COMPARISON OF USAGE-BASED CONGESTION PRICING SCHEMES

Joakim Ekström, Corresponding Author
Department of Science and Technology, Linköping University
Bredgatan 33, Norrköping, SE-601 74
Tel: +46 11 363011 Fax: +46 11 363270; Email: joakim.ekstrom@liu.se

Word count: Text (6042) + 6 tables/figure (6 x 250) = 7542

Submission date: 2018-10-29
This paper investigates how usage-based pricing can be applied in order to improve the efficiency of the transportation system. Distance-based, travel time-based and delayed-based pricing schemes are investigated, both from a one-link and network perspective. A static modeling framework is adopted, based on user-equilibrium, and a simplified emission model based on the HBEFA emission model is used. Optimal usage-based fees are calculated using a surrogate-based optimization framework, and determination of zone layout is done using k-means clustering. The results for a Stockholm region network show that already with network wide pricing a large proportion of the maximum improvement of social surplus achievable with first-best pricing, can be reached. Delay-based pricing, in which the users are charged based on the excess travel time in comparison to free flow travel times, achieves 99% of the maximum improvement, already with one single fee across the whole network. With clustering of links based on their first-best optimal fees, the benefits of both distance and travel time-based fees can significantly be improved.
INTRODUCTION

Reduced traffic congestion and traffic related greenhouse gas emissions, and improved air quality and city life are important current themes within the transportation research area. Congestion pricing, in various forms, has over the year’s attained interest, as a key approach for reducing congestion through presenting the drivers with economic incentives so that they can make decisions based on the full cost associated with their journey. Thus, improving the overall efficiency of the transportation system, reducing greenhouse gas emission and other traffic related pollutants. Yet, few cities have adopted congestion pricing, and among the successful examples are Singapore, London and Stockholm.

In theory, a first-best pricing scheme can be achieved by letting the road users pay for the additional external costs that their trip is generating for the society, e.g. delay to other road users, emissions, noise, accidents and road wear. In practice, such a pricing principle would be difficult to implement, as it requires congestion dependent fees (tolls), differentiated for each road segment. In addition, for a pricing scheme to be successful, the structure of the scheme needs to be clear to the users (1). Thus, one often rely on so-called second-best pricing schemes, which are based on optimal prices given some additional restrictions on fee levels and structure. Commonly applied fee structures are cordon-based pricing, in which the users are paying a toll when passing specific toll points, and area-based tolling, in which a fee is paid per day to allow for driving within a specific area. A third type of fee structure, which has attained less attention, is usage-based pricing, in which the drivers are continuously charged based on their use of the infrastructure. This can be achieved through distance-based and time-based pricing.

In one of the early studies on usage-based pricing, May and Milne (2) evaluate network efficiency under cordon, distance, travel time and delay based pricing. Their conclusion are that, for their case study, cordon pricing is the least effective price mechanism. There are mainly two attributes of usage-based pricing which makes them less appealing in practice: the ability of road users to accurately estimate the total charge for a trip, and the technological requirements for implementing such fee structures. To be able to accurately determine the total fee for a trip is important in order for the pricing scheme to have the desired effect (3). But as is pointed out by Bonsall et al. (1), it is difficult for drivers to correctly estimate distances. In terms of implementing usage-based pricing schemes, de Palma and Lindsey (3) list several studies on technological feasibility of distance-based pricing schemes, and in Europe several countries have introduced distance-based pricing for heavy goods vehicles. Assuming that the technological barrier for implementing and enforcing usage-based pricing schemes can be overcome, there is still the difficulty for the users to determine the total charge of a trip. With the development of smartphone applications for navigating in congested cities, the prediction of usage-based charges for a specific trip has become realistic, and with introduction of connected and partly automated vehicles, such pricing schemes are becoming even more attractive.

There are several studies considering mainly distance-based pricing as a usage-based pricing scheme. In (4) a distance-based (denoted km-based) pricing scheme for the Netherlands is outlined, and in (5) and (6) distance-based pricing scheme for reducing congestion on freeways and congestion zones, respectively, are presented. A combination of a distance and a travel time-based pricing scheme for Dublin is presented in (7) and user response is evaluated. Further, in (8) a travel time-based pricing scheme for Paris is evaluated. Although the benefits of usage-based pricing is demonstrated in several publication, there are no studies on methodologies for defining zone structures to be used with such pricing schemes. There are, however, several studies on optimal usage-based pricing, e.g. (9), (10) and (11), which present methodological approaches to optimal distance-based pricing, and (12) which considers the formulation of optimal joint distance and travel time-based fees.

For usage-based pricing schemes, there are mainly three key design features for the pricing mechanism; fee levels, differentiation between roads/zones/areas, and differentiation in time. This paper investigate the first two of these features, adopting a static user equilibrium model as basis for the analysis. The paper contributes with an investigation of the socio-economic efficiency of different usage-based pricing principles, and compares usage-based pricing with first-best based pricing schemes. Further, zone designs, for the introduction of differentiated usage-based charges, is discussed, and numerical results are presented for a Stockholm network model.
The remainder of the paper is outlined as follows. First, the modeling of the road users’ response to road pricing, and the modeling of environmental effects are presented, followed by a presentation of the evaluation criteria. Next, the three different usage-based fees are analyzed from a first-best pricing perspective, and evaluated from a second-best perspective in a Stockholm network. Finally, the results are summarized in a concluding section, and further research is discussed.

**EVALUATION OF PRICING SCHEMES**

This section presents the modeling of the road users’ response to a given pricing scheme, the modeling of environmental external effects, and the social surplus evaluation measure.

**The user equilibrium model**

A static macroscopic modeling approach will be adopted for predicting the road users’ response to different pricing schemes. In macroscopic models, the travelers’ choices and resulting traffic conditions are described at an aggregated level, in terms of mean demand, flows and travel costs.

The model formulation is based on the standard user equilibrium model, see e.g. (13), with the network described by a set of links $A$, each link $a$ with flow $v_a$ in the unit of vehicles per hour, and a set of origin destination (OD) pairs $I$. For each OD pair $i$ the total demand $q_i$ is distributed over a set of routes $\Pi_i$, each route $p \in \Pi_i$ with flow $f_p$ in the unit of travelers per hour. Note that the route flow is given in terms of travelers per hour, and link flows as vehicles per hour. To convert between the two the car occupancy factor $\beta$ is used. The relationship between route and link flows for a link $a$ given by

$$v_a = \frac{1}{\beta} \sum_{i \in I} \sum_{p \in \Pi_i} f_p \delta^a_p,$$

where $\delta^a_p$ is the link-route incident matrix, which each element taking on the value of 1 if route $p$ traverses link $a$, and 0 otherwise.

For each link $A$ there is a volume delay function, which gives the link travel time, in minutes, as function of link flow, and in this paper the function has the form

$$t_a(v_a) = \alpha_a l_a (1 + \left(\frac{v_a}{k_a n_a}\right)^\gamma), \quad (1)$$

where $\alpha_a, k_a$ and $\gamma_a$ are link type specific parameters (given in Table 1), $l_a$ is the link length, and $n_a$ is the number of lanes. This functional form is similar to the US Bureau of Public Roads (BPR) function and to the Swedish TU71 function (14). In practice, the parameters of the travel time function are not link specific, but link class specific, and here the seven link classes for which the TU71 function is defined will be used (presented in Table 1). Also, note that travel times are defined for flow above capacity. Thus, for flow below capacity we interpret the increase in travel time as link delay, while the additional delay for flow above capacity is interpreted as additional delay due to spillback effects in the traffic network. For a thorough description of the role of travel time functions in user equilibrium models, the reader is referred to (14) and (15).

The travel cost function $c_a(v_a)$, gives the link generalized travel cost, and can include components related to both travel time and monetary costs. The value of time (VoT) is used for converting travel time to a monetary cost, and vice-versa. In reality, the VoT is perceived differently by individual travelers, but for the travel cost functions used in this model a mean value across the population is used.

In this paper, three different types of usage-based fees are included in the travel cost function:

- $\tau^{(d)}$, a distance-based fee, charged per kilometer drive, with actual link fee equal to $\tau^{(d)} l_a$.
- $\tau^{(t)}$, a travel time-based fee, charged per minute driven, with actual link fee equal to $\tau^{(t)} t_a(v_a)$.
\[ \tau^{(e)}, \text{a delay-based fee, charged per minute excess travel time in comparison to the free} \]
\[ \text{flow travel time, with actual link fee equal to } \tau_a^{(e)}(t_a(v_a) - t_a(0)). \]

Here it is assumed that only non-negative fees are considered, i.e., no subsidies. Note that the travel time and delay-based fees are flow dependent, and may thus be more difficult for the travelers to accurately predict.

The travel cost function can now be expressed as
\[ c_a(v_a) = \theta t_a(v_a) + \tau_a^{(d)} l_a + \tau_a^{(t)} t_a(v_a) + \tau_a^{(e)} (t_a(v_a) - t_a(0)), \]

where \( \theta \) is the VoT. The user equilibrium conditions state that no user in OD pair \( i \) will travel on a route with a travel cost higher than the minimum travel cost. In OD pair \( i \) the minimum travel cost, \( \pi_i \), will equal the cost of traveling, given by the inverse of the demand function \( D^{-1}_i(q_i) \).

The travel demand function then becomes
\[ \pi_i = D^{-1}_i(q_i) = \pi_i^0 + \frac{1}{\alpha} \ln \frac{A_i}{T_i - A_i} + \ln \left( \frac{T_i}{q_i} - 1 \right), \]

where \( \alpha \) is the dispersion parameter in the multinomial logit model, \( T_i \) is the total demand for traveling in OD-pair \( I \), \( A_i \) is the car travel demand in the reference scenario and \( \pi_i^0 \) is the car travel cost in OD-pair \( i \), in the reference scenario.

The user equilibrium problem can now be formulated as
\[
\begin{align*}
\min_{v_a, q_i} & \beta \sum_{a \in A} \int_0^{v_a} c_a(x)dx - \sum_{i \in I} \int_0^{q_i} D^{-1}_i(w)dw \\
\text{subject to} & \sum_{p \in \Pi_i} f_p = q_i, \forall i \in I \\
& v_a = \frac{1}{\beta} \sum_{i \in I} \sum_{p \in \Pi_i} f_p \delta_p^a, \forall a \in A \\
& q_i \geq 0, \forall i \in I \\
& f_p \geq 0, \forall i \in I, p \in \Pi_i.
\end{align*}
\]

The model is solved using a partial linearization method \((18)\), in which fixed demand equilibrium problems are solved repeatedly, updating the demand in between.

**Modeling external environmental effects**

Exhaust emissions of carbon dioxide will be considered as the main source of external environmental effects related to traffic congestion. Carbon dioxide emissions will be modeled based on the output from the macroscopic traffic model described in previous section, i.e., based on mean travel times and flows. For such aggregated data, mean speed and traffic situation based emission models are suitable. Emission models applied on such an aggregated level are based on emission factors, which gives emissions of a specific pollutant in the unit of grams per vehicle kilometer. Emission factors are based on detailed simulation of emissions for specific driving cycles, aiming to mimic driving behavior for the specific studied traffic situation. By multiplying the emission factor with the road length and associated flow per time unit (commonly hour) for the same road, total emissions for a specific time period can be computed. In Europe, COPERT 4 \((19)\) and HBEFA \((20)\) are commonly applied models. The two models mainly differ in how emission factors are determined. In COPERT 4 emission factors are determined based on mean speed and in HBEFA it is the combination of road type and traffic situation that determine emission factors. Although
Ekström 6

the definition of traffic situations are rather qualitative in HBEFA, it is usually required to introduce
quantitative description of the traffic situations. Such description can include both mean speeds and
flow/capacity ratios. Here we will proceed with the HBEFA traffic situation based model, since it is
available and calibrated for Swedish conditions. Our approach is, however, not limited to HBEFA, and can
be applied to COPERT 4 as well as other aggregated emission models. It can also be noted that for passenger
cars following the EURO I and higher legislative classes, both models inherit their emission factors from
the Artemis project (21).

HBEFA emission factors correspond to a specific traffic situation for a specific road type
and for a specific vehicle category. It is usually applied with flow or speeds threshold that are used for
determining each traffic situation. As noted in (22) this application of HBEFA does not take spillback
effects into account, but treat emissions as if they are solely connected to traffic activities within the specific
link segment, which is not true when flow exceeds demand. In (23) it is further illustrated that this may
underestimate total emissions.

Here we derive a total carbon dioxide emission function from HBEFA emission factors. For each
traffic situation and road type, the corresponding weighted (across the different vehicle categories) mean
speed for each driving cycle is known. By combining the volume delay function for each road type with
the mean speed of the driving cycles, we can compute the corresponding flow for each traffic situation.
Based on this flow the total emissions can be calculated. A total emission function is then derived for each
road category, by fitting the \( \eta \) and \( \sigma \) parameters in a function of the form

\[
E_a(v_a) = v_a l_a \left( \eta_a + \sigma_a \left( \frac{v_a}{k_a \eta_a} \right)^\gamma_a \right),
\]

(3)
to the total emissions at flows corresponding to mean speeds of the different driving cycles. Note that the
functional form is similar to the total travel time function (volume delay function multiplied with flow)
when the volume delay function is given by equation (1). The \( \gamma \) and \( k \) parameters will be the same in both
functions, in order to have a similar growth in emissions and travel time when flow exceeds capacity. Up
to capacity, the total emission function will provide an interpolated version of the HBEFA emission factors,
thus providing a continuous growing function, in comparison to the original HBEFA emission factors,
which are given as step functions. To determine the most suitable function form and parameter settings are
beyond the scope of this paper, and should be investigated further. In Table 1, the parameters used within
this work, for the different road categories are presented.

<table>
<thead>
<tr>
<th>TU71 road class</th>
<th>Speed limit (km/h)</th>
<th>( \alpha )</th>
<th>( \gamma )</th>
<th>( k )</th>
<th>( \eta )</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>50</td>
<td>1.277</td>
<td>6</td>
<td>1162</td>
<td>0.178</td>
<td>0.064</td>
</tr>
<tr>
<td>51</td>
<td>50</td>
<td>1.428</td>
<td>4</td>
<td>928</td>
<td>0.197</td>
<td>0.064</td>
</tr>
<tr>
<td>52</td>
<td>50</td>
<td>2.000</td>
<td>3</td>
<td>802</td>
<td>0.209</td>
<td>0.101</td>
</tr>
<tr>
<td>70</td>
<td>70</td>
<td>0.895</td>
<td>4</td>
<td>1700</td>
<td>0.142</td>
<td>0.047</td>
</tr>
<tr>
<td>71</td>
<td>70</td>
<td>0.985</td>
<td>4</td>
<td>1400</td>
<td>0.143</td>
<td>0.056</td>
</tr>
<tr>
<td>90</td>
<td>90</td>
<td>0.713</td>
<td>7</td>
<td>2091</td>
<td>0.138</td>
<td>0.039</td>
</tr>
<tr>
<td>100</td>
<td>110</td>
<td>0.593</td>
<td>8</td>
<td>2091</td>
<td>0.133</td>
<td>0.032</td>
</tr>
</tbody>
</table>

The social surplus measure
The efficiency of a pricing scheme will be evaluated by the well-adopted social surplus measure, which
reflects both the changes in consumer surplus, budget revenues and social costs. The social surplus is given
by the consumer surplus, \( CS \), minus costs of additional external effects, \( E \), plus the total revenues, \( R \). Here
it is assumed that all benefits and costs are given in the unit of SEK\(^1\), and in practice, this means that travel
time and additional external effects must be transformed into monetary values. The VoT is in this study set
to 1.2 SEK per minute. Valuation of carbon dioxide emissions is still an open question within the research
community, and here the valuation used for cost-benefit analyses of infrastructure projects in Sweden will
be used. This valuation is equal to the Swedish carbon dioxide tax of 1.14 SEK (24).

Further, the no-toll scenario will serve as reference scenario, and \(\Delta\) indicates the changes between
current and reference scenario. The change is social surplus can then be formulated as (25)

\[
\Delta S = \Delta CS - \Delta E + \Delta R.
\]  

The consumer surplus is given by the difference between the user benefits and user costs. In this paper, the
demand is modeled through a pivot point version of the multinomial logit model, and the change in
consumer surplus is then given by the logsum (17)

\[
\Delta CS = \frac{1}{\alpha} \sum_{\alpha \in \ell} T_i \ln \left( \frac{A_i}{\bar{T}_i} e^{\alpha (\pi_i - \pi_i^0)} + \frac{T_i - A_i}{\bar{T}_i} \right).
\]

Note that the change in consumer surplus, external effects and revenues are implicit functions of the
congestion charging fees, and for each set of fees, the user equilibrium model needs to be solved to
determine corresponding flows and link costs.

**ANALYSIS OF USAGE-BASED FEES**

It is well known that by charging the road users with a fee equal to the marginal cost increase in congestion
externalities, they will make decisions based on the full, social, cost of their action, resulting in a system-
optimal distribution of road users in the traffic network, see e.g. (26) and (27). Common costs included in
congestion externalities of road traffic are travel time for the fellow road users and environmental cost
related to vehicle emissions.

In this study we limit the congestion externalities to travel time and carbon dioxide pollution, and
the marginal increase in externality costs, \(MC\), can for a specific link, with flow \(v_a\), be expressed as

\[
MC = \theta v_a \frac{d}{dv_a} t_a(v_a) + \rho \frac{d}{dv_a} E_a(v_a),
\]

where \(\rho\) is the valuation of carbon dioxide emissions.

Setting a link fee (toll) equal to \(MC\) results in a first best pricing scheme with a system optimal
distribution of travelers in the traffic network, maximizing the social surplus (4). Such tolls would be
difficult to implement and would result in a pricing scheme in which it is difficult for the road users to
predict their travel cost.

Usage-based pricing, in terms of distance and travel time based fees, are more appealing as it is
easier for road users to relate to distances and travel time, rather than marginal cost curves. Especially with
the help of travel planners, which are becoming widely used in smartphones.

For a travel cost function and a total emission function given by equations (2) and (3) respectively,
the marginal cost takes the form

\[
MC = l_a \left( \theta \sigma_a \gamma a \left( \frac{v_a}{k_a} \right)^{\gamma a} + \rho \sigma_a \frac{v_a}{k_a} \left( \frac{v_a}{k_a} \right)^{\gamma a - 1} \right).
\]

\(^1\) The exchange rate between SEK and USD is 8.37 SEK for 1 USD (102817).
Let $v^*_a$ be the system optimal flow under marginal social cost (MSC) pricing. Assuming usage-based fees, differentiated on link level (each link a zone), the optimal distance-based fee can then be computed as

$$
\tau^{(d)} = \theta \alpha_a y_a \left( \frac{v^*_a}{k_a} \right)^{\gamma_a} + \rho \beta_a \left( \frac{v^*_a}{k_a} \right)^{-1},
$$

and optimal time-based fee as

$$
\tau^{(t)} = \frac{l_a \left( \theta \alpha_a y_a \left( \frac{v^*_a}{k_a} \right)^{\gamma_a} + \rho \beta_a \left( \frac{v^*_a}{k_a} \right)^{-1} \right)}{t_a(v^*_a)}.
$$

Note that both the optimal distance and time-based tolls are independent of the link length (as $t_a(v^*_a)$ includes $l_a$). This is an important feature, as it means that links of similar characteristics (here defined by belonging to the same road class), for the same flow, will have the same optimal usage-based fee.

In Figure 1, the travel time function and marginal cost curves for travel time, carbon dioxide, and total external costs are illustrated for road class 52 (see Table 1). Also, the optimal distance ($\tau^{(d)}$) and travel time ($\tau^{(t)}$) based fees are included, assuming a system optimal flow equal to 800 vehicles per hour, as functions of flow. Now assume that the usage-based fees are set to erroneous values. Then it is obvious that the difference between the travel time-based fee and the total MSC curve is always smaller than the difference between the distance-based fee and the total MSC curve. Thus, a time-based fee, offset from the actual MSC price will be more similar to the real MSC compared to a distance-based fee, under the assumption that the time and distance-based fees are determined on similar grounds.

Now consider the travel time part of the total MSC curve, $MC_t = \theta \alpha_a y_a \left( \frac{v_a}{k_a} \right)^{\gamma_a}$. Its form is similar to that of the travel time function, $t_a(v_a) = \alpha_a l_a \left( 1 + \left( \frac{v_a}{k_a} \right)^{\gamma_a} \right)$. May and Milne (2), use a delay-based fee, which correspond to the excess travel time to the free flow travel time. They further conclude
that the delay-based fee is the most effective one in comparison to distance and time-based fees, but provide
no further analysis.

A delay based fee, $x_a$, would correspond to a total congestion charge, $x_a \alpha_a l_a \left( \frac{v_a}{k_a} \right)^{\gamma_a}$. By setting
$x_a = \theta \gamma_a$, it is possible to fully recreate travel time part of the MSC curve. Note that this fee would not be
depending on either link length or capacity. Thus, the fees will be equal within each link class. While this
would only account for the travel time part of the MSC curve, it can be approximated with a factor $h$, with
the combined fee computed as

$$\tau = h \theta \gamma_a t_a(v_a).$$

This analysis is under the assumption of specific forms of the travel time and total emission functions.
Nevertheless, although functional form can differ, many travel time functions have a similar structure.

**NETWORK EXAMPLE**

To evaluate the three different usage-based fee structures, a traffic network of the Stockholm region in
Sweden will be used. The network model of Stockholm region is an aggregated representation of the traffic
system of the Stockholm region, with 392 links and 40 zones (1 560 OD-pairs). The demand model
describes the morning rush hour, including the choice between private car and public transport, and it is
roughly calibrated for the year 2006. Therefore, the model should not be used for actual policy evaluation
purposes for the Stockholm region, but for a comparison between the different usage-based fee structures
it is believed to be a good representation of a congested urban area, including the complexity of travel
choices related to both route and mode. In addition, this network model has been used in several evaluations
of road toll optimization methods (28), (29) and (30), and a detailed description of the model is given in
(28), including demand model parameters.

**Scenario setup**

Three different zone design principles will be evaluated and compared with first-best MSC pricing.
- Network wide pricing, in which one single fee is applied for all links in the network.
- Clustering based on spatial similarities.
- Clustering based on MSC pricing similarities.

In the previous section, it is shown that links with similar characteristics will have the similar usage-
based charges. There are two determinants for what is similar characteristics: road class and level of
congestion. When clustering based on spatial similarities, it is assumed that such characteristics is related
to the link location as well, and the network is divided into zones based on proximity to the city center and
road class, e.g. it is assumed that links close to each others have similar road class and level of congestion.
Here proximity to the central business district and road class are used for determining three
different zones, illustrated in Figure 2. The three zones can be described as (1) the central business district
(CBD) within the current Stockholm cordon, (2) the E4 bypass highway through the city, and (3) all other
remaining links (OD connectors excluded).

By finding links with similar MSC prices, either per kilometer (for distance-based fees) or per
minute (for time-based fees), we can determine clusters of links that have small variation in MSC prices,
compared to the variation between all links. The assumption here is that such clusters is a good basis for
zone layout. Here the $k$-means clustering method (31) is adopted, which determines the $k$ cluster which
minimize the within zone variation of MSC prices and maximize the between zones variations of MSC
prices. This will, however, not result in spatially connected links, but still is interesting to investigate as it
gives understanding to the importance of MSC price variations within a zone. For both distance-based and
travel-time-based fees, alternative clustering ranging from 2 up to 5 clusters will be evaluated, and single
fee structures, as well as combined structures will be included in the analysis.
As is later shown in the results, delay-based pricing is highly efficient already with network wide pricing, and this fee structure is thus excluded from the clustering-based analysis, as division into zones will have a very limited effect on the improvement in social surplus.

Determining optimal usage-based fees

Given a specified zone structure, there is still the problem of determining the zone fees that maximize the social surplus. Assuming a rather limited number of zones, the resulting fee optimization problem will be a low dimension problem, for which surrogate-based optimization is a suitable methodological approach. In short, surrogate-based optimization samples a number of initial combinations of zone fees and a response surface is fitted to the sampled fees and their corresponding objective function values. In the next phase, additional fees are sampled based on an infill strategy, which is an iterative process where additional samples are chosen based on both the determination of less explored areas of the feasible space and of areas where good objective function values have been found previously. The optimization of fees is then done for the response surface, which provides an approximation of the real problem, for which global optimization algorithm can be efficiently applied. Each sampling, however, require the user equilibrium problem to be solved. The methodology does not guarantee the global optimal solution to be found, but high quality results have been obtained for a similar problem of similar dimension in (30).

The optimization technique has been applied in several studies on optimal congestion charges; see e.g. (30), (32), (33) and (34). Here initial samples are chosen based on a maximin latin hypercube design, a conical radial basis function is adopted as the response surface, and the infill strategy is based on candidate point sampling. Details on all of these components can be found in (30). As a basis for the implementation of the surrogate-based optimization, the Surrogate Model Optimization Toolbox for MATLAB (35) has been used. Depending on the number of the number of clusters, the total number of sampled fees will vary between 32 and 178 for the different investigated number of clusters. As comparison, the mean MSC price for the links belonging to the same cluster has also been evaluated.

FIGURE 2 Network model of the Stockholm region, with the three zones based on spatial similarities. The CBD is marked with green links, the bypass highway is marked with red links, and the zone including all other links is marked with blue.
RESULTS

This section presents results for the different scenarios outlined in the previous section. As comparison, the first-best MSC prices provide an improvement of the social surplus with 1 129 421 SEK (denoted $\Delta S^{MSC}$) which thus is the maximum improvement that any pricing scheme can achieve on this network. Table 2 and 3 presents the fees and resulting improvement in social surplus when applying mean distance or time MSC prices and optimal fees respectively, for both network wide and cluster-based fee structures.

### Table 2 Fees, in SEK per kilometer or minute, and resulting improvement of social surplus based on mean MSC prices

<table>
<thead>
<tr>
<th>Fee structure</th>
<th>Fees in SEK</th>
<th>$\Delta S$ in SEK</th>
<th>$\Delta S/\Delta S^{MSC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance-based network wide</td>
<td>2.16</td>
<td>588 447</td>
<td>52%</td>
</tr>
<tr>
<td>Time-based network wide</td>
<td>1.05</td>
<td>925 856</td>
<td>82%</td>
</tr>
<tr>
<td>Delay-based network wide</td>
<td>1.95</td>
<td>895 625</td>
<td>79%</td>
</tr>
<tr>
<td>Distance-based 2 clusters</td>
<td>1.25;11.32</td>
<td>631 027</td>
<td>56%</td>
</tr>
<tr>
<td>Distance-based 3 clusters</td>
<td>0.99;49.53;8.03</td>
<td>812 989</td>
<td>72%</td>
</tr>
<tr>
<td>Distance-based 4 clusters</td>
<td>0.91;49.53;6.48;17.59</td>
<td>887 046</td>
<td>79%</td>
</tr>
<tr>
<td>Distance-based manual 3 clusters</td>
<td>1.76 (bypass);4.11(CBD);0.98 (rest)</td>
<td>697 555</td>
<td>62%</td>
</tr>
<tr>
<td>Time-based 2 clusters</td>
<td>0.57;2.54</td>
<td>1 051 702</td>
<td>93%</td>
</tr>
<tr>
<td>Time-based 3 clusters</td>
<td>0.52;4.31;2.14</td>
<td>1 089 297</td>
<td>96%</td>
</tr>
<tr>
<td>Time-based 4 clusters</td>
<td>1.33;0.39;2.59;5.03</td>
<td>1 109 467</td>
<td>98%</td>
</tr>
<tr>
<td>Time-based 5 clusters</td>
<td>0.35;1.98;3.01;1.11;5.28</td>
<td>1 114 683</td>
<td>99%</td>
</tr>
<tr>
<td>Time-based manual 3 clusters</td>
<td>1.46 (bypass);1.35(CBD);0.77 (rest)</td>
<td>940 230</td>
<td>83%</td>
</tr>
</tbody>
</table>

### Table 3 Fees, in SEK per kilometer or minute, and resulting improvement of social surplus based on optimal fees

<table>
<thead>
<tr>
<th>Fee structure</th>
<th>Fees in SEK</th>
<th>$\Delta S$ in SEK</th>
<th>$\Delta S/\Delta S^{MSC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance-based network wide</td>
<td>1.52</td>
<td>704 895</td>
<td>62%</td>
</tr>
<tr>
<td>Time-based network wide</td>
<td>1.39</td>
<td>967 687</td>
<td>86%</td>
</tr>
<tr>
<td>Delay-based network wide</td>
<td>6.83</td>
<td>1 114 102</td>
<td>99%</td>
</tr>
<tr>
<td>Distance-based 2 clusters</td>
<td>1.29;5.83</td>
<td>777 253</td>
<td>69%</td>
</tr>
<tr>
<td>Distance-based 3 clusters</td>
<td>1.02;26.99;5.87</td>
<td>852 559</td>
<td>75%</td>
</tr>
<tr>
<td>Distance-based 4 clusters</td>
<td>1.06;59.77;4.63;24.32</td>
<td>925 779</td>
<td>82%</td>
</tr>
<tr>
<td>Distance-based manual 3 clusters</td>
<td>3.31;0.86;14.69;6.52;48.05</td>
<td>982 670</td>
<td>87%</td>
</tr>
<tr>
<td>Distance-based manual 3 clusters</td>
<td>1.42 (bypass);2.80 (CBD);1.19 (rest)</td>
<td>752 145</td>
<td>67%</td>
</tr>
<tr>
<td>Time-based 2 clusters</td>
<td>0.80;2.45</td>
<td>1 063 696</td>
<td>94%</td>
</tr>
<tr>
<td>Time-based 3 clusters</td>
<td>0.63;3.79;2.22</td>
<td>1 092 593</td>
<td>97%</td>
</tr>
<tr>
<td>Time-based 4 clusters</td>
<td>1.32;0.44;2.59;5.01</td>
<td>1 109 991</td>
<td>98%</td>
</tr>
<tr>
<td>Time-based 5 clusters</td>
<td>0.38;2.03;2.98;11.12;4.93</td>
<td>1 115 158</td>
<td>99%</td>
</tr>
<tr>
<td>Time-based manual 3 clusters</td>
<td>1.47 (bypass);1.85 (CBD);1.25 (rest)</td>
<td>983 578</td>
<td>87%</td>
</tr>
</tbody>
</table>

Network wide pricing

First, considering the use of mean MSC prices, travel time and delay-based fees provide improvements of the social surplus in the region of 80% of the first-best pricing scheme, in contrast to distance-based fees which only achieve 52%. Applying the surrogate-based optimization method, considerable improvements of the social surplus is possible. One single distance-based fee of 1.52 SEK per kilometer driven will achieve 62% of the first-best social surplus, and a time-based fee of 1.39 SEK per minute driven will achieve 86%. The delay-based fee provide the maximum improvement of social surplus, and achieve 99% of the first-best social surplus improvement with a fee 6.83 SEK per minute delay.

Different combinations of fee structures have also been evaluated. Some further improvements are possible by using combined fee structures, but the improvements are rather small, in the region of 1000 to...
3000 SEK. Compared to the size of the overall social surplus improvements these gains are negligible and therefore not presented here in more detail.

**Cluster based pricing**

When determining clusters based on distance or time MSC price similarities, the case with time-based fees achieves 94% of the first-best social surplus already with two clusters of links, which is considerably higher in comparison to the distance-based case with five clusters. The results also show that cluster distance or time mean MSC prices provides improvements of the social surplus close to what is achieved by the optimization process. For the case of time-based fees, the improvement from mean values to optimal fees is overall rather small.

Table 5 presents the standard deviation of distance or time MSC prices within each cluster. For the manually determined three clusters, which provides zones with connected links according to Figure 2, the standard deviation of the MSC prices is considerably higher, in comparison to the three clusters based on MSC price similarities. It also follows that the improvement in social surplus is considerably smaller for the manually determined clusters, with a larger difference for the case if time-based fees. Comparing distance and time-based clusters, the standard deviation is considerably smaller for the case of time-based pricing, suggesting that the MSC prices within the cluster will deviate less from the cluster mean MSC price.

For the case of time-based fees, inclusion of distance-based fees in the optimization process has also been evaluated for both the manually determined clusters and the ones based on MSC price similarities. For all cases, the optimal decision is to set the distance-based fee to zero, and the time-based fee to the same level as when distance-based fees are not included.

**TABLE 4 Standard deviation of MSC prices within each cluster**

<table>
<thead>
<tr>
<th>Fee structure</th>
<th>Standard deviation (cluster separated by semicolon)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance-based 2 clusters</td>
<td>1.43;8.46</td>
</tr>
<tr>
<td>Distance-based 3 clusters</td>
<td>1.02;0.389</td>
</tr>
<tr>
<td>Distance-based 4 clusters</td>
<td>0.91;0.91;2.38</td>
</tr>
<tr>
<td>Distance-based 5 clusters</td>
<td>1.05;0.56;2.38;1.38;0</td>
</tr>
<tr>
<td>Distance-based manual 3 clusters</td>
<td>2.40 (bypass); 6.04 (CBD); 1.43 (rest)</td>
</tr>
<tr>
<td>Time-based 2 clusters</td>
<td>0.39;0.90</td>
</tr>
<tr>
<td>Time-based 3 clusters</td>
<td>0.33;1.06;0.52</td>
</tr>
<tr>
<td>Time-based 4 clusters</td>
<td>0.30;0.19;0.48;1.10</td>
</tr>
<tr>
<td>Time-based 5 clusters</td>
<td>0.15;0.26;0.40;0.22;1.10</td>
</tr>
<tr>
<td>Time-based manual 3 clusters</td>
<td>1.51 (bypass); 0.95 (CBD); 0.81 (rest)</td>
</tr>
</tbody>
</table>

**CONCLUSIONS AND FURTHER RESEARCH**

This paper investigates usage-based congestion pricing, with three different fee structures. For the standard one-link analysis, it is shown that optimal delay-based pricing is only depending on link type, but not capacity (i.e. the number of lanes) or link length. Thus, from a theoretical point, it is an attractive fee structure. It is also illustrated that a travel time-based fee will better approximate the MSC price curve, in comparison to distance-based pricing.

Applied to a network model of the Stockholm region, results from the one-link analysis holds. With delay-based pricing being the most efficient one, followed by time-based pricing. These results are in line with the results in (2), and in this paper, it has further been shown that both travel time and delayed-based pricing can achieve improvements in social surplus very close to what is achieved with first-best pricing, but with a fee structure more easily understandable. With delayed-based pricing, 99% of the maximum social surplus improvement can be achieved, using the same fee across the whole network. While such a fee structure impose other difficulties related to its application, the results show that the benefits can be very high. Combined distance and time-based fees have previously been proposed in the literature; see e.g. (7) and (12). The results in this paper has shown that the benefits of such joint fees are very limited.
To be able to determine optimal zone structures is an important area for further research. This paper has shown that it is possible to find clusters of links with similar distance or time MSC fees, and that such clustering, for the case of time-based pricing, can provide improvements of the social surplus, close to what is achieved with first-best pricing, already with only two clusters. It is, however, also important to include restrictions on spatial connectivity between links when determining such clusters, in order to determine zones which can be used in practice for usage-based pricing.

ACKNOWLEDGMENTS

This research has been supported by the Swedish Energy Agency (grant number 38921-1).

REFERENCES


