Estimating welfare effects of congestion charges in real world settings

Maria Börjesson and Ida Kristoffersson – KTH


Abstract
According to the standard textbook analysis, drivers as a group will be worse off with congestion charging if not compensated by revenues. This result is confirmed by an analysis of the Stockholm congestion charging scheme using a static model with homogeneous users. However, both this static model and the standard textbook analysis omit three important factors: taste heterogeneity, effects of charges on the larger network arising from less blocking back of upstream links and the possibility for drivers to reschedule. Taking account of these factors, using a dynamic scheduling model with heterogeneous users estimated and calibrated for Stockholm, we find that drivers as a group benefit from the charging scheme in Stockholm without recycling of revenues. This paper further investigates the importance of the three mentioned factors. We find that all three factors add significantly to the benefit of the charges and that the most important is heterogeneity in the value of travel time savings. This paper also provides an update on the consumer benefits of the Stockholm charges.

Keywords: Congestion charges, Road Pricing, Acceptability; Evaluation; Urban Transport Policy; Transport externalities

JEL Codes: D61, R41, R42, C25, J22
Introduction

It is generally known that congestion charging can be an effective measure to solve environmental and congestion problems in urban areas, but there is still low political and public acceptability in many urban areas. One possible reason for low acceptability is the notion that drivers are worse off if not compensated by the revenues. In fact, the standard textbook analysis of congestion charging (Walters, 1961), using a static model of one origin and one destination connected by one link and homogeneous users, shows that drivers are worse off with congestion charges if not compensated by return of revenues. This result is confirmed by an analysis of the Stockholm congestion charging scheme using the national traffic model “Sampers”, which applies a single value of travel time savings (VTTS) for each trip purpose and travel mode1 (Engelson and van Amelsfort, 2011).

Both the Sampers and the standard textbook analysis disregard three important factors increasing the benefits of congestion charges: network effects, heterogeneity in the value of time and scheduling costs and dynamics in the temporal dimension. First, network effects imply that travellers who do not pay charges may benefit from them. Network effects are disregarded in the standard textbook analysis and greatly underestimated in static network assignment models that do not model blocking back of upstream links as queues builds up. Second, ignoring heterogeneity in VTTS in a system with a free parallel road leads to great underestimation of benefits because efficiency gains due to separation of traffic are ignored (Verhoef and Small, 2004). Third, in a one-link dynamic bottleneck setting with homogeneous users Arnott et al. (1994) show that an optimal time varying congestion charge is welfare neutral for drivers if not compensated by return of revenues, since the reduction in queuing costs exactly compensates the charge.

The purpose of this paper is to estimate to what extent the three factors discussed above add to the social benefit of the Stockholm congestion charging scheme, and a slightly modified scheme, using the model “Silvester”, which links a dynamic network assignment model with a mode and departure time choice model assuming heterogeneous users (Börjesson, 2008; Kristoffersson and Engelson, 2009). We find that each factor adds significantly to the benefit of the charging scheme. In fact, the Silvester analysis indicates that drivers as a group benefit directly from the congestion charging scheme, without compensation. Direct benefits for many drivers could be one factor explaining the current high public support for the congestion charges in Stockholm, which has even increased since the congestion charging scheme was introduced in 2006 (Börjesson et al., 2011). The Silvester model is estimated for Stockholm and calibrated using actual traffic flows before and after the introduction of the congestion charges. This paper therefore also provides an update on the benefits of the Stockholm congestion charges.

There is a large literature studying welfare effects of congestion charges. Most of the literature is theoretical (Verhoef and Small, 2004; Arnott et al., 1994; Glazer and Niskanen, 2000; Evans, 1992). There are also a few studies on real-world congestion charging schemes, most of them based either on observed travel times or on travel times from static assignment models. Eliasson (2009) provides an a posteriori cost-benefit analysis of the Stockholm congestion charges, based on observed travel times. Eliasson’s study results in a net benefit of about 80 M€/year and as much as 40% of the time gains arise on links outside the cordon. Santos and Shaffer (2004) present and discuss a cost benefit analysis of the London congestion charging scheme undertaken by Transport for London (TfL), which is also based on observed travel data. Prud’homme and Bocarejo (2005) have undertaken another cost-benefit analysis of the London congestion charging scheme based on observed data. The results of the two analyses for London are very different: TfL finds a net benefit of the charging system of about 70 M€/year (similar to the result for Stockholm given above), whereas Prud’homme and Bocarejo find a net loss of about the same size. The main difference in the two studies lies, according to Mackie (2005) and Raux (2005), in the

1 For drivers the value of time is 6.7 €/h for work trips and 4.4 €/h for other trips.
calculated travel time savings and the VTTS. Prud’homme and Bocarejo (2005) do not consider travel time savings outside of the charging zone and apply a lower VTTS.

Two independent cost benefit analyses have been made of a proposed marginal social cost pricing scheme in the Oslo-Akershus metropolitan region, which has a population of about one million. Grue et al. (1997) find a social benefit of 49 €/capita/year\(^2\) for the Oslo-Akershus area. The benefit is somewhat higher, 75 €/capita/year, in Fridström et al. (2000). The higher benefit in Fridström et al. (2000) is according to Vold et al. (2001) due to the fact that the transport models differ and because the cost-benefit analysis is somewhat simpler in Grue et al. (1997). The difference in the transport models is not in assignment; both use the static Emme/2 model. Rather, the two models differ on the demand side, where the model used in Fridström et al. (2000) includes trip frequency, destination and mode choice, whereas the model used in Grue et al. (1997) includes on the demand side route and departure time choice.

Rich and Nielsen (2007) provide social benefit calculations of four proposed charging schemes in Copenhagen using advanced route choice models but not dynamic assignment. Maruyama and Sumalee (2007) compare social benefit of different charging schemes in Utsunomia in Japan using volume-delay functions to calculate link travel times, i.e. static assignment. The conclusion in Maruyama and Sumalee (2007) is that area-based schemes are in general socially more beneficial, but also more inequitable than cordon-based schemes. Kickhöfer et al. (2010) provide social benefit calculations of several proposed distance-based charging schemes in Zürich using the activity-based model Matsim.

It is well known that congestion pricing normally will generate a net welfare surplus, but different researches have come to different conclusions as to whether congestion charges will be progressive or regressive. The conclusion is largely dependent on what assumptions are made about the distributions of VTTS, scheduling preferences and flexibility (Verhoef and Small, 2004; Arnott et al., 1994). Fosgerau and de Palma (2010) show that introducing special effects in the bottleneck models will lead to different conclusions as to who will win and who will lose. It is also dependent on whether citizens have access to and actually patronize public transport and “slow modes” and how the revenues are used (Verhoef and Small, 2004; Glazer and Niskanen, 2000; Armelius and Hultkrantz, 2006). Note also that in case there are positive external labour market effects (Anderstig et al., 2012), including taxation and agglomeration effects, or in case the productivity of firms increase due to shorter travel times, this will benefit the entire society and not only individual travellers. The distributional effects are not the main focus of this paper. However, in the study of welfare effects we are touching upon distributional effects when analysing gains and losses of different groups of travellers. The distributional effects are important, since it is critical to design congestion charges such that a large proportion of the population perceives themselves as winners in order to gain acceptability from the public, as noted by several authors (Arnott et al., 1994; Eliasson, 2008; Schaller, 2010).

The Stockholm congestion charging scheme consists of a cordon around the inner city, reducing traffic through the bottlenecks located at the arterials leading into the inner city. In this paper we also analyse a slightly modified version of the scheme, which there is political consensus to introduce at a later stage\(^3\). The modified scheme generates substantially higher benefits than the present scheme. The charging trial has been described in detail elsewhere (Börjesson et al., 2011; Eliasson, 2008; Eliasson et al., 2009). Since it is usually assumed that drivers are not fully compensated by shorter travel times, the recommendation in the acceptance literature is that congestion charges must be part of a “package” and that the acceptance will be strongly dependent on how the revenues are used (Eliasson, 2008). The present study suggests that this may not always be the case, since many drivers may benefit even if not compensated by the revenues. One of the most interesting and encouraging results of the Stockholm congestion charges in Stockholm has been the positive trend in

---

\(^2\) 380 NOK converted to € with the conversion rate 1 NOK = 0.13 €.

\(^3\) There is political consensus to introduce the modified scheme when a new western bypass has been built around Stockholm. The new bypass is however not included in the forecasts of this paper.
public and political support of the charges to the extent that acceptability is no longer an issue. In Stockholm municipality the public support has gradually risen from less than 30% in favour of the charges to slightly more than 50% at the end of the trial, and to almost 70% at the end of 2007.

The outline of this paper is as follows. A theoretical background is given in the next section, followed by a description of the dynamic modelling system Silvester in section 3. Section 4 describes the congestion charging scheme in place in Stockholm today and the modified version that is analysed in this paper. Section 5 describes the methods and results are given in Section 6. Section 7 concludes.

**Theory**

From the standard textbook analysis (Walters, 1961), it can easily be shown that all drivers paying a congestion charge are not fully compensated by shorter travel times. Let the total generalized cost of travelling on the link be given by Equation (1):

\[ C = fc(f) \]  

(1)

In Equation (1), flow is \( f \) and \( c \) is average generalized cost. The marginal generalized cost of additional traffic on the link is then given by Equation (2):

\[ \frac{dC}{df} = c(f) + fc'(f) \]  

(2)

The traffic flows will, in the case without the congestion charge, increase up to the point of intersection between the average generalized cost function and the demand function (point III in
Figure 1). This is not an efficient allocation of resources. To reach efficiency, traffic flow must decrease to the point $f^*$, which is the flow corresponding to the point of intersection between the marginal generalized cost function and the demand function (point $I'$ in
Figure 1).
To lower the flow to \( f^* \), the average generalized cost, \( C/f \), can be increased to equal the marginal generalized cost at flow \( f^* \) by adding an optimal charge \( \tau^* \). At flow \( f^* \) the marginal generalized cost is \( dC/df \big|_{f=f^*} = c(f^*) + f^*c'(f^*) \) and the average generalized cost is \( C^*/f^* = c(f^*) + \tau^* \). The optimal charge is thus \( \tau^* = f^*c'(f^*) \). Drivers staying on the road will pay \( \tau^* \) and their generalized travel cost (except the congestion charge) decreases by \( c^0 - c^* \). Since \( \tau^* > c^0 - c^* \) the drivers remaining on the road will be worse off than without charges; they will lose the area \([c^0-IV-I-c^1]\). Drivers priced off the road lose the area under the demand curve \([IV-I-III]\). The net benefit of the congestion charges is the area \([c^*-V-I-c^1]\) (revenues) minus the area \([c^0-IV-I-c^1]\) (loss for the remaining drivers) and minus the area \([IV-I-III]\) (loss for drivers priced off the road), which equals the area \([I-II-III]\) (see for instance Johansson and Mattsson (1995)). The standard textbook analysis described above disregards several important factors, opening for the possibility that some drivers become better off in a situation with congestion charges.

First, the standard analysis assumes that there is one single origin-destination pair (OD-pair) connected by one link. In a network setting the benefit could increase due to route choice effects (still assuming one OD-pair) or due to the fact that there are many OD-pairs. If route choice in user optimum is different from that of the system optimum, first-best pricing may add to the welfare of the drivers (assuming no compensation by revenues). Braess’ paradox (Braess, 1968) constitutes a good example, stating that an additional link in a network in some cases increases the total travel time in the network due to inefficient route choice at user optimum. Total travel time increases because traffic using the additional link produces additional external costs in the network. If a congestion charge on the additional link internalizes the external cost in the network, then traffic flow at the additional link and the total travel time in the system will decrease, so that drivers as a group could benefit (without paying any charge). Another situation arises if travellers have different OD-pairs.
Verhoef and Small (2004) note that the benefit of first-best pricing on one link in a network is usually underestimated if the traffic on this link gives rise to external costs for traffic having other OD-pairs. The simple network of Figure 2 demonstrates the insight.

Figure 2: A simple network for illustration of network effects.

Some drivers travel (A,B) and other drivers travel (A,C). An optimal congestion charge on the link (B,C) could benefit drivers travelling from A to B. If there is blocking back of upstream links and signal plans at intersections in (A,B), which builds up from bottlenecks in (B,C), the benefit for drivers travelling from A to B may be large even if there is no capacity constraint on the link (A,B).

Second, ignoring heterogeneity in VTTS in a system with a free parallel road (or an efficient public transport system) leads to great underestimation of social benefits, by disregarding the efficiency gains due to separation of traffic (Verhoef and Small, 2004). Specially, congestion charging tends to sort trips between routes and modes with respect to VTTS. Using numerical simulations and a simple network, Verhoef and Small (2004) analyse the benefit of the second-best charging scheme with one of two parallel links charged. They find that travellers with the highest VTTS benefit from congestion charging without compensation by return of revenues. Travellers with the trade-off VTTS incur the greatest losses.

Third, in a dynamic setting the congestion charge may be time varying and the drivers’ possibility to reschedule is taken into account. Arnott et al. (1994) show, assuming a one-link bottleneck model and homogeneous users, that with an optimal time varying charge users adjust their departure time such that queuing is completely avoided. Moreover, they show that the optimal time varying congestion charge is welfare neutral for the users if they are not compensated by return of revenues; the reduction in queuing cost exactly compensates the charge. If drivers have heterogeneous scheduling preferences, still assuming constant VTTS, benefits from the charges would increase even more, due to sorting of travellers with high and low scheduling costs. Hence, drivers as a group could even benefit without return of revenues. Applying a one-link bottleneck model, De Palma and Lindsey (2002) show that if the value of schedule delay (VSD) is homogeneous and only the VTTS heterogeneous, all drivers lose from first best pricing, except those with the highest VTTS. Still, there is an efficiency gain due to the sorting of trips with respect to VTTS between departure times, compared to the situation when the VTTS is homogeneous. Lindsey (2004), analyse the case when both VTTS and VSD vary between discrete groups. Van den Berg and Verhoef (2011) extend the work by Lindsey (2004) and consider a situation with continuously distributed VTTS and VSD. They find that travellers with an intermediate VSD and the lowest VTTS for this VDS suffer the greatest loss, both in the first-best pricing case and the second-best case with a free parallel road. In the second-best case, also those with low VSD may substantially benefit from the second-best scheme, attracted to the earliest and latest departure times on the tolled road.

Fourth, the textbook analysis neglects the benefit of improved travel time reliability due to congestion charging. We do not, however, examine this benefit in this paper, due to the difficulty of analysing reliability in a dynamic large-scale scheduling model. Many travel times must be simulated, and then departure times of the drivers will adjust depending on the distribution of the travel times. It is not clear whether convergence can be reached⁴. However, Fosgerau and Karlström (2010) derive a reduced-form model approach from an underlying scheduling model, assuming that travel times follow a known random

---

⁴ There are scheduling models, for instance the one suggested by Fosgerau and Engelson (2011), where the optimal departure time does not depend on the standard deviation of the travel time distribution, and only on the mean travel time. The analysis becomes simpler in such cases.
distribution and that travellers choose their departure time optimally given this distribution. Different underlying scheduling models imply different statistical measures of reliability, but assuming the classical scheduling model originating from Vickery (1969) and Small (1982) the disutility of travel time unreliability is proportional to the standard deviation of travel time distribution\textsuperscript{5}. This implies that in a static model, travel time reliability may be introduced in the indirect utility function as a standard deviation (or some other statistical measure depending on the underlying scheduling model) under the assumption that the travellers depart at optimal departure time\textsuperscript{6}. Introducing travel time reliability as a reduced form measure can, however, not change the basic feature of the textbook example that drivers as a group become worse off with congestion charges assuming compensation by the revenues. The effect is only that the generalized cost function in

![Figure 1](image_url) becomes steeper when also including travel time reliability.

Fifth, benefits from improved urban environment may arise also from the congestion charges (Eliasson, 2008). This benefit is difficult to value and we do not analyse this benefit further.

In summary, there are a number of reasons why the welfare effects of congestion charging may be underestimated in the standard textbook analysis. In the following we concentrate on the first three reasons described above.

The model

For the analyses we apply the Silvester model, which is calibrated for Stockholm and includes dynamic assignment, heterogeneity in preferences, scheduling and mode (car or public transport) choice. The model includes a supply model and a demand model linked in an iterative procedure. The mesoscopic dynamic assignment model Contram calculates

\textsuperscript{5} Given that scheduling preferences and standardized travel time distribution remains constant.

\textsuperscript{6} Given a relationship between travel time delay and the appropriate measure of reliability.
route choice and resulting travel time and monetary cost for trips in each OD-pair, given the
demand for car trips departing in each fifteen minute interval. A mixed logit model then
takes the travel times and costs from the assignment model and generates the demand for
car trips departing in each fifteen minute interval. The generated demand is then fed back
into the assignment model. Hence, the travel times calculated in the assignment impact
departure time choice in the demand model, and vice versa, in the feed-back loop. A
description of the implementation of the Silvester demand model and the procedure of
connecting the demand model to Contram is found in Kristoffersson and Engelson (2009).

**Departure time and mode switch model**

The demand model in Silvester has been estimated on stated and revealed preferences
(using survey data and observed travel time data) of car drivers in Stockholm before the
introduction of charges (Börjesson, 2008). In the combined SP and RP survey information
concerning trip purpose and whether the drivers had fixed or flexible working hours was
collected. Since scheduling flexibility and cost sensitivity differs between trip purpose and
work schedule flexibility, the trips were segmented with respect to this information in the
estimation. There are three trip purpose segments in Silvester, each with one demand
model: 1) commuting trips with fixed working hours and school trips (short: fixed, 2)

business trips (business) and 3) commuting trips with flexible working hours and other trips
(flexible), where “other trips” includes e.g. shopping and leisure trips.

The demand model is a mixed logit model which builds on the scheduling models of Small
(1982) and Vickrey (1969), assuming that drivers trade-off travel costs (travel time, distance-
based cost, charge etc.) against scheduling delay costs. Equation (3) shows the utility
functions, which are similar for the three trip purpose segments, except that for business
trips the public transport alternative ($1_g3338$) is not available.

$$U_c^{kt} = \beta_1^{kt}SDE^{yt} + \beta_2^{kt}SDL^{yt} + \beta_3^{kt}M^{t\omega} + b_1^kT^{t\omega} + b_2^k\sigma^{t\omega} + \epsilon_c^t,$$

$$U_p^{kt} = C_p^k + b_3^kT_p^{t\omega} + b_4^k\delta^k + \epsilon_p$$

$$SDE^{yt} = \max(y - t, 0)$$

$$SDL^{yt} = \max(t - y, 0)$$

In Equation (3), $U_p^{kt}$ is the utility function for public transport mode $p$, trip purpose $k$ and
OD-pair $\omega$, $U_c^{kt}$ is the utility function for car mode $c$, trip purpose $k$ preferred departure
time interval $y$, actual departure time interval $t$ and draw from parameter distribution $d$.
Index $t = 0$ denotes departure times before 6.30 am, $t = 1, \ldots, 12$ denotes departure times
in the twelve quarters from 6.30 to 9.30 am respectively and $t = 13$ departure times after
9:30 am. $SDE^{yt}$ and $SDL^{yt}$ are schedule deviation early and late respectively for preferred
departure time $y$ and actual departure time $t$. Since time is divided into 15 minute time
intervals, $SDE^{yt}$ and $SDL^{yt}$ are multiples of 15 minutes. $M$ is monetary cost which includes
both the congestion charge and a distance-based cost, $T$ is travel time, $\sigma$ is standard
deviation of travel time, $\epsilon$ is a Gumbel distributed error term, $C_p$ is an alternative specific
constant for public transport and $\delta$ is the share of car drivers who also possess a public
transport monthly card (in the estimation $\delta$ was a dummy variable equal to 1 if the driver
had a public transport monthly card and 0 otherwise).

The demand model forecasts the share of car drivers switching to public transport due to
changes in the network, such as introduction of congestion charges. Since a logit model
always will predict that at least a small share of the population chooses each alternative, some users choose the public transport alternative already in the situation without congestion charges. This share is however small in the Stockholm application and amounts to only 1.6 per cent of all users.

Parameters labelled $\beta$ are heterogeneous in the population following a Johnson’s $S_B$ distribution bounded on $[-1,0]$, whereas parameters labelled $b$ are assumed to be constant in the population. Heterogeneous parameters are simulated using 50 random draws from Johnson’s $S_B$ distribution. The original demand models included a normally distributed error term for the alternative specific constant $C_p$ which had a large spread. Using 50 draws, the effect of this parameter depended too much on the specific draws made. The public transport error term was therefore removed and the alternative specific constant $C_p$ adjusted accordingly.
Table 1: Parameter values in the departure time and mode choice models for the three trip purposes

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Flexible</th>
<th>Fixed</th>
<th>Business</th>
</tr>
</thead>
<tbody>
<tr>
<td>β1 (Schedule delay early, mean)</td>
<td>-0.46</td>
<td>-0.35</td>
<td>-0.25</td>
</tr>
<tr>
<td>β2 (Schedule delay late, mean)</td>
<td>-0.52</td>
<td>-0.56</td>
<td>-0.35</td>
</tr>
<tr>
<td>β3 (Cost, mean)</td>
<td>-0.30</td>
<td>-0.26</td>
<td>-0.12</td>
</tr>
<tr>
<td>b1 (Travel time)</td>
<td>-0.23</td>
<td>-0.08</td>
<td>-0.19</td>
</tr>
<tr>
<td>b2 (Travel time uncertainty)</td>
<td>-0.09</td>
<td>-0.09</td>
<td>-0.13</td>
</tr>
<tr>
<td>b3 (PT travel time)</td>
<td>-0.19</td>
<td>-0.24</td>
<td>-</td>
</tr>
<tr>
<td>b4 (PT season ticket)</td>
<td>18.33</td>
<td>16.74</td>
<td>-</td>
</tr>
<tr>
<td>C_p (PT constant)</td>
<td>-3.00</td>
<td>3.00</td>
<td>-</td>
</tr>
<tr>
<td>Mean VTTS in €/h</td>
<td>10.5 [1.6, 82.5]</td>
<td>8.3 [0.6, 77.2]</td>
<td>36 [1.6, 100]</td>
</tr>
<tr>
<td>Overall mean VTTS</td>
<td>12.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of users</td>
<td>60</td>
<td>30</td>
<td>10</td>
</tr>
</tbody>
</table>

Parameter values for the different trip purposes are reported in
Table 1. For random parameters the reported value corresponds to the mean of the draws used in simulation. The Johnson’s S_B distribution is censored in order to be able to calculate a robust welfare measure based on VTTS (which is the time parameter divided by the random cost parameter). The censoring is made at VTTS equal to 100 €/h, which corresponds to a cost parameter value of $-0.014$ for flexible trips, $-0.005$ for fixed trips and $-0.011$ for business trips. The simulated parameter values that are smaller than the threshold are set to the threshold. The mixed logit choice probabilities are simulated by calculating the logit formula for each draw and averaging the result (Train, 2003). From table 1 it is also clear that the flexible commuters and other trips is the largest segment and that business trips are just 10 percent of all trips.

Assignment

Contram takes as input a time-sliced origin-destination matrix (OD-matrix) and a network specification, assigns vehicles to the network in the form of packages and calculates the shortest path for each package by assigning them one by one to the network (Taylor, 2003). Iteration of assignment is needed since the shortest path and corresponding travel time of a package may be affected by subsequent packages travelling between other OD-pairs. The iteration process can be compared to a day-by-day learning of network conditions. The naïve user chooses the shortest route under free-flow conditions, which creates sever congestion on some routes. New routes are then chosen on the second day (second iteration), given the experienced travel times from day one. After a number of days the user equilibrium is reached. Contram uses deterministic assignment such that results are always the same given the same input and scenario settings.

Model validation

The Silvester model is calibrated in two steps. First, using the method of reverse engineering, where demand in each preferred departure time interval is adjusted such that Silvester produces correct traffic flows in the situation without congestion charges (Kristoffersson and Engelson, 2008). In the second step Silvester was calibrated to better reproduce the observed effects when charging was introduced (Kristoffersson, 2011).

Figure 3 shows validation of average traffic flow between 6.30-9.30 am at 59 links (red dots in the left picture) including the cordon links after the two-step calibration.

Figure 3: Validation of Silvester traffic flow against field measurements

The congestion charging schemes

The Stockholm congestion charges were first introduced as a trial 3 January – 31 July 2006, followed by a referendum in the City of Stockholm. The referendum was pushed through by parties against congestion charges but in the end a majority in the City of Stockholm voted for keeping the charges. The new Liberal-Conservative government reintroduced the charges in August 2007. However, the revenues from the permanent system were earmarked to a partially government-funded transport investment package including both road and public transport investments in the county.

The congestion charging scheme consists of a cordon around the inner city of Stockholm with time-differentiated charges. The congestion charge is a tax levied on certain vehicles for passages in and out of Stockholm’s inner city weekdays 6.30 am to 6.30 pm. During the trial, the traffic flows across the cordon were reduced by on average 22 percent during charged hours. The charge varies between €1 and €2 depending the time of day, with a maximum

---

7 About half of the neighbouring municipalities also had referendums.
daily charge per vehicle of €6. The charge does not apply overnight, at weekends and on public holidays or during the month of July.

Figure 4: The cordon around the inner city of Stockholm (dashed line), the bypass E4/E20 (solid line) and the location of the charging points (red dots). Source: Eliasson et al (2009).

The area inside the cordon is around 30km$^2$. The location of the cordon is depicted in

Figure 4. The dashed line is the charging cordon, the dots are charging points and the solid line is the non-charged bypass E4/E20 west of the inner city. There is no congestion charge
for journeys to and from the island of Lidingö (see Figure 4) which pass in and out of the charging zone within 30 minutes. Bypass E4/E20 is heavily congested but is not charged at present for political reasons. There is, however, political consensus for charges on the Bypass E4/E20 in 2020, when a new bypass further west is built. In this paper, we analyse primarily the modified scheme with charging levied also on the bypass E4/E20, because the welfare effect of this modified scheme is considerably higher.

Throughout this paper, we refer to the scheme in operation today as the “current scheme” and to the current scheme plus a charge levied on the bypass E4/E20 as the “modified scheme”. There are moderate route choice effects of the schemes, in particular in the modified scheme. In the current scheme, the main route choice effect is that drivers travelling between Northern and Southern Stockholm may choose to travel through the city centre or to divert to the bypass E4/E20 to avoid paying the charge. This effect is small in the current scheme; the traffic flow on the bypass was not affected when the charging scheme was introduced, abolished and reintroduced. The route choice effect would be even smaller in the modified scheme, since a congestion charge is then also levied on the bypass E4/E20.

The revenues from charges in 2008 amounted to approximately 85 M€ in 2008. The operating cost of the current cordon scheme is approximately 25 M€ per year.

**Method of analysis**

The social benefits of the Stockholm congestion charging scheme, and the modified scheme, have previously been analysed with the national forecast model Sampers (Engelson and van Amelsfort, 2011), which is static and has several common features with the standard textbook analysis: network effects from congestion charges are small because the blocking back of upstream links are not modelled, drivers are assumed to have a single VTTS (within trip each purpose) and scheduling is left out. The implications from the standard textbook analysis, that drivers are considerably worse off with congestion charges (before return of revenues), are also confirmed in the Sampers analysis. For these reasons, the Sampers analysis can be interpreted as an application of the standard textbook analysis of the Stockholm congestion charging schemes (even if network effects and value of time heterogeneity are included to some extent).
The analysis in Silvester extends the standard textbook and the Sampers analysis by including, or improving the representation of, network effects, heterogeneity in the value of time and the temporal dimension. The aim of the following analysis is to assess to what extent these effects add to the benefit of the modified scheme in Stockholm. This is accomplished by first calculating the consumer surplus (CS) of the scheme with the standard Silvester model. We then investigate how CS changes when the three effects extending the textbook analysis are taken out from the standard Silvester model successively in different steps. These steps are described in sub-section 0-0. When the three effects are taken out of Silvester, the model resembles the forecast model Sampers. We therefore also compare CS predicted by Silvester and by Sampers. We focus on the modified scheme in this analysis.

CS calculation

First, we compute the consumer surplus of the current and modified congestion charging schemes. The CS is defined as the difference between the logsums computed with and without the scheme. The logsum is given by Equation (4):

\[ L_{k\omega y} = \frac{1}{50} \sum_{d=1}^{50} \ln \left( \sum_{t=0}^{13} e^{\nu_{c}^{k\omega t d}} + e^{\nu_{p}^{k\omega}} \right) \]

(4)

\( \nu_{c}^{k\omega t d} \) and \( \nu_{p}^{k\omega} \) are the representative utility functions for a traveller departing at \( t \) with trip purpose \( k \), OD-pair \( \omega \), preferred departure time interval \( y \) and set of preferences \( d \), defined in Equation (3). For each \( \omega \) and \( y \), travellers are divided into 50 groups with different sets of preferences for \( SDE, SDL \) and \( M \). The 50 sets of preferences are constructed by combining (for each trip purpose \( k \)) 50 random draws from each of the random preferences \( \beta_{1}, \beta_{2} \) and \( \beta_{3} \). The draws are the same each time the model is applied.

When referring to CS in the rest of the paper it is assumed that no revenues are recycled.

Network effects

Next, we assess the benefits arising from network effects because travellers have different origins and destinations. The network effects are assessed by exploring to what extent benefits accrue to drivers in trip relations where charges do not apply. To identify trips in uncharged trip relations, all trips are categorized with respect to the origin and destination zone. The zones used for this categorization are depicted in
Figure 5, including Inner city (I), North zone (N), South zone (S) and Lidingö Island (L). The zones are chosen such that the OD-category indicates if the trips are charged or not. Out of the sixteen resulting OD-relation categories, eight are uncharged.
The trips to and from the city centre are charged (I-N; I-S; I-L; N-I; S-I; L-I). Trips from the North zone to the South zone are charged once or twice depending on the route choice (via bypass E4/E20 or inner City) in the modified scheme (N-S; S-N). Trips to and from Lidingö and the North or South zone are uncharged (N-L; S-L; L-N; L-S), as well as trips within zones (I-I; N-N; S-S; L-L).

Since the effect of route choice is limited in the modified scheme, the route choice effects would not be the main reason for the large benefits of congestion charging scheme and this is not analysed further in this paper.

**Heterogeneous VTTS**

Next, we explore to what extent the assumption about heterogeneity in VTTS adds to the benefit of the congestion charging scheme due to sorting of trips over modes and departure times with respect to VTTS. This is explored by comparing the CS computed by Equation (4) with the CS computed under the assumption that the VTTS is constant across trips. We carry out the latter computation by constraining the cost parameters ($\beta^k$) to be constant over all trips with the same trip purpose ($k$). The cost parameters for different trip purpose segment are chosen such that the VTTS is equal across purposes (the travel time parameter is not randomly distributed but varies across trip purpose segments). The cost parameters are chosen such that the model predicts the same share of traffic across the cordon diverting to public transport as in the charged situation in the standard Silvester model, which implies a VTTS of 4.8 €/h. We may think of this VTTS as that which would have been estimated based on traffic flows before and after the introduction of congestion charges. The logsum, with and without charges, is computed as described by:

---

8 Average VTTS is 12.5 €/h in Silvester and in Sampers VTTS is 9.5 €/h.
In Equation (5), the cost parameter is no longer specific for each group \( d \), but the scheduling parameters are still varying between different groups \( d \).

Efficiency gains arising from heterogeneity in VTTS will apply primarily to traffic in OD-relations where congestion charges are levied. Sorting of trips with respect to VTTS will not occur for uncharged trip relations, since these do not face the trade-off between money and time when making the choice of mode and departure time. For this reason, the benefits arising from heterogeneous VTTS are assessed only for the charged trip relations (I-N; I-S; I-L; N-I; S-I; L-I; N-S; S-N)\(^9\).

**Scheduling benefits**

In the next step we investigate to what extent the benefits predicted by the Silvester model arise because travellers can reschedule. Part of the scheduling benefit, however, is captured in the previous analysis step because they arise in combination with the heterogeneity in the cost parameter (or VTTS). The heterogeneity in the cost parameter sorts trips not only between modes but also in the temporal dimension, and the latter implies additional efficiency gains. In this analysis step we concentrate on the benefits over and above the rescheduling benefits arising due to heterogeneity in the cost parameter captured in the analysis of the previous section. For this reason the VTTS is held constant when further investigating the scheduling benefits.

To calculate the CS under the assumption that drivers cannot reschedule consistently, a new demand model, would have to be estimated. However, we approximate the outcome of such a constrained model by computing the CS using the parameters of the existing Silvester model, but constraining the model such that drivers cannot reschedule, only adjust mode and route, when the scheme is introduced.

For each group of travellers with identical preferences \( (d) \), OD-pair \( (\omega) \), trip purpose \( (k) \) and preferred departure time \( (y) \), the share of drivers departing in interval \( t \) in the uncharged situation, \( P^u(t|d, y, k, \omega) \), are in the charged situation constrained to go by car in the same departure time interval \( t \) or to switch to public transport. Hence, the possibility to change departure time and still drive when the charges are introduced is no longer available in the model. Drivers are thus constrained to the binary choice of driving (in departure time interval \( t \)) or divert to public transport. The share of drivers travelling with public transport in the situation without charges \( (P^u(p|d, y, k, \omega)) \) is assumed to remain on public transport (this share is small in the present population, as discussed in Section 0). Equation (6) shows the logsum computed in the scenario with charges. Note that the cost parameter \( (\beta_3^k) \) does not depend on \( d \), since what we evaluate here is the benefit of scheduling assuming constant VTTS.

\[
L_{k|\omega} = \frac{1}{50} \sum_{d=1}^{50} \left[ \ln \left( \sum_{t=0}^{13} e^{v^u_{ct|\omega} + \epsilon_{p|\omega}} \right) \right] 
\]

\[ (5) \]

---

\(^9\) In reality, the benefits for uncharged trips can obviously be affected by heterogeneity in the VTTS, if these effects change the total traffic flows. However, since we are assuming that the traffic flow remains unchanged when the VTTS changes from being randomly distributed to being constant, this is not the case in the present study.
\[ L^{kow} = \frac{1}{50} \sum_{d=1}^{50} \left[ \sum_{t=0}^{13} P^u(t|d, y, k, \omega) \ln \left( e^{\nu^{kady}_c} + e^{\nu^{kaw}_p} \right) + P^u(p|d, y, k, \omega) \ln \left( e^{\nu^{kaw}_p} \right) \right] \]  \hspace{1cm} (6)

For the situation without congestion charging the logsum is computed by Equation (5). The CS of the scheme (assuming homogeneous VTTS and no scheduling benefits) is computed as the difference between these two. We do not re-run the assignment model in this analysis, and make thus the assumption that the traffic flows (and travel times) in each time period remain the same when rescheduling is constrained. As will be shown, this assumption can be reasonably well justified.

**Results**

**CS of the schemes**

We find that the total CS (without recycling of revenues) calculated from Equation (4) is positive, both for the current scheme and the modified scheme. **Table 2** shows the total CS of the two schemes and the CS divided on trip purpose segments. **Table 3** shows CS per trip for all trips and for each trip purpose segment.

**Table 2: Resulting CS forecasted by Silvester**

<table>
<thead>
<tr>
<th>CS, all trips (M€/year)</th>
<th>Total</th>
<th>Flexible</th>
<th>Fixed</th>
<th>Business</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current scheme</td>
<td>13</td>
<td>-2</td>
<td>-5</td>
<td>20</td>
</tr>
<tr>
<td>Modified scheme</td>
<td>32</td>
<td>4</td>
<td>-5</td>
<td>33</td>
</tr>
</tbody>
</table>

Business travellers gain most from the charging scheme although these make up only ten percent of the trips. Fixed schedule travellers lose most and flexible schedule travellers are in between. This is consistent with the finding of van den Berg and Verhoef (2011), who show that it is the drivers with intermediate value of schedule delay (VSD) and low VTTS for this VSD that incur the greatest losses. Drivers with fixed working schedule have a VSD that is on average higher than travellers with flexible schedule and lower than business travellers; they have also the lowest average VTTS. Furthermore, **Table 3** shows that the benefit is considerably larger for the modified scheme. With this scheme, only travellers with fixed schedule become as a group worse off with the charges.

While these numbers show some large distributional effects, it is worth pointing out that it is not necessarily the business travellers who are the winners, but rather their employers. This would have a positive effect on the efficiency of the economy (see for instance Anderstig et al. (2012)) and therefore gain the entire society.

**Table 3: CS per trip**

<table>
<thead>
<tr>
<th>CS, all trips (€)</th>
<th>Average</th>
<th>Flexible</th>
<th>Fixed</th>
<th>Business</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current scheme</td>
<td>0,04</td>
<td>-0,01</td>
<td>-0,06</td>
<td>0,67</td>
</tr>
<tr>
<td>Modified scheme</td>
<td>0,11</td>
<td>0,03</td>
<td>-0,06</td>
<td>1,11</td>
</tr>
</tbody>
</table>

For comparison, CS of the two schemes computed with the rule of a half and based on Sampers forecasts can be found in **Table 4**. The Sampers forecast gives, as mentioned before, a large negative CS. The implications from the standard textbook analysis, that drivers are worse off without return of revenues, are thus confirmed. In fact, the revenues in the Sampers forecast barely balance the large negative consumer surplus.
Table 4: Resulting CS based on Silvester and Sampers forecasts\textsuperscript{10}.

<table>
<thead>
<tr>
<th>CS, all trips (M€/year)</th>
<th>Silvester</th>
<th>Sampers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current scheme</td>
<td>13</td>
<td>-76</td>
</tr>
<tr>
<td>Modified scheme</td>
<td>32</td>
<td>-97</td>
</tr>
</tbody>
</table>

In the following we concentrate on the modified scheme. We refer to the CS of the modified scheme computed with the standard Silvester model, 32 M€/year, as the base case CS.

Network effects

Table 5 shows the resulting CS arising for the sixteen different trip relations. The largest benefits arise in the uncharged trip relations: inside the cordon (I-I), within the North zone (N-N) and within the South zone (S-S), with benefits of 9.0, 14.0 and 5.2 M€/year, respectively. Travellers in the uncharged relations connecting North and South to Lidingö (N-L, L-N, S-L, L-S) also benefit 3.9 M€/year. Note also that travellers going from the south (S) to inner city (I) or North (N) benefit as a group although they are charged. It is striking that these are the relations with highest congestion levels in the situation without charging. This suggests that benefits of congestion charging increase if initial congestion levels are high.

Table 5: CS divided on charged and uncharged trip relations

<table>
<thead>
<tr>
<th>Origin</th>
<th>Destination</th>
<th>Inner city (I)</th>
<th>North (N)</th>
<th>South (S)</th>
<th>Lidingö (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inner city (I)</td>
<td>Consumer surplus</td>
<td>9.0</td>
<td>-4.0</td>
<td>-8.5</td>
<td>-0.4</td>
</tr>
<tr>
<td>North (N)</td>
<td>all trips = 31.8 M€/year</td>
<td>-0.5</td>
<td>14.0</td>
<td>-3.3</td>
<td>1.8</td>
</tr>
<tr>
<td>South (S)</td>
<td></td>
<td>7.7</td>
<td>11.1</td>
<td>5.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Lidingö (L)</td>
<td></td>
<td>-2.2</td>
<td>0.6</td>
<td>0.3</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 6 compares the base case CS (from Table 2 and Table 3) with the CS of the charged relations. The left chart shows total CS and the right shows average CS. In the charged relations, CS is negative with a total of −0.3 M€/year.

\textsuperscript{10} Sampers results are taken from Engelson and van Amelsfort (2011).
Only business trips still have on average a positive CS when looking solely at charged relations. Trips with fixed schedule still incur the greatest loss both in total and per trip. The difference in CS between business trips and trips with fixed schedule decreases from 1.17 to 0.85 €/trip, but the difference is still substantial. The reason for the larger difference when uncharged relations are included is that business trips have a higher average VTTS, which leads to a higher valuation of the travel time savings in uncharged relations.

Total CS decreases from a large benefit (32 M€/year) to a small loss (−0.3 M€/year) when disregarding the uncharged trip relations. The network effects, arising because there are many OD-pairs and extensive blocking back of upstream links building up from bottlenecks on the cordon in the uncharged case, is thus a major reason behind the positive CS.

Interestingly, network effects explains to a large extent also the additional benefits of the modified scheme compared to the current scheme (32 compared to 13 M€/year). This is because the extensive queues that currently build up upstream the congested bypass E4/E20, reaching far out in the network, reduces when charges are levied also on the bypass E4/E20.

In the following two steps in the analysis we leave out the uncharged trips, since only these trips face a trade-off between money and time when making the choice of mode and departure time, as explained in section 5.3.

Heterogeneous VTTS
This section assesses the benefits arising because of heterogeneous VTTS. As shown in the previous section, the total CS is −0.3 M€/year for charged trip relations. Figure 7 shows that CS declines to −67 M€/year for these trips when VTTS is held constant across trips. Hence, the heterogeneity in the VTTS adds more to the total CS than the network effects analysed in the previous section. The large decrease in CS is in accordance with the findings of Verhoef and Small (2004), who note that the benefits of second-best pricing may be dramatically underestimated if heterogeneity in VTTS is not taken into account.

The CS declines for all trip purposes, but mostly for business trips. The CS declines most for business trips because they have, on average, the highest VTTS in the base case. In fact, when all trip purposes have the same VTTS, the average CS per trip are similar between the trip purposes. The small differences in average CS that remain arise because of differences in scheduling and travel time sensitivity. On average trips with fixed schedule incur the greatest loss because they have the highest scheduling costs (compared to time and cost sensitivity). Trips with flexible schedule now lose most as a group, but this is simply because this is the largest group of travellers.

Figure 7: CS with the standard Silvester model for all trips (black), charged trip relations (chequered) and charged trip relations with constant VTTS (striped).

Scheduling effects
In addition to constant VTTS, we now constrain the drivers travelling in time period \( t \) in the situation without charges to the binary choice of driving (in departure time period \( t \)) or divert to public transport. The CS for the situation with charges is now computed by Equation (6).

The benefits arising from scheduling flexibility are relatively small, 10 M€/year, in the situation with constant VTTS. The effect is largest for business trips. Interestingly, taking out the possibility for drivers to reschedule when still assuming heterogeneous VTTS as in the standard Silvester model, the benefit reduces 36 M€/year. Hence, a large benefit of the heterogeneity in VTTS arises in combination with the possibility for the drivers to reschedule. The benefit of rescheduling adjustments thus increase substantially when the cost parameter (or VTTS) is heterogeneous.

Figure 8 shows that simultaneously taking out the possibility for drivers to reschedule, the network effects and heterogeneity VTTS, reduces the base case CS from +32 M€/year to −77 M€/year. A CS of −77 M€/year is in the same magnitude as the CS calculated by Sampers (−97 M€/year). The higher benefits in this constrained Silvester analysis, compared to the Sampers analysis is mostly due to the fact that the static model predicts smaller travel time gains for charged traffic.

We may compare these numbers to the benefits of the current scheme. For the current scheme the base case CS is 13 M€/year. The combined effect of heterogeneity in VTTS and rescheduling adjustments induce a benefit of 51 M€/year and network effects induce a benefit of 22 M€/year. Hence, simultaneously taking out the possibility for drivers to reschedule, the network effects and heterogeneity VTTS, reduces total CS for all trips from +13 M€/year to −60 M€/year. A CS of −60 M€/year is also in the same magnitude but a bit smaller than the CS calculated by Sampers (−76 M€/year).
In this analysis travel times are not re-calculated by the assignment model and we end this section by justifying this simplification. We thus assume that traffic flow, and therefore the travel times, in each 15 minutes period remain approximately unchanged when rescheduling adjustments are taken out of the model. To test this assumption we compare the number of departures in each 15 minutes period calculated with and without rescheduling adjustments included in the model. The result is shown in Figure 9, which shows the number of departures in the situation without charging.

First, note that on the aggregate level, the time varying congestion charges have a limited effect on the shape of the flow profile in the standard Silvester model, contrary to the expectation. This is consistent with the revealed flow profiles before and after the introduction of the Stockholm congestion charges, showing an almost negligible effect on the shape of the flow profile; merely shifting the flow curve down. This does not necessarily
imply that the possibility to reschedule is not important at the individual level; on the contrary, the analysis in this section indicates a substantial benefit from scheduling effects when the VTTS are heterogeneous and implies that some drivers move from and others to the peak. Hence, even if the aggregate flow profile does not demonstrate any significant effect on departure time choice, there second order effects increase the benefits of rescheduling.

The key point of interest here is that the departure time profiles calculated with and without rescheduling adjustments included in the model are small, which is consistent with the limited effect on the aggregate time profile in the standard Silvester model when charges are introduced. The difference in number of departures in each 15 minutes period calculated with (dashed) and without (dotted) rescheduling adjustments included in the model is at most 3.6 per cent. These differences in traffic flow would have some, but a minor effect on travel times, and is of the same magnitude as the day-to-day variation. Hence, the assumption that travel times remain approximately the same when scheduling is taken out of the model would not have any major impact on the result of the analysis.

Figure 9: Number of departures in each time period.

Conclusions

The standard static textbook analysis of congestion charges implies that drivers as a group will be worse off with congestion charges if they are not compensated with return of revenues. This result is confirmed when evaluating the two different schemes, the current and a modified version of the Stockholm congestion charging scheme, with the static national forecast model Sampers. In fact, the revenues barely balance the consumer surplus calculated with Sampers. The Sampers model has several common features with the standard textbook analysis: network effects from congestion charges are small because the blocking back of upstream links are not modelled, users are assumed to have homogeneous values of time within the same trip purpose and scheduling is left out.

When analysing the two schemes applying the dynamic model Silvester, estimated and calibrated for Stockholm and including scheduling and mode choice, heterogeneous users and dynamic traffic assignment, we find, however, a positive benefit for drivers even without return of revenues for both the current and modified charging scheme. We find the
consumer benefits of the Stockholm charges to be 13 £/year. This benefit would, however, increase substantially to 32 £/year if the scheme was modified to charge also the heavily congested E4 bypass.

The key factors explaining the difference in calculated consumer surplus between the dynamic and static models are:

A) Sorting of trips between modes, routes and departure times according to heterogeneous VTTS adds 77 £/years to the benefit in the modified congestion charging scheme we analyse.

B) Benefits for drivers travelling in non-charged OD-pairs arising from less blocking back of upstream links and intersections, where queues used to build up from the charged bottleneck links adds another 32 £/years to the benefit.

Taking out these effects from Silvester gives a consumer surplus of the same magnitude as calculated by Sampers. The combination of these factors is thus crucial for an accurate evaluation of benefits of a congestion charging scheme and can change the sign of welfare estimates. This result provides a strong warning for estimating consumer benefits of congestion charges with inappropriate models in real world settings.

The result of this paper is also relevant for the recent theoretical literature rising concerns that congestion charging, in the presence of distortive income taxation, induces negative effects on the labour market, because congestion charges reduce accessibility (Parry and Bento, 2001; Van Dender, 2003; Pilegaard and Fosgerau, 2008; De Borger, 2009). The insight that a substantial share of drivers gain directly from congestion charges suggests that congestion charging has a positive effect on accessibility and would thus not induce negative effects on the labour market.

The direct benefits for many drivers indicated by this study could be one factor explaining the high public support for the congestion charges in Stockholm, which has even increased since the congestion charging scheme was introduced in 2006. This finding could help to increase the low political and public acceptability in many urban areas that currently consider introducing congestion charging.

References


