{p_w Y_P E_P}{2} + \frac{p_v\, l_p E_P t Y_P{}^2}{L} \tag{6.22}$$

$$G_N = tY_Pc_{KC}Y_N\frac{l_N}{L} + tY_PE_Pc_{KO}Y_N\frac{l_N}{L} + tY_N{}^2E_N \quad c_{KO}\frac{l_N}{L} + \frac{p_wY_NE_N}{2} + \frac{p_v\, l_NE_NtY_N{}^2}{L} \tag{6.23}$$

$$\varepsilon = T_Pc_{KC}\frac{Y_Nl_N}{L} + T_Pc_{KO}\frac{Y_Nl_N}{L}E_P \tag{6.24}$$

$$W = tY_Pc_{BC} + tY_PE_Pc_{BO} + T_NY_Nc_{KO}\frac{l_N}{L} + \frac{p_v}{L}(Y_PE_Pl_PT_P + Y_NE_Nl_NT_N) \tag{6.25}$$

What about constraint 6.2a? There are two possible cases: active or non-active, i.e. buses run either full or not at the peak. The following proposition shows that they indeed run full: the peak constraint is also active when the off-peak constraint is.

Proposition: Buses run full at the peak when they run full at the off-peak.

Proof: by contradiction. Let us assume that the constraint (6.2a) is not active. In that case, we just take the derivatives in equation (6.19) and make them both equal to zero, which yields:

$$f_N = \sqrt{\frac{G_N+\varepsilon f_P}{A_N}} \text{ and } f_P = \sqrt{\frac{G_P}{A_P+\varepsilon/f_N}} \tag{6.26}$$

These expressions can be compared with the frequencies obtained for each period in isolation, i.e. $f_{i1} = \sqrt{\frac{G_{i1}}{A_{i1}}}$, with $A_{i1} = T_i(c_{BC} + E_ic_{BO})$ and $G_{i1} = \frac{Y_i^2l_it}{L}(c_{KC} + E_ic_{KO} + E_ip_v) + \frac{p_wE_iY_i}{2}$.

It is quite direct to observe that:
$A_P = A_{P1}$
$A_N < A_{N1}$: no capital costs c_{BC} at the off-peak.
$G_P < G_{P1}$: costs associated with bus size, $c_{KC} + E_ic_{KO}$, do not appear in G_P (as bus size is given by the off-peak, changing frequency does not reduce the bus size).
$G_N > G_{N1}$: the first two terms in G_N, involving Y_P and costs associated with bus size, do not appear in G_{N1}(choosing large buses is costlier than in the isolated case, because the same buses will run at the peak; so it is better to have smaller buses).

These relationships show directly that
i) peak frequency f_P is lower than f_{P1} (isolated case);
ii) off-peak frequency $f_N$is larger than f_{N1} (isolated case).

Conclusion ii) implies that bus size is lower than in the isolated off-peak case, which in turn is lower than in the isolated peak case (because K increases with Y). Therefore, the assumption of constraint (6.2a) not active when (6.2b) is, leads to a contradiction: at the peak, frequency and bus size are simultaneously lower than in the isolated case, which cannot occur and proves the assumption wrong. **Q.E.D.**

The proposition implies that buses run full at the peak under every circumstance (it is only the off-peak that may present both cases). Therefore it always holds that

$$f_P = \frac{Y_Pl_P}{KL} \tag{6.27}$$

Combining this with equation (6.17) we obtain an expression for f_P as a function of f_N:

$$f_P = \frac{Y_P l_P}{Y_N l_N} f_N \qquad (6.28)$$

This shows that not only the peak frequency has to be larger than the off-peak one, but also that their ratio replicates the ratio between the corresponding passengers-kilometers traveled per hour. Thus, contrary to Jansson's intuition, it is never optimal that frequencies are equal across periods.

Using (6.28) the value of the resources consumed can be expressed as a function of f_N only:

$$VRC_2 = \left(A_P \frac{Y_P l_P}{Y_N l_N} + A_N\right) f_N + \left(G_N + G_P \frac{Y_N l_N}{Y_P l_P}\right) \frac{1}{f_N} + \varepsilon \frac{Y_P l_P}{Y_N l_N} + W \qquad (6.29)$$

Minimizing $VRC_2$and using equations (6.20) to (6.24) we get (6.30) to (6.32).

$$f_N^* = \sqrt{\frac{tY_P c_{KC} Y_N \frac{l_N}{L} + \frac{p_w Y_N (E_N + E_P l_N / l_P)}{2} + [E_P Y_P + E_N Y_N] t Y_N (l_N / L)(c_{KO} + p_v)}{T_N E_N c_{BO} + T_P c_{BC} Y_P l_P / (Y_N l_N) + T_P E_P c_{BO} Y_P l_P / (Y_N l_N)}}, \qquad (6.30)$$

$$f_P^* = Y_P \sqrt{\frac{tY_P c_{KC} \frac{l_P^2}{L} + \frac{p_w (E_N l_P^2 / l_N + l_P E_P)}{2} + [E_P Y_P + E_N Y_N] t (l_P^2 / L)(c_{KO} + p_v)}{T_N E_N c_{BO} Y_N l_N + T_P c_{BC} Y_P l_P + T_P E_P c_{BO} Y_P l_P}} \qquad (6.31)$$

$$K^* = \sqrt{\frac{Y_N T_N l_N E_N c_{BO} + T_P c_{BC} Y_P l_P + T_P E_P c_{BO} Y_P l_P}{tY_P c_{KC} L + L^2 \frac{p_w (E_N + E_P l_N / l_P)}{2} + L[E_P Y_P + E_N Y_N] t (c_{KO} + p_v)}} \qquad (6.32)$$

The comparison between the expressions for f_N^* and f_P^* with those given by the corresponding isolated ones can be done by looking at the (expanded) generic version of the optimal isolated frequencies for a period shown in Equation (6.33):

$$f_{i1} = \sqrt{\frac{Y_i (\frac{p_w E_i}{2} + tY_i (l_i / L)(p_v E_i + c_{KC} + c_{KO} E_i))}{T_i (c_{BC} + E_i c_{BO})}} \qquad (6.33)$$

The analytical comparisons between optimal frequencies in equations (6.30) and (6.31) with the corresponding expressions of equation (6.33), lead to inconclusive results because they depend on the values of all parameters. Instead, a numerical comparison is offered in Figures 6.6, varying Y_P from $60{,}000$ to $100{,}000$ keeping $Y_N = 50{,}000$ in Figure 6.6a, and varying Y_N from $50{,}000$ to $95{,}000$ keeping $Y_P = 100{,}000$ in Figure 6.6b.

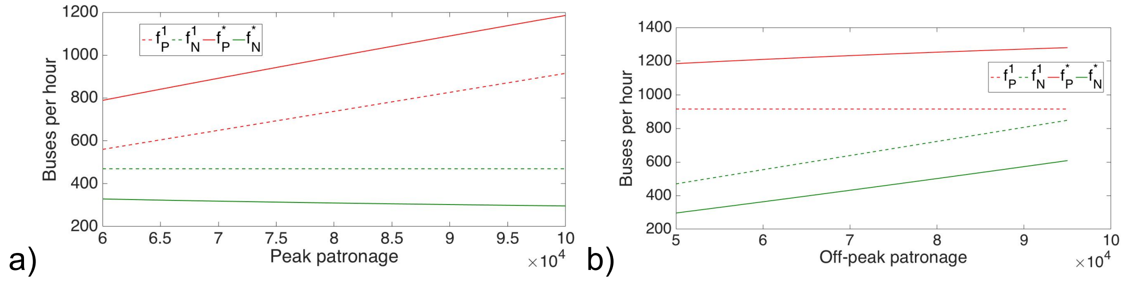


Figure 6.6. Comparison of peak and off-peak frequencies considering joint optimization against isolated optimization.

From Figures 6.6 we obtain that, numerically, $f_P^* > f_{P1}$ and $f_N^* < f_{N1}$, the same results obtained for the case where the off-peak constraint was not active (shown analytically for f_P^*) in section 6.2. Figures 6.6 also show that the relations analytically found for the crossed effects between frequencies and flows also hold in this case, i.e.

$$\frac{\partial f_P}{\partial Y_N} > 0, \quad \frac{\partial f_N}{\partial Y_P} < 0 \qquad (6.34)$$

As by definition $Y_P > Y_N$, these equations show that when flow i approaches flow j unilaterally, frequency j increases. The intuition behind this deals with the relative values of flows across periods: when they tend to converge the day gains homogeneity, approaching a single extended period, making capital costs weigh less.

Regarding bus size, although equation (6.32) shows that K could increase or decrease when either Y_P or Y_N increases, equations (6.34) imply that

$$\frac{\partial K}{\partial Y_P} > 0, \frac{\partial K}{\partial Y_N} < 0 \qquad (6.35)$$

In this case, when flow i approaches flow j unilaterally (i.e. when Y_N increases or Y_P decreases) bus size decreases. This could be interpreted in terms of the parameters involved in equation (6.32), particularly for the off-peak flow effect:

- When peak passengers increase, bus size has to increase just as in the single period case.
- When off-peak passengers increase, it is better to reduce bus size in order to reduce the time spent at bus stops.

If $Y = Y_P + Y_N$ remains fixed, the total effects of flow variations when they get closer can be examined. These effects can be seen in Figure 6.7, where $Y = 150{,}000$ and Y_N/Y varies from 5% to 45%. As we are increasing the off-peak flow while decreasing the peak flow, off-peak frequency increases because of a direct effect and the corresponding crossed effect (equation 6.34). In the case of peak frequency, it is pushed upwards because of the crossed effect and downwards because of the decrease of Y_P (direct effect); this last effect prevails resulting in a slight decrease. Bus size decreases because both effects represented in equations (6.35) work in the same direction.

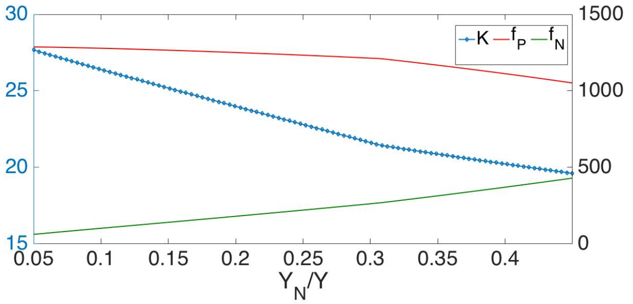


Figure 6.7. Bus size and peak and off-peak frequencies as a function of Y_N/Y .

Figure 6.7 reveals an evident change of slope in the three curves represented when the ratio Y_N/Y reaches 0.31, which is the value where constraint (6.2b) becomes active, i.e. when the off-peak buses start to be full. This motivates an examination of the conditions on the flows that make constraint (6.2b) active. Such conditions cannot be found explicitly, as they depend on the relationship between frequencies and flows when only constraint (6.2a) is active, and they are linked by an equation of degree 5. The conditions on the flows are represented numerically in Figure 6.8, where the yellow zone represents the combinations of flows that activate constraint (6.2b). For the buses in the off-peak to run fully loaded at the optimum, the flow in the peak has to be lower than two times the off-peak flow approximately.

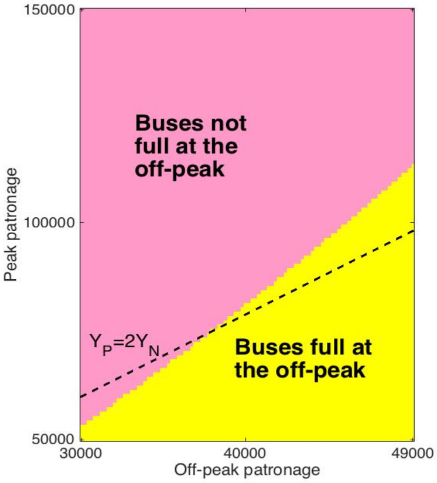


Figure 6.8. Flow conditions that make buses full at the off-peak.

6.4 Main conclusions

- Optimizing a single line with two periods yields equations that cannot be solved analytically. Nevertheless, some analytical conclusions are possible.
- At the peak, buses always run full; at the off-peak, they might run full or with idle capacity. Which is the case depends on the relative value of each period's flow.
- When buses do not run full at the off-peak, it is shown that peak frequency is larger than what it is obtained when optimizing this period in isolation. This happens because, as these vehicles will also run at the off-peak, decreasing its size is less costly.

- In this same context, a comparison between optimal off-peak frequency and the isolated optimal one is inconclusive, because there are effects pushing in opposite directions: having no capital costs pushes off-peak frequency upwards; but vehicles are larger than in the isolated case and their size do not diminish when frequency increases, which pushes frequency downwards. A numerical analysis reveals that off-peak frequency is lower than in the isolated case.
- In this same context, if peak flow increases, off-peak frequency gets lower because vehicles get larger. If off-peak flow increases, peak frequency increases as well because the size of the vehicles becomes more relevant.
- When buses run full at the off-peak, explicit expressions for the frequencies and bus size are obtained. All previous conclusions are shown (numerically) to remain valid.

Chapter 7. Systems that allow for two fleets.

When studying two periods, we have dealt so far with models that assume that all vehicles have the same size. Keeping the simple and intuitive model explained in chapter 6, but dropping this restriction, might yield better results. In this chapter two possible systems that considers two fleets are studied: one that adds extra buses at the peak and one that optimizes each period independently. A comparison between these systems and the one-fleet system is also provided.

7.1 Joint optimization allowing for two fleets

First, let us study the case in which some buses are added for the peak. This requires designing two fleets simultaneously: one that runs alone in the off-peak and another that runs as a complement during the peak only to satisfy the (larger) demand. Let us denote the buses that run all day as small (represented by a subindex S when needed) and those that are used only at the peak as large (represented by L); it is shown below that the buses that run exclusively during the peak period are indeed larger than the ones that run all day. The fact that large buses operate during one period only (and, therefore, depreciate at a slower rate) will be reflected by the operating costs coefficients that are multiplied by the duration of the period.

At the peak, both fleets are running. This poses a new and relevant operational difficulty. As large buses are carrying more passengers, they spend more time at bus stops, which implies that their cycle time is also larger. This could induce at least three undesirable effects:

- It is a well-known fact that different time at stops might induce *bunching* (see, for example, Newell, 1974). The explanation in our case is quite simple: if a large bus is followed by a small bus, the temporal headway between them is reduced after each stop, just because the large bus is spending more time there.
- Headways between consecutive buses are not constant in time, which means that the system becomes much more irregular. Headway variability is an index of unreliability and has been shown to be a very relevant (negative) quality factor for users by many researchers (like Friman *et al.*, 1998, Beirão *et al.*, 2007, or Redman *et al.*, 2013), such that regularity has merit in itself.
- A particular aspect of the headway variability condition is the chance that some buses might get full, forcing passengers to wait for the next bus. This is also likely to happen when headways change over time because longer headways induce a larger amount of passengers accumulated at the bus stop.

To avoid these nocuous effects, some complementary strategies need to be applied to induce equal cycle times across fleets in the peak. In this model we use a *holding* strategy (that has been well studied in other contexts, such as Osuna and Newell, 1972 or Daganzo, 2009), simply consisting in forcing small vehicles to have the same time at stops than large vehicles. This can be done by waiting with their doors closed for a fixed time (after passengers board) before leaving the stop[20]. This is illustrated in Figure 7.1.

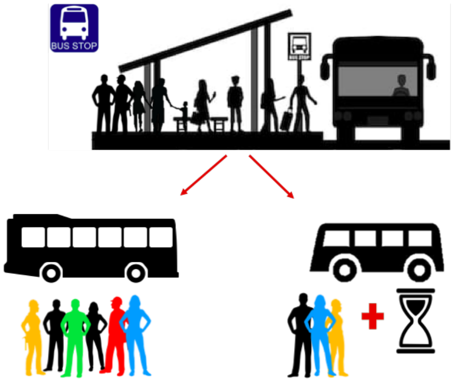


Figure 7.1. Graphical description of the holding strategy.

To describe the equations that govern the operation of this system, let us first introduce some notation. In this case, B_S and B_L denote the size of the fleets such that $B_S + B_L$ buses are running at the peak and B_S at the off-peak. Vehicles capacities are K_S and K_L respectively. We do not describe operational rules at the off-peak, as this period can be represented by the traditional one-line model provided capital costs are correctly adjusted to avoid double-counting. At the peak, headways h_S and h_L are going to be used instead of frequencies (which are not particularly informative in this case); h_i represents the time elapsed between the instant in which the previous vehicle closed its doors and the instant in which the current type i vehicle does. In other words, h_S and h_L represent the period during which passengers arrive at the bus stop to board the corresponding vehicle. Please note that, under this definition, the time spent by a small bus at the stop with its doors closed is not part of its headway in that stop; it corresponds to the headway of the next bus.

During the peak period a total of Y_P passengers arrive per unit time such that the number of passengers that ride a particular vehicle type during a cycle can be expressed as $Y_P h_i$; then total time spent at the stops is $tY_P h_i$. Equality of cycle times t_c of both types of vehicles imply:

$$T_P + tY_P h_S + H = t_c = T_P + tY_P h_L \tag{7.1}$$

Where H is the total holding time spent by each small vehicle in a cycle. We will show that the VRC can be expressed depending only on B_L and B_S, as vehicle sizes and headways can be written as functions of the fleet sizes. To do this, let us begin noting that vehicle sizes are given by:

$$K_L = Y_P h_L \frac{l_P}{L}, \quad K_S = Y_P h_S \frac{l_P}{L} \tag{7.2}$$

[20] A literal application of this rule could raise complaints from the users. Nevertheless, it can be changed to some equivalent rules that do not affect the equations here deduced, such as diminishing the speed of the small buses during the peak period.

To calculate the average waiting time, note that if the next bus is of type i, passengers will wait from zero to h_i, such that in average they wait for $h_i/2$. A proportion $\frac{K_iB_i}{K_LB_L+K_SB_S}$of the users take a bus of type i, which yields an average waiting time given by

$$\bar{t_w} = \frac{B_LK_L}{B_LK_L+B_SK_S}\frac{h_L}{2} + \frac{B_SK_S}{B_LK_L+B_SK_S}\frac{h_S}{2} \qquad (7.3)$$

Equations (7.2) and (7.3) yield

$$\bar{t_w} = \frac{1}{2}\frac{B_Lh_L^2+B_Sh_S^2}{B_Lh_L+B_Sh_S} \qquad (7.4)$$

Under this formulation waiting time for passengers boarding a small vehicle includes the time waiting for other passengers to board, as is usually done, but once the doors are closed time until departure is considered as in-vehicle time. Following Jansson (1980) and Jara-Díaz and Gschwender (2003b), time waiting for passengers in other stops is always in-vehicle time for those already aboard the bus. Then in-vehicle time can be expressed as:

$$\bar{t_v} = \frac{l_P}{L}t_c = \frac{l_P}{L}(T_P + tY_Ph_L) \qquad (7.5)$$

So far, we have expressed the components of the users' costs as functions of both fleet sizes and headways. Operators' costs - that depend on fleet sizes and vehicle sizes - can also be expressed as functions of B_i and h_i using equation (7.2). Two additional equations will allow us to have fleet sizes as the only variables. The first one is a different way of calculating the cycle time:

$$t_c = B_Lh_L + B_Sh_S \qquad (7.6)$$

To understand where does this come from, imagine a user observing which bus to take during a lapse of time that lasts t_c. This user will observe each of the B_L large buses as the next one coming during a lapse h_L (and equivalently with each small bus). Equation (7.6) can also be written as $t_c = (B_L + B_S)\bar{h}$, i.e., cycle time equals total fleet times average headway.

The second equation comes from the fact that small buses have to fulfill off-peak conditions. The relationship between fleet and vehicle capacity obtained for the single period model explained in chapter 2 does apply in this case; therefore

$$K_S = \frac{l_NY_NT_N}{L(B_S-tY_N)} \qquad (7.7)$$

Equations (7.2) and (7.7) reflect the fact that all buses are full in both periods. Combining those equations yields

$$Y_Pl_Ph_S\frac{B_S-tY_N}{T_N} = Y_Nl_N \qquad (7.8)$$

Equations (7.1)-(7.8) are sufficient to make the VRC a function of fleet sizes only, i.e. $VRC(B_L, B_S)$. In equations (7.9)-(7.11) we show the value of the resources consumed during each period, VRC_N and VRC_P, and the capital costs of small buses that are assigned to the whole day, VRC_D.

$$VRC_N = B_S E_N(c_{BO} + \frac{l_N Y_N t}{L(B_S - tY_N)} c_{KO}) + Y_N p_w \frac{T_N}{2(B_S - tY_N)} + Y_N p_v \frac{l_N}{L} \frac{B_S T_N}{B_S - tY_N} \tag{7.9}$$

$$VRC_P = B_L(c_{BC} + E_P c_{BO} + Y_P \frac{l_P}{L} \frac{B_S Y_N l_N T_N - T_P Y_P l_P (B_S - tY_N)}{Y_P l_P (B_S - tY_N)(tY_p - B_L)} (c_{KC} + E_P c_{KO})) + B_S(E_P c_{BO} +$$
$$\frac{l_N Y_N t}{L(B_S - tY_N)} E_P c_{KO}) + Y_P \frac{p_w}{2} \frac{B_L\left(\frac{B_S Y_N l_N T_N - T_P Y_P l_P (B_S - tY_N)}{Y_P l_P (B_S - tY_N)(tY_p - B_L)}\right)^2 + B_S\left(\frac{l_N Y_N T_N}{l_P Y_P (B_S - tY_N)}\right)^2}{B_L \frac{B_S Y_N l_N T_N - T_P Y_P l_P (B_S - tY_N)}{Y_P l_P (B_S - tY_N)(tY_p - B_L)} + B_S\left(\frac{l_N Y_N T_N}{l_P Y_P (B_S - tY_N)}\right)} +$$
$$Y_P p_v \frac{l_P}{L} B_L \frac{B_S Y_N l_N T_N - T_P Y_P l_P (B_S - tY_N)}{Y_P l_P (B_S - tY_N)(tY_p - B_L)} + B_S(\frac{l_N Y_N T_N}{l_P Y_P (B_S - tY_N)}) \tag{7.10}$$

$$VRC_D = B_S(c_{BC} + \frac{l_N Y_N t}{L(B_S - tY_N)} c_{KC}) \tag{7.11}$$

First order conditions for B_S and B_L yield a system of equations of degree at least 5, such that no analytical solutions are possible; a numerical approach is used, using the value of the parameters shown in the Appendix A. Figures 7.2 represents the optimal values for these variables as a function of the off-peak (7.2a) and the peak (7.2b) passengers flows; the Figure also shows the values of the corresponding vehicle sizes (7.2c and 7.2d), showing interesting results including some intuitive ones, e.g. buses added at the peak (the so-called "large" buses) resulted actually larger than those that run all day (the "small" buses).

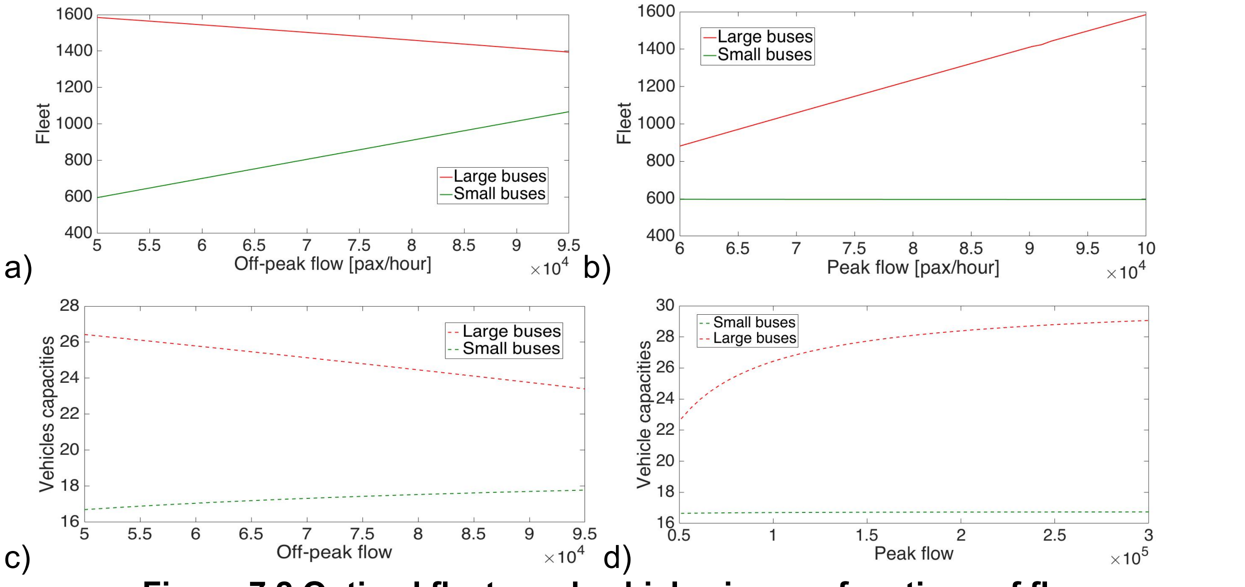


Figure 7.2.Optimal fleets and vehicle sizes as functions of flows.

As expected, the fleet of small buses and their size increase with the off-peak flow, and the fleet of large buses and their size increase with the peak flow, i.e.

$$\frac{\partial B_S}{\partial Y_N} > 0, \frac{\partial K_S}{\partial Y_N} > 0, \frac{\partial B_L}{\partial Y_P} > 0, \frac{\partial K_L}{\partial Y_P} > 0 \tag{7.12}$$

Additionally, as the off-peak flow increases not only the fleet of small buses increase but also their size. As these vehicles also run at the peak, less large buses of a smaller size are needed, i.e.

$$\frac{\partial B_L}{\partial Y_N} < 0, \frac{\partial K_L}{\partial Y_N} < 0 \tag{7.13}$$

Finally, Figures 7.2b and 7.2d also shows that the fleet of small buses and their size are practically insensitive to the peak flow, although actually the fleet decreases and vehicles size increases by a very small amount, i.e.

$$\frac{\partial B_S}{\partial Y_P} \approx 0, \text{ with } \frac{\partial B_S}{\partial Y_P} < 0; \frac{\partial K_S}{\partial Y_P} \approx 0, \text{ with } \frac{\partial K_S}{\partial Y_P} > 0 \tag{7.14}$$

The results reflected by equations (7.12) and (7.14) indicate that an increase in the peak flow is covered by having more and bigger large buses, and - in a less relevant way - by having less and larger small buses, which can be intuitively explained as follows. First, the combination of these changes contributes to increasing the average size of the vehicles: large buses become larger, small buses become larger and the percentage of large buses increases. Second, the model adjusts mostly the large buses in order to avoid an unnecessary impact on the off-peak period, whose conditions remain unchanged. Nevertheless, as the difference between the vehicle sizes of the two types of buses increases, so does the holding time for small buses, which is inefficient; this is why the number of small buses also increases slightly.

7.2. One or two fleets? Comparison of the models

In this section we compare the single-fleet design (buses of one size running all periods) against the two-fleets designs, including not only the one developed in section 7.1 but also the basic case of one independent fleet per period, whose design corresponds simply to solving twice the classic single period case.

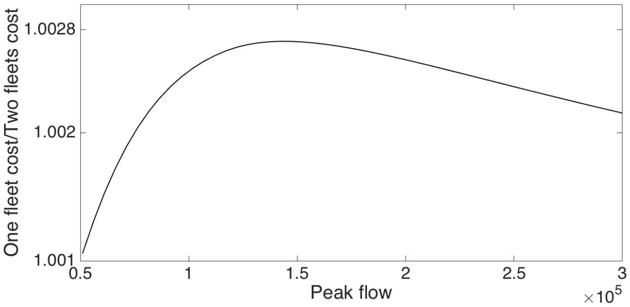


Figure 7.3. Ratio between total costs of the two systems.

Let us begin comparing the two-fleet system (chapter 6) and the one analyzed in section 7.2, recalling that in the single-fleet case buses may run full or not at the off-peak depending on the relative values of the flows. Does the two-fleets system imply an improvement over the single fleet case? Numerical results yield that total costs for the one-fleet system are always larger than the two-fleets case, but by less than 0.3% (with

$Y_N = 50{,}000$ and $Y_P \in [51{,}000;\ 300{,}000]$), as shown in Figure 7.3. There, the cost difference diminishes towards the extremes: at the beginning of the curve, both systems yield almost equal costs because peak and off-peak flows are almost equal; when the peak flow is much larger than the off-peak, it is the peak that dominates the cost calculations in both cases.

Figure 7.4 shows that although operators and users costs are also almost equal, the difference between systems is mostly explained by operators' costs. Actually, the two-fleets system presents systematically a lower fleet of buses that are larger in average, which diminishes operators' costs. That is to say, the flexibility of the two-fleets system allows for having a fleet that is more efficient for the operators. On the other hand, off-peak users are in a better situation with the two-fleets system, because it allows for smaller vehicles that increase frequency and decrease in-vehicle time; the opposite happens with peak users, because the holding strategy increases their in-vehicle time.

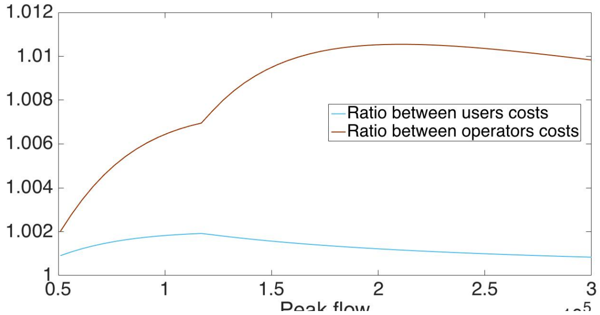


Figure 7.4. Ratios between users' and operators' costs for the two systems.

As discussed, the two-fleets system requires the introduction of some operating rules that mean an additional planning effort that has not been accounted for, e.g. the holding strategy for small buses during the peak (or vehicle speed reduction), which in turn induces some possibly unwanted effects on the users (like waiting inside full buses). Then the small cost savings for both users and operators do not seem to justify this system when compared to the one-fleet case.

What about the two-fleet system optimizing each period independently? The system that optimizes each period in isolation does not take advantage of the obvious scope economies associated to running one bus during two periods, but fleets can be adapted to demand exactly. In Figure 7.5 we compare the total costs of the three designs using the two-fleets system with holding as the reference to construct the ratios (same parameters as in Figure 7.3): "independent periods" system (red lines) and the one-fleet system (blue lines). The independent periods design yields the worst results, although differences are quite small for the three systems (less than 1%). As the time spent by passengers in the buses in-motion (i.e. not in the stops) is independent of the design, it can be subtracted to highlight the differences, which yields that optimizing periods independently can increase costs by about 5%. Numerical analysis shows that this difference is explained mostly by operators' costs (losing economies of scope); actually, users costs are slightly lower in the system that optimizes each period in isolation. Note that when the peak flow is very large, all the systems yield similar designs (and cost results).

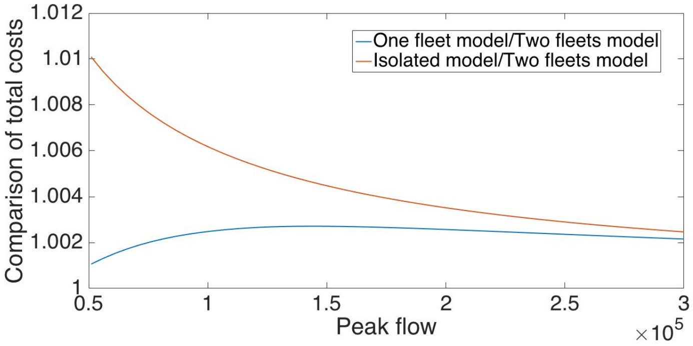


Figure 7.5: Comparison of the three systems.

7.3 Main conclusions

- It is possible to design a model with two fleets, with only “small” buses running at the off-peak, complemented by “large” buses at the peak.
- The equations that govern this two-fleets system have been obtained. A numerical analysis shows that the fleet of small buses is almost insensitive to changes in the peak period; on the other hand, if the off-peak flow increases, as we will have more and larger small buses, we will need less and smaller large buses.
- This two-fleets system is always better than the single-fleet one, but the difference is extremely low, such that the difficulties induced by having two fleets do not seem to be justified.
- Using two fleets that operate independently at each period is worse than the other two systems.

Chapter 8. Second-best strategies.

In this chapter, we investigate some second-best strategies for the public transport design for a single line considering two periods. We will optimize for one of the periods, and then consider the resultant size of the vehicles as a fixed parameter when optimizing the frequency for the other period. Asking which of the two emerging strategies is better was first done by Walker (2012) in a general scheme; it is worth saying that Villalobos (2018) faced this same question over a simplified version of Santiago, Chile, that included lines structure design, finding that it would be better to design for the off-peak (and adapt for the peak) in that particular case.

8.1 Optimizing one line with fixed capacities

Let us begin posing a general optimization problem that will appear throughout this chapter. As vehicles' capacity is going to be fixed form some stages of the analysis, it is useful to study the problem that optimizes frequency for a predefined capacity K:

$$(P_K)\colon \min_{\mathrm{f}\geq\frac{\mathrm{Yl}}{LK}} VRC(f) = B(f)(\,c_0 + c_1K) + \frac{p_wY}{2f} + \frac{p_vYl}{L}t_c(f) \tag{8.1}$$

Using the same analysis developed in chapter 2, Problem (P_K) may be easily written explicitly as a function of the frequency:

$$(P_K)\colon \min_{\mathrm{f}\geq\frac{\mathrm{Yl}}{LK}} (fT + tY)(c_0 + c_1K) + \frac{p_wY}{2f} + \frac{p_vYl}{L}(T + \frac{tY}{f}) \tag{8.2}$$

Let us call (P) the traditional single-line problem with vehicles capacity being optimized too. Define (f^*, K^*) as the optimal solutions of (P), and f_K as the optimal solution of (P_K). How to find f_K depends on if it lies in the interior of the feasible zone or in its border, i.e., if $f_K > \frac{Yl}{LK}$ (buses not full) or if they are equal (and buses run full); which is the correct option depends on K. If buses do not run full, then f_K can be found just by making the derivative of VRC equal to zero in (P_K), yielding

$$f_K = \sqrt{\frac{\frac{Yp_w}{2}+\frac{Y^2p_vtl}{L}}{T(c_0+c_1K)}} \tag{8.3}$$

Defining $r(K) = \sqrt{\frac{\frac{Yp_w}{2}+\frac{Y^2p_vtl}{L}}{T(c_0+c_1K)}}$, we conclude that buses do not run full iff $r(K)\cdot K > \frac{Yl}{L}$. Otherwise, if $r(K)\cdot K \leq \frac{Yl}{L}$, optimal frequency is given by

$$f_K = \frac{Yl}{LK} \tag{8.4}$$

It is easy to see that $r(K) \cdot K$ is an increasing function, such that there exists some K_0 that acts as a threshold: $f_K = r(K)$ iff $K \geq K_0$ (that is to say, buses do not run full when their size is larger than K_0). Note that if $K = K^*$, then $f_K = f^*$, which make buses to run full, implying that

$$K^* \leq K_0 \tag{8.5}$$

The inequality in (8.5) is strict, which follows from the following straight-forward calculation

$$[r(K^*) \cdot \ K^*]^2 = \frac{l^2Y^2}{L^2} \frac{\frac{p_W}{2}+\frac{Yp_vtl}{L}}{\frac{p_W}{2}+\frac{Ytl(p_v+c_K)}{L}} \frac{Tc_BY}{T(c_B+c_KK^*)Y} < \frac{l^2Y^2}{L^2} \tag{8.6}$$

Synthetizing, it has been shown that when we optimize only the frequency in the single-line design, considering the capacity as an exogenous value, then the solution for the frequency can be given by expressions (8.3) or by (8.4). Which is the correct one depends on whether the fixed value of the vehicle size is larger (8.3) or smaller (8.4) than some K_0, which happens to be strictly larger than K^*.

8.2 Optimization of one period and adaption of the other

First, let us study the case in which the peak period is optimized as in a classical single-line model, regardless of the characteristics of the off-peak period. After this, a sub-fleet of an optimal size is used at the off-peak. The peak optimization is not going to be explained, as it is the same procedure explained in chapter 2 for a single period.

To optimize the off-peak frequency, it is necessary to decompose operators' costs as in chapters 6 and 7 (capital and operating costs). The value of the resources consumed at the off-peak is expressed in (8.7), which has no capital costs because the size of the fleet is determined by the peak characteristics.

$$VRC_N^P = T_Nf_NE_N(c_{BO} + c_{KO}K_P) + tY_NE_N(c_{BO} + c_{KO}K_P) +$$

$$\frac{p_WY_NE_N}{2f_N} + \frac{p_vl_NtY_N^2E_N}{Lf_N} + \frac{p_vl_NY_NE_NT_N}{L} \tag{8.7}$$

Now, we need to determine if f_N is determined by equation (8.3) or (8.4), i.e., if buses run full or not. K_P is larger than the size obtained if the off-peak is optimized in isolation; further, here we are considering an off-peak without capital costs, which yields even smaller vehicles. Nevertheless, it is not evident if K_P is large enough to induce not-full vehicles at the off-peak (i.e., its comparison with the K_0 defined above is inconclusive). We assume that buses do not run full, which we verify numerically afterwards. Off-peak frequency is then:

$$f_N^P \ = \sqrt{\frac{\frac{Y_NE_Np_W}{2}+\frac{Y_N^2E_Np_vtl_N}{L}}{T_N(c_{BO}E_N+c_{KO}E_NK_P)}} \tag{8.8}$$

With these expressions, total daily costs can be calculated. They should be compared with total daily costs under the alternative system, in which the off-peak period is optimized in isolation, and its fleet is expanded (but preserving vehicles' size) to fulfill peak requirements. Please notice that this means that fleet size is still given by the peak, i.e., capital costs should be assigned there. Hence, the size of the buses is smaller than in the isolated off-peak case (due to no capital costs), which presents in turn smaller buses than the isolated peak: this implies that buses will run full, as shown in section 8.1, and peak frequency can be expressed as:

$$f_P^N = \frac{Y_P l_P}{L K_N} \tag{8.9}$$

8.3 Comparison between the two alternatives

Which of these two systems results more effective depends on all the parameters. Nevertheless, there are two parameters that represent the relevance of each period: on the one hand, optimizing for the peak is based intuitively on the fact that this is the most loaded period, which makes Y_P a key variable. On the other hand, the off-peak period lasts longer, such that if it is served inefficiently, it might have a strong impact on the total costs. This makes E_N another key variable. Let us show now that extreme cases for these variables yield to these intuitive results.

Proposition 8.1: if $Y_P \to \infty$, it is better to optimize for the peak; if $E_N \to \infty$, it is better to optimize for the off-peak.

Proof: First, recall that by equations (2.3), K_P converges to an upper bound when $Y_P \to \infty$. Let us analyze now the impact of Y_P over the off-peak costs. If the bus size is given by the off-peak, Y_P does not affect at all. In the other case, off-peak costs are affected through K_P, which is a convergent function of Y_P. Hence, off-peak costs are constant in one system and convergent to a finite quantity in the other. On the other hand, peak costs are asymptotically linear (and divergent to infinity) in both cases; but they are obviously lower when the peak period is optimized, precisely because we are optimizing it.

The proof for E_N is analogous. Bus size does converges to an upper bound when $E_N \to \infty$, such that peak costs are convergent as well. This means that the relevant comparison is between the linear and divergent off-peak costs, which are obviously lower when the off-peak is being optimized. **Q.E.D.**

One could look at the total number of passengers at each period $E_i Y_i$ to determine which period should be considered for the optimization. Although this rule is not exact, numerical results shows that this a remarkably good approximation rule, as shown in Figure 8.1.

The difference in costs between these two strategies is never larger than 1.3%. Nevertheless, if we exclude the costs associated to passengers over the vehicle in-motion (which are fixed and do not depend on the design), this difference reaches 6% in the extreme cases. When compared with the one-fleet system that optimizes both periods jointly, up to 2% might be saved in the first-best.

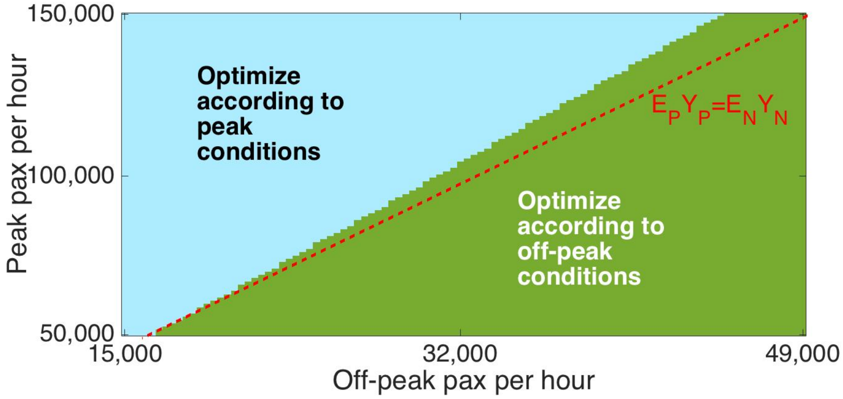


Figure 8.1. Which second-best strategy is better?

8.4 Main conclusions

- Optimizing a single line in one period with fixed capacities yields two possible solutions depending on the number of passengers: buses might run full or with idle capacity.
- When the off-peak period is optimized in isolation, and the peak period is optimized using the resulting bus capacities, vehicles do run full. In the inverse situation (optimizing the peak and adapting the off-peak), a numerical analysis shows that they do not run full.
- If the number of passengers at the peak is large enough, it is better to optimize the peak and adapt the off-peak than to proceed inversely.
- If the off-peak period is long enough, it is better to optimize the off-peak and adapt the peak than to proceed inversely.
- In general, an approximated rule is to optimize the period that presents a larger total number of passengers.

Chapter 9. Synthesis and conclusions.

In this thesis we have studied several aspects regarding public transport design and its economic implications. Mohring (1972), Jansson (1980, 1984) and Jara-Díaz and Gschwender (2003b, 2009) have provided clear and explicit solutions for the optimal design of a public transport system in only one temporal period and over a single line. However, when more daily periods are considered, or when different lines are optimized jointly over a network, the problem becomes more complex.

The emergence of a network induces new challenges. Design decisions are not only how many buses and of which size, but also following which routes: *lines structure* is now an additional variable to be optimized. This variable is discrete, and optimizing it by itself is already an NP-Hard problem for many. Further, users' behavior becomes less predictable, as they may have many possible routes to go from their origins to their destinations. These new difficulties make this problem a complex one even when trying to pose it, as several assumptions need to be considered; solving it exactly is beyond current possibilities.

This is why several strategies have been developed to face this design problem. In this thesis we studied in deep one of these strategies (heuristics), we analyzed some of the economic topics of this spatial design, and we zoomed over one partial aspect: lines density in space.

To study the performance of the heuristics, we selected four of them and we applied them over the scheme proposed by Fielbaum *et al* (2016) and following its methodology. The frequencies of the lines structures emerging for each of them were optimized by minimizing the value of the resources consumed, which considers operators' costs (including total number of vehicles and total number of seats) and users' costs (including waiting and in-vehicle times, and total number of transfers). This procedure allows us to compare these heuristics-lines structures with the basic structures defined by Fielbaum *et al* (2016): hub and spoke, feeder-trunk, no transfers and no stops.

By doing so, we found that heuristics create mostly direct structures, but with some flexibility (through some parameter σ) that might consider routes that do not follow shortest paths in order to collect more passengers. These structures were better than the basic structures for most combinations of the parameters that define the OD-matrix. Two heuristics in particular (the ones proposed by Dubois *et al*, 1979, and by Ceder and Wilson, 1986) generated the most competitive structures. These heuristics were precisely the most flexible ones through σ. Nevertheless, none of these heuristics was very responsive to changes in the OD-matrix. Actually, the basic hub and spoke structure and exclusive structure dominate when the total flow of passengers is too low or too large, respectively.

Analyzing scale economies and their relationship with lines structures enlightens the effects of this design variable. As the number of passenger increases (preserving the internal distribution of the trips), there are some discrete levels of patronage where the optimal lines structure changes. If we study a segment of total flow where lines structure is constant, the scale economies analysis over one line remains valid, i.e., there are scale economies but they become less relevant for a larger number of passengers. This is true because each line preserves the scale effects existing in a single-line system: the Mohring effect for users, and the chance of using larger vehicles for operators, which are more relevant than the diseconomies of scales induced by larger times at bus stops.

In the exact points where line structure changes, there is a discrete source of scale economies (i.e., the degree of scale economies DSE jumps discretely). We first show this analytically: average costs of the "old" and the "new" lines structures are equal, but the marginal cost of the new one is lower. We then studied these changes in detail, analyzing which are the positive externalities induced by the increase in the number of passengers and that show up when lines structures changes.

For this, we defined a three-dimensional concept: the "directness", encompassing the number of transfers per trip, the average length of the trips (compared with the case in which every passenger travels across shortest paths) and the number of stops per trip. If these indices decrease, we say that directness increases. We studied the evolution of these indices in a quite simple and ad-hoc network, and then in the model proposed by Fielbaum *et al* (2016). In both cases we found that directness increases[21] with patronage. The explanation is quite intuitive: when these indices are low, passengers are benefited by lower in-vehicle time and less transfers, which are two of the three components of the users' cost. But lines become useful for less OD pairs, such that a lower percentage of the users will use each line, which reduces frequencies, increasing waiting times (the other component of users' costs). This increase in waiting times becomes less relevant when the number of passengers is high, inducing changes in lines structures.

The numerical analysis of users' costs in these models verified that waiting time increased after each change in lines structures. Regarding operators, numerical analysis showed that as lines become more OD-specific, their idle capacity is reduced, i.e., there is another source for scale economies regarding the total number of seats.

Fares and subsidies are coherent with these results: subsidies are needed, but subsidy per passenger decreases continuously until lines structure changes inducing a discrete increase. The opposite happens with the fares. Total subsidy increases but converges when the number of passengers goes to infinity.

How dense in space should a transit network be is not a novel topic. We deepened its understanding in two directions: first, we improved the model studied by Chang and Schonfeld (1991) that optimize frequencies and density in a set of parallel lines that includes access time in users' costs, by dropping some unnecessary simplifications: operators costs are now dependent on bus size, and cycle time depends on the number of passengers through time needed to board and alight the vehicles. The resulting model cannot be solved analytically, but we proved that each of the parallel lines preserves the relationship between number of passengers and frequencies found in the

[21] In some of the lines structures changes, there is a trade-off between these indices, such that one might increase locally while the others decrease. Nevertheless, in the global picture all the three indices decrease clearly.

single-line model. This result helped us showing that the number of passengers per line increases with flow, and to show that the scale analysis for the single-line model is valid in this context, but adding here a new source of scale economies: the diminishing in access times. We proved that frequency and density are adjusted to make access costs equal to waiting costs.

Second, we extended the model proposed by Fielbaum *et al* (2016), allowing each arc to represent many arcs. This was done by splitting each arc into a number (to be optimized) of parallel streets per unit area. We showed that when this new decision variable is added, total costs are reduced. Directness still increases with patronage, but at a slower rate, because each change in lines structures increases not only waiting times but also access times. Access and waiting costs are also equal under this scheme.

On the other hand, extending a single-period analysis to a model that incorporates a peak and an off-peak is less complex in its conception than the evolution of routes structures. There are no discrete variables neither user's choices. However, solving the optimization problem that emerges is still an analytical challenge.

We first considered the model in which all buses have the same size. In the single-period model, the size of the buses is adjusted to fit exactly the constant load that they carry. When there are two periods, the loads at each period might be different. To model this situation, two capacity constraints need to be considered, namely that buses are large enough to carry each period's load. As having larger buses increases the cost of the system, at least one of them is going to be active. The fleet is always given by peak conditions, because this period's flow is larger and trips are longer due to congestion.

We divided the analysis by cases. We first studied what happens when peak constraint is active but off-peak is not, i.e., when buses run full only at the peak. In this case, the optimization problem can be expressed as a function of two variables: peak and off-peak frequencies. The first order conditions, however, lead to equations of degree five, which cannot be solved analytically. This means that explicit expressions for the solutions are not achievable; nevertheless, managing the equations yield some interesting analytical conclusions:

- Peak frequency is always larger than what is obtained designing for the peak in isolation. This happens because as buses are also running at the off-peak, having large buses is too costly (for example, due to the gasoline expenses). Reducing their size requires increasing their frequency in order to be able to carry every passenger.
- An analogous comparison for the off-peak is inconclusive. On the one hand, capital costs must not be considered because the size of the fleet is given by peak conditions (which pushes off-peak frequency upwards). But on the other hand, buses are larger than in the isolated case, and increasing the frequency does not imply savings by means of reducing their size anymore (both effects push off-peak frequency downwards). Numerical analysis showed that these last effects prevail, i.e., off-peak frequency happens to be lower than in the isolated case.
- The effects of changes in one period's number of passengers over the other period's frequency were also deduced. If peak's flow increases, buses get larger

inducing a reduction in the off-peak frequency. If the off-peak flow increases, the off-peak frequency increases in response, making costs associated to bus size weigh more; bus size then decreases, which forces peak frequency to increase.

We then studied what happens if the off-peak constraint is active, i.e., if buses run full at the off-peak. It was proved that if the peak constraint was not active as well, then at the peak period buses would be smaller and would have a lower frequency than in the isolated peak case, which is a contradiction because peak users would not fit into the system. The conclusion is that buses also run full at the peak in this case (i.e., they always run full at the peak). When both constraints are active, it is possible to express the optimization problem depending on only one variable, such that explicit expressions for the solutions can be found. These solutions show that all relationships found in the previous case depend on the values of the different cost-related parameters. Nevertheless, the numerical analysis reveals that all these conclusions are still valid in this case.

We also studied numerically in which cases the buses run full at both periods and when only at the peak. It was found that this depends on both flows, and that there is a lineal boundary; a mnemonic (but not exact) rule is that buses run full at the off-peak if and only if off-peak's hourly flow is larger than half peak's hourly flow.

Dropping the assumption that all buses need to be of the same size introduces a new difficulty: if buses of different size run at the same time, the lapse of time spent at the stops is going to be different (because larger buses will have more passengers boarding and alighting), such that headways will not be constant anymore. This can be solved in two different ways: having separated fleets for each period, which can be analyzed as the sum of two independent single-period cases and does not take advantage of the scope economies, or by means of a *holding* strategy, meaning that smaller buses are forced to wait unmoved at the stops after passengers finish boarding and alighting, to reproduce exactly the time spent by the larger ones. This second system was modeled (with the holding strategy used at the peak), and the equations that govern it were obtained. No explicit solutions were possible, but we were able to compare this system with the one-fleet system and with the system obtained by operating each period independently.

We found that the two-fleets system with holding was always the best one. However, the difference with the one-fleet system was extremely low, such that it does not seem to justify the more intricate rules of operation. Operating each period independently is always the worst option.

Although we were able to understand which were the impacts of considering more than one period when designing public transport, this was possible only because we were considering a single line. When a network is considered, some second-best strategies might be necessary in order to find solutions; a very usual one is to optimize considering only peak conditions, and then make some adaptions to serve the off-peak with same fleets and lines. The following question arises: is this better to proceed in the opposite direction, i.e., to optimize according the off-peak conditions and then adapt for the peak?

We studied this question in the very simple context of a single line, in order to get some insights regarding under which conditions each of these second-best strategies is better. To do so, we compared a single-line system in which the peak period is optimized in isolation, and the off-peak period is served with an optimal subset of the fleet; versus the same single-line scheme in which the off-peak period is optimized in isolation and the fleet is then expanded (preserving the size of the vehicles) optimally to carry peak passengers.

In this comparison, we showed that if the peak flow is very large, then the strategy that optimizes for the peak conditions is better; the opposite happens if the off-peak period lasts for too long. A numerical comparison reveals that a good approximation is that the strategy that optimizes the period with the maximum total patronage (i.e. flow per time unit multiplied by length of the period) should be used.

The complexity and relevance of transit design suggests many lines for future research. In this thesis we considered spatial and temporal complexity in parallel, without exploring what happens when a network with different periods is considered. This design problem can be faced with several different strategies: adjusting current heuristics to consider different periods, developing new algorithms or relaxed linear programming problems, comparing second-best strategies as done in chapter 8 but over the whole network, or performing an analysis similar to what we did in chapter 6 but over simple networks rather than a single line. In any of these strategies, the interaction between considering different periods and lines structures should be studied with special care: are lines structures necessarily constant in time?

All of the analyses performed here can be extended to a scheme in which more than one mode is considered. When real public transport networks are designed (or extended), a crucial aspect is to decide which modes of transport to use. Considering various modes in the simple design problem is a first step to analyze later schemes with spatial and/or temporal complexity. This appears to be particularly relevant for scale analysis, because gathering enough passengers to justify some faster and more expensive transit modes seems to be a new discrete source for scale economies.

When studying heuristics, we found that they all created direct-type lines structures. Developing heuristics that create other type of structures, such as feeder-trunk or hub and spoke, is also a relevant challenge. There are combinatorial techniques that could be useful for this, as the *p-hub* problem (Ernst and Krishnamoorhty, 1996) that seeks for the best set of hubs of size p; these techniques would need to be adapted to represent the specific characteristics of transit design.

Finally, new technologies are expected to induce relevant transformations in transport systems in general, and in public transport in particular. Studying how to design a transit system that includes autonomous vehicles, massive coordination ability between vehicles (and between vehicles and passengers), electric combustion and other new features is crucial and will inspire new research challenges regarding their economic aspects.

Bibliography

Alekseev, V.B., 2004. Abel's Theorem in Problems and Solutions: Based on the Lectures of Professor VI Arnold. Springer Science & Business Media, Dordrecht.

Alonso, W. (1964). Location and land use. Toward a general theory of land rent. Harvard University Press, Cambridge.

Badia, H., Estrada, M., and Robusté, F. (2014). Competitive transit network design in cities with radial street patterns. *Transportation Research Part B: Methodological*, *59*, 161-181.

Basso, L., and Jara–Díaz, S. (2005). Calculation of economies of spatial scope from transport cost functions with aggregate output with an application to the airline industry. *Journal of Transport Economics and Policy (JTEP)*, *39*(1), 25-52.

Basso, L., and Jara-Díaz, S. (2006a). Are returns to scale with variable network size adequate for transport industry structure analysis? *Transportation Science*, *40*(3), 259-268.

Basso, L., and Jara-Díaz, S. (2006b). Distinguishing multiproduct economies of scale from economies of density on a fixed-size transport network. *Networks and Spatial Economics*, *6*(2), 149-162.

Basso, L., and Jara-Díaz, S. (2010). The case for subsidisation of urban public transport and the Mohring effect. *Journal of Transport Economics and Policy*, 44(3), 365-372.

Basso, L., and Jara-Díaz, S. (2012). Integrating congestion pricing, transit subsidies and mode choice. *Transportation Research Part A: Policy and Practice*, 46(6), 890-900.

Basso, L., and Silva, H. (2014). Efficiency and substitutability of transit subsidies and other urban transport policies. *American Economic Journal: Economic Policy*, *6*(4), 1-33.

Beirão, G., and Cabral, J. S. (2007). Understanding attitudes towards public transport and private car: A qualitative study. *Transport policy*, *14*(6), 478-489.

Borndörfer, R., Grötschel, M., and Pfetsch, M. E. (2007). A column-generation approach to line planning in public transport. *Transportation Science, 41(1)*, 123-132.

Caves, D., Christensen, L., and Tretheway, M. (1984). Economies of density versus economies of scale: why trunk and local service airline costs differ. *The RAND Journal of Economics*, 471-489.

Ceder, A. and Wilson, N. H. (1986). Bus network design. *Transportation Research Part B: Methodological, 20(4)*, 331-344.

Cenek, J. (2010). Line routing algorithm. *Journal of Information, Control and Management Systems*, *8*(1), 3-10.

Chang, S. K. and Schonfeld, P. M. (1991). Multiple period optimization of bus transit systems. *Transportation Research Part B: Methodological, 25(6)*, 453-478.

Chriqui, C., and Robillard, P. (1975). Common bus lines. *Transportation Science*, *9*(2), 115-121.

Cominetti, R., and Correa, J. (2001). Common-lines and passenger assignment in congested transit networks. *Transportation Science*, *35*(3), 250-267.

Currie, G. (2005) The demand performance of Bus Rapid Transit. *Journal of Public Transportation*, 8, 41-55.

Daganzo, C.(2009). A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. *Transportation Research Part B: Methodological*, *43*(10), 913-921.

Daganzo, C. F. (2010). Structure of competitive transit networks. *Transportation Research Part B: Methodological*, *44*(4), 434-446.

De Borger, B., Mayeres, I., Proost, S., and Wouters, S. (1996). Optimal Pricing of Urban Passenger Transport: A Simulation Exercise for Belgium. J*ournal of Transport Economics and Policy 30 (1)*, 31–54.

Dubois, D., Bel, G. and Llibre, M. (1979). A set of methods in transportation network synthesis and analysis. *Journal of the Operational Research Society, 30(9)*, 797-808.

Ernst, A. T. and Krishnamoorthy, M. (1996). Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location science*,*4*(3), 139-154.

Farsi, M., Fetz, A., and Filippini, M. (2007). Economies of scale and scope in local public transportation. *Journal of Transport Economics and Policy*, *41*(3), 345-361.

Fernández, J., De Cea, J., and De Grange, L. (2005). Production costs, congestion, scope and scale economies in urban bus transportation corridors. *Transportation Research Part A: Policy and Practice*, *329*(5), 383-403.

Fielbaum, A., Jara-Diaz, S., and Gschwender, A. (2016). Optimal public transport networks for a general urban structure. *Transportation Research Part B: Methodological*, *94*, 298-313.

Fielbaum, A., Jara-Diaz, S. and Gschwender, A. (2017) A parametric description of cities for the normative analysis of transport systems. *Networks and Spatial Economics 17*, 343-365.

Fielbaum, A., Jara-Diaz, S. and Gschwender, A. (2019a). Beyond the Mohring effect: scale economies induced by transit lines structures design. Submitted to: Economics of Transportation.

Fielbaum, A., Jara-Diaz, S., and Gschwender, A. (2019b). The role of lines density in the strategic design of transit networks. Submitted: Transportation Research Part B: Methodological.

Friman, M., Edvardsson, B., and Garling, T. (1998). Perceived service quality attributes in public transport: Inferences from complaints and negative critical incidents. *Journal of Public Transportation*, *2*(1), 4.

Garcia-Martinez, A., Cascajo, R., Jara-Diaz, S. , Chowdhury, S., and Monzon, A. (2018). Transfer penalties in multimodal public transport networks. *Transportation Research Part A: Policy and Practice, 114*, 52-66.

Glaister, S., and Lewis, D. (1978). An integrated fares policy for transport in London. *Journal of Public Economics*, *9*(3), 341-355.

Gschwender, A., Jara-Díaz, S. and Bravo, C. (2016). Feeder-trunk or direct lines? Economies of density, transfer costs and transit structure in an urban context. *Transportation Research Part A: Policy and Practice 88*, 209-222.

Hamilton, B., and Röell, A. (1982). Wasteful commuting. *Journal of political economy*, *90*(5), 1035-1053.

Hörcher, D., and Graham, D. J. (2018). Demand imbalances and multi-period public transport supply. *Transportation Research Part B: Methodological*, *108*, 106-126.

Hurdle, V. F. (1973). Minimum cost locations for parallel public transit lines. *Transportation Science*, *7*(4), 340-350.

Jansson, J. O. (1980). A simple bus line model for optimization of service frequency and bus size. *Journal of Transport Economics and Policy*, 53-80.

Jansson, J. O. (1984). *Transport System Optimization and Pricing (Chichester: Wiley)*

Jara-Díaz, S. (1982a). The estimation of transport cost functions: a methodological review. *Transport Reviews*, *2*(3), 257-278.

Jara-Díaz, S. (1982b) Transportation product, transportation function and cost functions. *Transportation Science* 16, 522-539.

Jara-Dıaz, S., and Basso, L. (2003). Transport cost functions, network expansion and economies of scope. *Transportation Research Part E: Logistics and Transportation Review*, *39*(4), 271-288.

Jara-Díaz, S., and Cortés, C. (1996). On the calculation of scale economies from transport cost functions. *Journal of Transport Economics and Policy*, 157-170.

Jara-Díaz, S. R. and Gschwender, A. (2003a). From the single line model to the spatial structure of transit services: corridors or direct?. *Journal of Transport Economics and Policy, 37(2)*, 261-277.

Jara-Díaz, S., and Gschwender, A. (2003b). Towards a general microeconomic model for the operation of public transport. *Transport Reviews*, *23*(4), 453-469.

Jara-Díaz, S. R. and Gschwender, A. (2009). The effect of financial constraints on the optimal design of public transport services. *Transportation* 36 (1), 65-75.

Jara-Díaz, S., Fielbaum, A., and Gschwender, A. (2017). Optimal fleet size, frequencies and vehicle capacities considering peak and off-peak periods in public transport. *Transportation Research Part A: Policy and Practice*, 106, 65-74.

Jara-Díaz, S., Fielbaum, A., and Gschwender, A. (2019). Public transport design considering two periods: some necessary extensions. Working paper.

Jara-Díaz, S., and Tirachini, A. (2013). Urban bus transport: open all doors for boarding. *Journal of Transport Economics and Policy*, 47(1), 91-106.

Keaton, M. (1990). Economies of density and service levels in U.S. railroads: an experimental analysis. *Logistics and Transportation Review* 26, 211-227.

Kocur, G. and Hendrickson, C. (1982). Design of local bus service with demand equilibration. *Transportation Science, 16(2)*, 149-170.

Kraus, M. (2008). Economies of scale in networks. *Journal of Urban Economics*, *64*(1), 171-177.

Laporte, G., Mesa, J.A., Ortega, F.A., and Perea, F., (2011). Planning rapid transit networks. *Socio-Econ. Plann. Sci.* 45 (3), 95-104.

Lin, J., and Ban, Y. (2013). Complex network topology of transportation systems. *Transport reviews*, *33*(6), 658-685.

Louail, T., Lenormand, M., Picornell, M., Cantú, O., Herranz, R., Frias-Martinez, E., Ramasco, J. and Barthelemy, M. (2015). Uncovering the spatial structure of mobility networks. *Nature Communications*, *6*, 6007.

Masucci, A. P., Smith, D., Crooks, A., & Batty, M. (2009). Random planar graphs and the London street network. *The European Physical Journal B*, *71*(2), 259-271.

Medina, M., Giesen, R., and Muñoz, J. (2013). Model for the optimal location of bus stops and its application to a public transport corridor in Santiago, Chile.

Transportation Research Record: Journal of the Transportation Research Board, (2352), 84-93.

Mohring, H. (1972). Optimization and scale economies in urban bus transportation. *The American Economic Review, 62(4)*, 591-604.

Newell, G. F. (1974). Control of pairing of vehicles on a public transportation route, two vehicles, one control point. *Transportation Science*, *8*(3), 248-264.

Osuna, E., and Newell, G. (1972). Control strategies for an idealized public transportation system. *Transportation Science*, *6*(1), 52-72.

Parry, I., and Small, K. (2009). Should urban transit subsidies be reduced?. *American Economic Review*, *99*(3), 700-724

Pattnaik, S. B., Mohan, S. and Tom, V. M. (1998). Urban bus transit route network design using genetic algorithm. *Journal of Transportation Engineering, 124(4)*, 368-375.

Proost, S., and Van Dender, K. (2008). Optimal urban transport pricing in the presence of congestion, economies of density and costly public funds. *Transportation Research Part A: Policy and Practice*, *42*(9), 1220-1230.

Quak, C.B. (2003). A passenger-oriented approach of the construction of a global line network and an efficient timetable. *Delft University, Netherlands*.

Raveau, S., Guo, Z., Muñoz, J. C., and Wilson, N. H. (2014). A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. *Transportation Research Part A: Policy and Practice*, *66*, 185-195.

Redman, L., Friman, M., Gärling, T., and Hartig, T. (2013). Quality attributes of public transport that attract car users: A research review. *Transport Policy*, *25*, 119-127.

Schöbel, A. and Scholl, S. (2005). Line planning with minimal transfers. *In 5th Workshop on Algorithmic methods and Models for Optimization of Railways* (No. 06901).

Schonfeld P. M. (1981) Minimum Cost Transit and Paratransit Services. Transportation Studies Center Report, Dept. of Civil Engineering, University of Maryland, College Park.

Small, K.A. (2004). Road pricing and public transport, in G. Santos (eds.), Road Pricing: Theory and Evidence, Research in Transportation Economics, Vol. 9, Elsevier Science, 133-158

Tirachini, A., Hensher, D., and Jara-Díaz, S. (2010a). Comparing operator and users costs of light rail, heavy rail and bus rapid transit over a radial public transport network. *Research in transportation economics*, 29(1), 231-242.

Tirachini, A., Hensher, D. and Jara-Díaz, S. (2010b). Restating modal investment priority with an improved model for public transport analysis. *Transportation Research Part E: Logistics and Transportation* Review, 46(6), 1148-1168.

Tirachini, A. and Hensher, D. (2011). Bus congestion, optimal infrastructure investment and the choice of a fare collection system in dedicated bus corridors. *Transportation Research Part B: Methodological*, 45(5), 828-844.

Tirachini, A. and Hensher, D. (2012). Multimodal transport pricing: first best, second best and extensions to non-motorized transport. *Transport Reviews*, 32(2), 181-202.

van Nes, R., Hamerslag, R., and Immers, L. H. (1988). The design of public transport networks (Vol. 1202). National Research Council, Transportation Research Board.

Villalobos Zaid, M. (2018). Estructuras óptimas de líneas de transporte público considerando distintos periodos en Santiago. (Civil Engineering Thesis). Universidad de Chile, Santiago, Chile.

Viton, P. (1992). Consolidations of scale and scope in urban transit. *Regional Science and Urban Economics*, *22*(1), 25-49.

Walker, J. (2012). *Human Transit*, Island Press, Washington D.C.

Appendix

Appendix A: Numeric values of the parameters

Parameter	c_0	c_1	T_0	g	t	p_v	p_w	a	n	p_R
Value	10.65 [US$/h]	0.204 [US$/h]	30 [min]	1/3	2.5 [sec]	1.48 [US$/h]	4.44 [US$/h]	0.8	8	0.59 [US$]

Table A1: Parameters used in chapters 2 and 3.

Parameter	α	β	c_0	c_1	T_0	g	t
Value	0.5	0.25	8.5 [US$/h]	0.204 [US$/h]	30 [min]	1/3	2.5 [sec]

Parameter	p_v	p_w	a	n	p_R	L_0	Q	V
Value	1.48 [US$/h]	2.96 [US$/h]	0.8	8	0.59 [US$]	30 [km]	2 [km]	13 [km/h]

Table A2: Parameters used in chapter 4.

Parameter	α	β	c_0	c_1	T_0	g	t
Value	0.25	0.22	10.65 [US$/h]	0.204 [US$/h]	30 [min]	1/3	2.5 [sec]

Parameter	p_v	p_w	p_a	a	n	p_R	P	v_a
Value	1.48 [US$/h]	4.44 [US$/h]	5.33 [US$/h]	0.8	8	0.37 [US$]	2 [km]	5 [km/h]

Table A3: Parameters used in chapter 5.

Parameter	E_P	E_N	T_P	T_N	l_P	l_N	t
Value	5 [h]	13 [h]	2 [h]	1.5 [h]	10 [km]	5 [km]	2.5 [sec]

Parameter	L	c_{BC}	c_{KC}	c_{BO}	c_{KO}	p_v	p_w
Value	40 [km]	4.14 [US$]	0.45 [US$]	1.32 [US$/h]	0.1 [US$/h]	1.48 [US$/h]	4.44 [US$/h]

Table A4: Parameters used in chapters 6,7 and 8.

Appendix B: Some details on the application of each heuristic to the city model (section 3.1).

Dubois *et al* (1979)

The following pseudo-code describes the heuristic:

```
(0) Define the value of the tolerance $\sigma$
(1) Define the set $M$ that contains all the minimum-length (i.e. sum of the distances of each arc) routes that cannot be extended preserving the length condition.
```

(2) Build the set M^*, which starts empty:
For each route $L \in M$, denote x its initial node and y its final node.
For each z not in L, define L_z the route that starts in x, goes to z and then to y always following minimum-length paths.
If the length of L_z exceeds the length of L by a fraction smaller than an exogenously fixed tolerance σ, then add L_z to M^* and stop searching for L. Start searching for L_z, but always comparing lengths with respect to L.

(3) Define $\bar{M}$ as the union of M^* and the paths in M connecting origins and destinations that are not connected by paths in M^*. The routes in this set are the candidates.

(4) Arrange the routes in $\bar{M}$ according to some rule.

(5) Define M' as an empty set. Add the paths in $\bar{M}$ to M' until the city is fully connected (admitting transfers).

(6) Calculate the portion of trips that need one or more transfers. If this portion is “small”, go to step (8). Otherwise, go to (7).

(7) Add the route in $\bar{M}$ that minimizes the number of required transfers. Go to (6).

(8) Stop if the difference between the average travel time in the system and the average time in a system where all the trips go through the minimum-length path is “small”. Otherwise, go to (9).

(9) Add the route in $\bar{M}$ that decreases the most the average travel time. Go to (8).

In the parametric city, the set M is composed by all the routes connecting two peripheries following an optimal path (crossing the CBD or through the subcenters ring, depending how far are the respective zones). First we assume $\sigma < 1/3$. We check all the possible routes L_z; we verify that M^* happens to be composed only by the routes that go from one periphery to a periphery that is 3 zones away through the subcenters ring (instead of crossing the CBD). So the candidate set $\bar{M}$ is composed by paths going through the subcenters ring that reach peripheries that are 1, 2 or 3 zones away, and paths that go the opposite subcenter (i.e., the one that is 4 zones away) crossing the CBD. To arrange $\bar{M}$ (step 4), the authors propose criteria such as the most used or the least costly routes. We adapt the last one, arranging the set from the shortest to the largest route. In step (6), if we only added the shortest routes in $\bar{M}$ (i.e., the routes that go the neighbor zones) we would connect almost the whole graph: only the CBD would remain unconnected. For this not to happen we add the routes that go from each periphery to its opposite crossing the CBD. The graph is now connected, but all the trips that go to two or more zones away (with the exception of the opposite zone) require a transfer, so the number of transfers is at least[22] $(a\gamma + b\tilde{\gamma})(n-4)Y$. If $\gamma < 0.1$, we skip step (7); otherwise we add the route that go from one periphery to the periphery that is three zones away and we eliminate all the transfers. As the travel time cannot be reduced adding more routes, the final line structure has been reached.

Afterwards, we solve for $\sigma > 1/3$. Doing so, the line that goes from a periphery to the opposite subcenter presents a small variation: after going to the CBD, it goes to one neighbor of the opposite subcenter and then finishes in the opposite subcenter. Notice that in this case the number of transfers will always be smaller than $\frac{n-5}{n}(a\gamma + b\tilde{\gamma})$, and

[22] The routes that minimize the number of transfers are those in which the user goes to the CBD first and take the second bus there. Other routes require more transfers.

when $n = 8$ this is smaller than $\frac{2}{7}$; therefore, for $\sigma > 1/3$ a single structure is obtained. It is worth commenting that in the case $\sigma < 1/3$ with $\gamma > 0.1$ passengers whose destination is located three zones away from their origin could go either through the subcenters ring, without making transfers, or take a shorter trip changing buses in the CBD; as their choice depends on the frequencies finding the optimal frequency requires iterations. We first assign all these passengers with no transfers and calculate the resulting optimal frequencies; then we verify whether the choice with no transfers is in fact the min user cost. If not, we re-calculate considering transfers. As expected the results depend on the parameters.

Ceder and Wilson (1986)

This algorithm builds routes that depart from a terminal (the first origin), searching for trees (i.e., connected graphs with no cycles). To do so, select any terminal as the first one and consider the following pseudo-code:

(1) Search for all the nodes that you can reach from the current origin; if no node is reachable, go to (2), otherwise, for each reachable node, if it has not been connected, and if the total length of the path does not exceed the length of the shortest path by a percentage σ, add it to the tree. If there is no such node, go to (2). Otherwise, select any of these nodes as the new origin for the same terminal and repeat (1).
(2) Select as the new origin the last node added to the tree that has not been an origin yet. If there is no such node, go to (3).
(3) If there are no more terminals, end. Otherwise, select a new terminal, define it as the origin and go to (1).

In our scheme, routes start from a periphery (predefined as terminals) and necessarily go to the own subcenter. Then it is possible to go to neighbor subcenters or to the CBD. We start exploring the routes through the CBD such that, depending on the value of σ, we will reach $H =$3, 5 or 7 subcenters. Doing so, $H = 3$ if $\sigma < 0.2517$ ($\approx \frac{1}{4}$) and $H = 5$ otherwise (recall that H is the number of foreign subcenters that are reached crossing the CBD). Once the routes through the CBD have been explored, the rest of the routes tour the subcenter ring until the whole graph is covered. Afterwards it is impossible to extend any route without reaching a previously built route that started from the same terminal.

Borndörfer *et al* (2007)

The optimization problem, in this scheme, turns into

Minimize $\sum_{R\in U} p_v t_R y_R + \sum_{l\in L} c_0 B_l$
s.t.
$\sum_{R:O(R)=s,D(R)=t} y_R = (OD)_{st}$ ∀ nodes s,t (A1)
$\sum_{R:a\in R} y_R \leq \sum_{R:a\in R} f_R K$ ∀ edge a (A2)
$\sum_{l:a\in l} f_L \leq \Gamma_a$ ∀ edge a (A3)
$f \geq 0$ (A4)
$y \geq 0$ (A5)

The variables are the vector of frequencies f and the vector of passenger flows y. The (total) social cost only considers travel times and a fixed cost per bus. In the original model there is a fixed cost per line, but to be consistent we put them equal to zero. Waiting times and transfers play no role.

The first equation imposes that passengers on all routes R serving a given O-D pair add up to the corresponding O-D demand. Note that this means that passenger assignment to routes is endogenous aiming at minimizing social cost, which may not coincide with the individual preference of each user. As K is the size of each bus, the second equation imposes bus capacity constraints. The third equation relates frequency to streets capacity, Γ_a, but in our model this is not considered so, in practice, we used $\Gamma_a = +\infty\ \forall a$.

Cenek (2010)

First assume that all passengers will take the shortest route to go from their origin to their destination; this permits the construction of arc-specific weights as the number of passengers that use that arc under this assumption. Then, the following procedure is applied:

(1) Select the arc with the largest weight from those that finish in a center (if that weight is 0, finish). Define the node u as the other extreme of that arc.
(2) Extend the line. Select the arc with the highest weight from those that are incident in u. Update u as the other extreme of that arc. If u is a center or has no other incident arc with positive weight, go to (3). Otherwise, repeat (2).
(3) The line has been built and added to the set of lines. Calculate w as the minimum weight of the arcs present in this line. Subtract w from the weight of all the arcs in the line. Go to (1).

To analyze the application to our scheme, some route notation is needed. A route will be denoted by its initial node (whose zone will be always denoted by i) and the final node (denoted by j if different from i). Note that if the destination is the own subcenter or the CBD, there is only one possible route. If the destination is a foreign subcenter, when the route goes through the CBD it is marked with an H; if it goes across the subcenters ring is marked with an F. For example LP_iCBD is a line that starts in a periphery, stops in the subcenter from the same zone and finish its tour in the CBD. In this case there is no need to specify H or F because there is only one route. LSC_iSC_jH is a route that starts in a subcenter and goes to another, but stopping previously in the CBD.

To apply this heuristic to our scheme, let us note that the weight of each arc depends only on the types of nodes that it links, e.g. w of $P_i - SC_i$ is a or w of $SC_i - CBD$ is $a\left(\alpha + \frac{3}{7}\gamma\right) + b\left(\tilde{\alpha} + \frac{3}{7} + \tilde{\gamma}\right)$. Depending on the values of the parameters a, α and β, the largest weight of an arc will be either $P_i - SC_i$ or $SC_i - CBD$. As all the arcs are incident in a center, it would be impossible to have lines that tour more than one arc, so we do not stop when arriving at another center.

<u>First case</u>: $a > a(\alpha + \frac{3}{7}\gamma) + b(\tilde{\alpha} + \frac{3}{7}\tilde{\gamma})$

The first line added is $LP_i - SC_i$; then, we start with an arc $SC_i - CBD$ and is then extended to $CBD - SC_j$, so the line is LSC_iSC_jH (we assume that the final subcenter is the opposite to the initial subcenter); finally, the last type starts again with arc $SC_i - CBD$ but is then extended to $SC_i - SC_{i+1}$: this line is denoted LSC_iCBDb.

As a result *a posteriori*, when finding optimal frequencies the line LSC_iCBDb will always present a null frequency. This means that the structure is a mixture between the feeder-trunk structure (because all the passengers from the periphery take the feeder bus to their subcenter) and the hub and spoke structure (because the CBD will be a hub where almost all the passengers that go to a foreign subcenter will transfer).

Second case: $a < a(\alpha + \frac{3}{7}\gamma) + b(\tilde{\alpha} + \frac{3}{7}\tilde{\gamma})$
In this case the first line starts with an arc $SC_i - CBD$, and it extends to $P_i - SC_i$, so the line is LP_iCBD; the second type starts again with arc $SC_i - CBD$ and it extends to $CBD - SC_j$, so the line is LSC_iSC_jH (as in the first case, we assume that the final subcenter is the opposite to the initial subcenter); finally, the last type of line starts with arc $SC_i - SC_{i+1}$, it extends to $SC_{i+1} - SC_{i+2}$ and so on, so finally the line is the circular line presented in some previous structures.

Appendix C: Analysis of line structures over the isosceles-city (section 4.3.2).

The analysis will be done for the single-line case because it is the more complicated of the two cases. The other one (two identical direct lines) can be solved based on the analysis of a single-line case that connects two points. Each of the components of the single-line cyclical system can be expressed as a function of frequency:

- Bus capacity (K): total passengers per unit time Y use f buses per unit time, such that the load of each bus is $K = Y/f$.
- Cycle time (t_c): regarding vehicle in motion, each bus needs to travel across a path whose length is $2L_0 + Q$, taking a time of $(2L_0 + Q)/V$; regarding time at stops, each passenger needs $2t$ to board and alight a bus whose load is Y/f passengers, which makes a total of $2tY/f$. Total cycle time is the sum of these two terms: $\frac{2L_0+Q}{V} + 2t\frac{Y}{f}$.
- Fleet (B): recalling that $f = \frac{B}{t_c}$, it becomes apparent that $B = f\frac{2L_0+Q}{V} + 2tY$.
- Waiting time (t_w): passengers arrive at an homogeneous rate to the bus stop, and buses exhibit a constant headway such that on average each passenger will wait half the headway $(1/2f)$.
- In-vehicle time (t_v): it needs to be calculated as the average between two types of OD-passengers. Passengers that alight from the bus at the first stop travel a distance L_0 such that time in-motion is L_0/V. At the first stop the bus stays $\frac{Y}{2f}t$, and users that alight there spend on average half of that time. Passengers that alight at the second stop travel a distance $L_0 + Q$; they stay in the vehicle $\frac{Y}{2f}t$ at the first stop, and - in average - half that time at the second stop. The average in-vehicle time for passengers is then $\frac{1}{2}\left[\left(\frac{L_0}{V} + \frac{Y}{4f}t\right) + \left(\frac{L_0+Q}{V} + \frac{Y}{2f}t + \frac{Y}{4f}t\right)\right]$.

Replacing all previous expressions in $VRC = B(c_0 + c_1K) + p_wYt_w + p_vYt_v$ yields:

$$VRC = (f\frac{2L_0+Q}{V} + 2tY)(c_0 + c_1\frac{Y}{f}) + p_w Y\frac{1}{2f} + p_v Y\frac{1}{2}\left[\left(\frac{L_0}{V} + \frac{Y}{4f}t\right) + \left(\frac{L_0+Q}{V} + \frac{Y}{2f}t + \frac{Y}{4f}t\right)\right] \tag{A6}$$

Making the derivative with respect to f equal to zero yields:

$$f^* = \sqrt{\frac{Y(2tc_1+p_w/2 + p_vYt/2)}{2c_0(L_0+Q)/V}},\ \ K^* = \sqrt{\frac{Y2c_0(L_0+Q)/V}{(2tc_1+p_w/2 + p_vYt/2)}} \tag{A7}$$

Both expressions increase with Y, with f^* tending to a linear function, and K^* tending to some constant when $Y \rightarrow \infty$.