Efficiency and acceptability of pricing policies and transport investments in distorted economies

Jonas Westin

Doctoral Thesis in Transport Science
with a specialisation in Transport Systems
October 2012

Division for Transport and Location Analysis
Department of Transport Science
School of Architecture and the Built Environment
KTH Royal Institute of Technology
SE-100 44 Stockholm, Sweden
Akademisk avhandling som med tillstånd av Kungliga Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen fredagen den 26 oktober klockan 13.00 i sal F3, Lindstedtsvägen 26, Kungliga Tekniska högskolan, Stockholm.
Abstract

This thesis contains five papers studying the economic efficiency and political acceptability of road pricing policies and transport investments in distorted economies.

Interactions between the transport market and other distorted markets, such as the labor market, can have a large impact on the welfare effect of a road pricing policy or a transport investment. Many road pricing studies therefore try to incorporate effects from other distorted markets in the analysis. Paper I analyzes how the economic efficiency of a road toll in a distorted economy depends on assumptions about the initial tax system. In the road pricing literature, the welfare effect of a road toll is often found to depend on revenue use. Using a simple general equilibrium model paper I shows that the relative efficiency of marginal revenue recycling policies depends more on assumptions regarding inefficiencies in the initial tax system than on the road toll per se. Paper II studies the effect on welfare, equity and labor supply from a road toll in a commuting population with heterogeneous value of time and endogenous labor supply. When explicitly taking into account that commuters have different value of time, the road toll can increase total labor supply even when the revenues are not recycled back to the commuters. The analysis stresses the importance of recognizing traveler heterogeneity when analyzing congestion pricing.

Road pricing policies are often characterized by conflicting interests between different stakeholders and different geographical areas. Papers III and IV study the economic efficiency and political acceptability of pricing and investment policies in different institutional and geographical settings. The main contribution of the papers is to explain how political constraints can lead to inefficient tolling strategies. The papers contribute to the existing literature on political acceptability of road pricing by analyzing the conflict and potential trade-off between political acceptability and economic efficiency.

A difficulty when assessing the welfare effect of a future transport policy is also that many factors and parameters needed for the analysis are uncertain. Paper V studies the climate benefit of an investment in high speed rail by calculating the magnitude of annual traffic emission reduction required to compensate for the annualized embedded emissions from the construction of the line. The paper finds that to be able to balance the annualized emissions from the construction, traffic volumes of more than 10 million annual one-way trips are usually required, and most of the traffic diverted from other transport modes must come from aviation.
Acknowledgments

I would like to express my gratitude to my supervisor Lars-Göran Mattsson who introduced me to Transport Science and has provided me with great support, advice and helpful comments during the completion of this thesis. I am also grateful to my assistant supervisors Joel Franklin and Jonas Eliasson for all your support and interesting discussions.

I wish to thank all colleagues, current and former, at the Division for Transport and Location Analysis (TLA), the Centre for Transport Studies (CTS) and the Swedish National Road and Transport Research Institute (VTI). To Marcus Sundberg, Daniel Jonsson, Tom Petersen, Roger Pyddoke, Mattias Lundberg, Per Kågeson and Stef Proost who all taught me different fields economics, and to Erik Jenelius and Åke J. Holmgren with whom I wrote my first (but not last) scientific paper. To Gunnar Isacsson, Jan-Eric Nilsson, Yusak Susilo, Karin Brundell-Freij, Inge Vierth, Carl Hamilton, Lars Hultkrantz, Anders Karlström, Gunnar Lindberg, Jake Whitehead, Kandice Kreamer Fults and all others whom I might have forgotten who have commented and given suggestions on improvements on the papers.

I also wish to thank Lars Westin, Fredrik Olsson-Spjut, Martin Eriksson, Andreas Forsgren, Per-Olov Marklund, Rolf Hugoson and everyone at the Centre for Regional Science (CERUM) at Umeå University for facilitating my continued research by inviting me after my relocation to the Northern Hemispheres.

Finally, I would like to express my gratitude to my mother, father and brother for all your support and to my family, Ida, Stella and Konrad for everything.

This research was financed by the Swedish Transport Administration (Trafikverket), the former Swedish Road Administration (Vägverket), the Swedish Governmental Agency for Innovation Systems (VINNOVA), the ERA-NET Transport's Surprice programme on road user charging for passenger vehicles and the Centre for Transport Studies (CTS), which is gratefully acknowledged.

Örnsköldsvik, September 2012
Jonas Westin
**List of papers:**


Contents

1 Introduction ................................................................................................................................. 1

2 Welfare effects of transport policies in distorted economies .................................................... 3
   2.1 The labor market effect of road tolls...................................................................................... 4
      2.1.1 The labor market response at the extensive margin......................................................... 5
      2.1.2 The labor market response at the intensive margin......................................................... 5
      2.1.3 The labor market response in a heterogeneous population.............................................. 6
   2.2 Revenue recycling and the welfare effect of road pricing .................................................. 8
      2.2.1 Revenue recycling revisited............................................................................................... 9
      2.2.2 The difference between marginal and non-marginal policies........................................ 11
      2.2.3 The distributional impact of a road toll............................................................................ 13
   2.3 Economic efficiency in a utilitarian perspective ................................................................. 14
      2.3.1 Welfare and utility – two approaches to welfare economics............................................ 14
      2.3.2 Pareto improvements and Kaldor-Hicks efficiency......................................................... 15
      2.3.3 Consequences for the policy making process................................................................. 17

3 Political acceptability of road pricing policies ...................................................................... 19
   3.1.1 The cost of acceptability.................................................................................................. 21

4 Conclusions ............................................................................................................................. 22

References .................................................................................................................................. 24
Efficiency and acceptability of pricing policies and transport investments in distorted economies

Tous les événements sont enchaînés dans le meilleur des mondes possible ...
Cela est bien dit, répondit Candide, mais il faut cultiver notre jardin.
- Voltaire, Candide, Chapitre XXX Conclusion.

1 Introduction

Road congestion is a large problem in many urban areas where a high demand for road traffic during morning and afternoon peak hours causes long delays. To deal with this problem, transport economists have for a long time argued for the use of road pricing to reduce congestion and improve the urban environment.

The papers in this thesis study the economic efficiency and political acceptability of road pricing policies and transport investments in distorted economies. The first two papers study the welfare effects of road tolls in distorted economies. The next two papers analyze the tension between economic efficiency and political acceptability of urban road pricing policies targeted at reducing congestion and improving the urban environment. The last paper finally analyzes the climate effect of investments in high speed rail.

In a standard textbook analysis of road pricing, Pigouvian taxes are used to internalize congestion and environmental externalities by equating the price of travel with its marginal social cost. If the economy has no other distortions and all prices in the economy are equal to their marginal costs, this pricing rule ensures a welfare improving Pareto efficient solution. This result does however not necessarily hold if there are distortions in other interconnected markets (Rouwendal and Verhoef 2006).

The main contribution of paper I is to analyze how the economic efficiency of a road toll in a distorted economy depends on assumptions about the initial tax system. In the road pricing literature, the welfare effect of a road toll is often found to depend on revenue use. Using a simple general equilibrium model paper I shows that the relative efficiency of marginal revenue recycling policies depends more on assumptions regarding inefficiencies in the initial tax system than on the road toll per se.

Since a transport market often has large feedback externality effects in other distorted markets, such as the labor market, the actual welfare effect of a transport policy may differ significantly from the effect predicted by a first-best analysis that ignores spillover effects in other distorted markets (Parry and Bento 2002). To evaluate the full effect from a transport policy, interactions between the transport market and other distorted markets therefore need to be incorporated in the analysis.
A critical assumption in many cost-benefit analyses of urban road tolls is that the whole population has a single value of time. Paper II studies the effect of a road toll on welfare, equity and labor supply in a commuting population with heterogeneous value of time and endogenous labor supply. When explicitly taking into account that commuters have different value of time, aggregate labor supply is found to increase even when the revenues are not recycled back to the commuters. This indicates that a road toll by itself does not automatically reduce aggregate labor supply. The paper hence stresses the importance of recognizing traveler heterogeneity and mode-choice self-selection when analyzing congestion pricing.

Although road pricing often is an efficient way of reducing urban congestion, the public and political support for efficient road pricing policies is often low. In the road pricing literature several factors for the low acceptability have been identified ranging from aversion of pricing, perceived loss of freedom and equity and fairness considerations to uncertainty about revenue use and awareness of problems caused by car traffic (Shade and Schlag 2003).

Urban road pricing policies are often characterized by conflicting interests between different stakeholders and geographical areas. The conflicting interests can make it difficult for the political process to achieve support for efficient road pricing policies. As a consequence, few successful implementations of efficient urban road pricing schemes can be found where the purpose of the toll is to reduce transport externalities such as congestion, emissions, noise and accidents. The use of road tolls is instead often associated with construction of new transport infrastructures where the main purpose of the toll is to finance the investment rather than to reduce existing externalities.

Many regions face problems with through-traffic that causes local negative externalities. A response to this problem is often to invest in new bypass transport infrastructure to divert transit traffic from the city center. Increasingly, such proposals are accompanied by tolling as a means of finance. An optimal policy from an efficiency point of view would be to consider tolling both on the new bypass and on the existing road that is relieved based on marginal cost pricing principles. Political constraints protecting influential interest groups can however lead to inefficient tolling where the political feasible solution is to only toll the bypass and not the city center road.

Paper III and IV study the economic efficiency and political acceptability of pricing and investment policies in different institutional and geographical settings. Using theories from political economy, the papers examine how political constraints can lead to inefficient tolling strategies. The papers contribute to the existing literature on political acceptability of road pricing by analyzing the conflict and potential trade-off between political acceptability and economic efficiency.

Paper V analyses a completely different area, the climate benefit of investment in high speed rail. The paper studies the climate benefit of an investment in high speed rail by calculating the magnitude of annual traffic emission reduction required to compensate for the annualized embedded emissions from the construction of the line. To account for uncertainties in underlying assumptions, a Monte Carlo simulation framework is used in the analysis. The paper finds that, to be able to balance the annualized emissions from the construction, traffic volumes of more than 10 million annual one-way trips are usually required, and most of the traffic diverted from other modes must come from aviation.
2 Welfare effects of transport policies in distorted economies

To analyze the welfare effect of transport policies in distorted economies, a research literature has emerged where interactions between the transport market and other distorted markets are studied (Calthrop et al. 2010; Fosgerau and Pilegaard 2008; Mayeres and Proost 1997; Parry and Bento 2001; Van Dender 2003). A general methodology in this literature is to incorporate effects in other distorted markets in the cost-benefit analysis by embedding a model of the transport market in a general equilibrium framework. This type of analysis is often referred to as second-best in the sense that it incorporates market imperfections and other type of restrictions outside of the transport system in the analysis.

One market that has been given special attention in the road pricing literature is the labor market. Since the worst congestion is usually related to peak hour travel, this has raised apprehensions about how the labor market might be affected by a road toll targeted at reducing peak hour travel, which is often associated with work-related commuting.

The main argument is that since a road toll raises the monetary cost of commuting, it can decrease employment at the extensive margin in a similar way as an income tax. Parry and Bento (2001) even argue that the resulting welfare losses in the labor market can exceed the Pigouvian welfare gain from internalizing the congestion externality in the transport market. However, if the collected revenues from the road toll are used to cut distortionary income taxes, they show that a double dividend can arise where welfare increases both in the transport market and in the labor market. The literature is hence related to the double-dividend literature and the idea that it sometimes is efficient to tax externalities higher than their partial equilibrium Pigouvian level, given that the revenues are used to cut distortionary taxes elsewhere in the economy (see Bovenberg 1999; Parry and Oates 2000; Ballard and Don Fullerton 1992; Schwartz and Repetto 2000; and Kim 2002).

A related argument, often found in the road pricing literature, is therefore that the welfare effect of a road toll critically depend on how the collected revenues are recycled back into the economy (see Parry and Bento 2001; Van Dender 2003; De Borger and Wuyts 2009; Fosgerau and Van Dender 2010). A reason for this is that the value of the collected revenues often is larger than the monetary value of the time gains from the reduced congestion. During the Stockholm congestion charging trial in 2006, the value of the shorter travel times were for example estimated to be around two thirds of the collected toll revenues (Eliasson 2009).

Comparing different revenue recycling schemes; a labor tax cut is in general preferred to a lump-sum transfer (Parry and Bento 2001; Verhoef and Ubbels 2002), unless equity considerations are explicitly included in the social welfare function as in Mayeres and Proost (2001). The argument is related to the weak double-dividend claim that an environmental tax in general improves welfare more if the revenues are returned through cuts in other distortionary taxes instead of being returned in a lump-sum transfer (see Goulder 1995; and Bovenberg 1999).

Using a simple analytical model, Parry and Bento (2001) also argue that the efficiency gains are larger if the toll revenues are recycled through reduced labor taxes than if the revenues are spent on increased public transport subsidies. An opposite result is however found in De Borger and
Wuyts (2009) who argue that since a labor tax cut and an increased public transport subsidy have very different effects on congestion, it may be more efficient to recycle the revenues via a targeted public transport subsidy, rather than to spend them on a general reduction of distortionary labor taxes. The choice of which markets to include in the model and how to model the interaction between them can therefore have a strong impact on the result of the analysis. The economy of scale in the public transport system can for instance be important for the overall welfare from a policy that induces a modal-shift from car to public transport.

A troublesome implication of the conflicting results in the literature is that the estimated welfare effects and the relative efficiency of different revenue recycling policies depend, not only on which markets that are included in the analysis, but also on how the interactions between different transport modes are modeled. However, as pointed out by Rouwendal and Verhoef (2006), since many of the needed relationships can be hard to measure, both the total welfare effect as well as the optimal toll levels can be difficult to assess empirically.

This has given rise to the question of how to evaluate the economic efficiency of a road toll or an infrastructure investment in a distorted economy? From the road pricing literature three arguments can be identified:

1. Since a road toll raises the monetary cost of commuting, it can reduce labor supply at the extensive margin. In combination with pre-existing distortions in the labor market, this can result in a welfare loss of the same magnitude as the welfare gains in the transport market.

2. The welfare effect of a road toll therefore needs to be analyzed in a second-best setting where interactions with other distorted markets are included in the analysis; and

3. How the revenues are recycled back to the economy are critical for the total welfare of a road pricing policy.

In the following subsections these three arguments will be discussed. The argumentation is based on the first two papers in the thesis.

2.1 The labor market effect of road tolls

Determining the labor market effect of a road toll is difficult empirically. In the evaluation of the Stockholm congestion charging trial, the labor market effect was discussed as a source of uncertainty but was not included in the final cost-benefit calculations (Eliasson 2009). While accessibility to the labor market can be reduced for commuters who are priced out by the road toll, accessibility can be improved for high-income earners and business related travelers who are willing to pay the charge. There are also long term and dynamic effects related to both the matching in the labor market and the housing and location market.
2.1.1 The labor market response at the extensive margin

A standard assumption in the transport literature on labor supply and commuting is that a road toll decreases labor supply (Calthrop et al. 2010; Mayeres and Proost 1997; Mayeres and Proost 2001; Parry and Bento 2001; Van Dender 2003). The result is generally based on the assumption that the commuters can only choose labor supply at the extensive margin (participation in the labor force modeled as the number of workdays), whereas labor supply at the intensive margin (work hours per day or work intensity) is assumed to be fixed. The assumption creates a strict complementarity between labor supply and the number of commuting trips. This means that, conditional on a chosen transport mode, a commuter can only reduce his or her transport cost by reducing labor supply.

A change in labor supply at the extensive margin can be interpreted in multiple ways; as an effect on the actual number of work days, as a change in the labor force participation rate or as a location effect. The models in paper I and II analyze commuting between two fixed geographical locations where the individuals are assumed to live in a suburb and commute to work in the central business district (CBD). Labor supply in this setting can therefore be interpreted as the share of the population who choose to work in the CBD, compared to those who choose to work in the local area or choose to exit the labor force altogether. A policy that increases (reduces) labor supply, would therefore correspond to the situation where more (less) people choose to work in the CBD and less (more) choose to work where they live or quit the labor market.

In many models in the road pricing literature, the road space during peak hour is assumed to only be used for commuting, thus ignoring the impact from other type of trips such as leisure trips and freight transports. The motivation for this assumption is often that work trips constitute the main part of the traffic during the morning and afternoon peaks when congestion is most severe. However, even a small share of non-work-related car trips can have a large impact on the marginal travel time.

Policies that reduce the amount of non-work-related traffic during rush hour can therefore have a positive effect on labor supply. Differentiating the road toll between trip purposes can therefore reduce congestion without having the same negative impact on commuting (see Van Dender 2003). But even a uniform road toll can increase the number of work trips (and hence labor supply) if the value of the time gains from the reduced number of leisure trips is large enough to outweigh the increased monetary cost of commuting. This can both happen if the initial share of non-work-related travel is large, or if the leisure trips are more cost-sensitive than the work trips and the congested speed-flow curve is steep.

2.1.2 The labor market response at the intensive margin

The assumption that labor supply only is chosen at the extensive margin is however a critical assumption when analyzing the effect of a road toll, since it implies that, “conditional on the choice of transport mode ... workers may only reduce their commuting cost by reducing their total labor supply” as pointed out by Gutiérrez-i-Puigarnau and van Ommeren (2010). The assumption is often motivated with results from the empirical labor market literature, where the labor supply response at the extensive margin (labor force participation) in general is seen as more important than the response at the intensive margin (hours of work) (Kleven and Kreiner 2006).
Many empirical studies indicate that the labor supply elasticity is larger along the extensive than the intensive margin, that is, the employees are more willing to adopt their participation than hours worked (Kimmel and Kniesner 1998). There might also be constraints and regulations in the labor market that limit the individuals’ possibilities to choose their work hours. Using socio-economic panel data for Germany, Gutiérrez-i-Puigarnau and van Ommeren (2010) study the effect of commuting cost on labor supply patterns. Using an exogenous change in commuting distance as a proxy for a change in transport cost, they find a small positive effect of distance on weakly labor supply. This implies that the positive effect on daily hours, at least for Germany, is stronger than the negative effect on the number of work days. Their analysis does however not include effects on labor force participation.

Once at work, a road toll makes it more attractive to work longer hours for example by changing from a part-time job to a full-time job or work overtime. Since a road toll can have a positive effect on labor supply at the intensive margin, it is nevertheless important to not beforehand disregard the effect on both the extensive and the intensive margin when estimating the total effect on labor supply from a road toll. Especially in a flexible labor market where the commuters are allowed to choose the number of work hours without restrictions (such as small business owners and self-employed), the commuters can compensate for the increased commuting cost by working longer hours each work day, making the overall effect on total labor supply, at least in theory, ambiguous.

2.1.3 The labor market response in a heterogeneous population

A critical assumption in a large part of the road pricing literature where the welfare effect of a road pricing policy is analyzed is that all commuters have the same value of time. This is a problematic assumption when it comes to road tolls, since one of the main features of the policy instrument is that it sorts people according to their value of time, given the existence of feasible alternatives.

The welfare effect of a road toll in a heterogeneous population is examined in paper II. In the paper, a population with continuously distributed value of time commutes between home and work in a static economy. Each commuter chooses labor supply at the extensive margin and mode of transportation based on his or her daily wage. The commuters can choose between two transport modes to get to work, a fast but expensive mode subject to congestion (car) and a slow but cheap mode with no congestion (public transport). Given a fixed wage distribution, there exists a unique modal-split wage level, such that all commuters with a wage above the split-point only choose car and all commuters with a wage below the split-point only choose public transport.

Since a road toll increases the monetary cost of car commuting, the toll will shift the modal-split wage upwards causing more people to commute by public transport instead of by car. The effect on total labor supply therefore depends on both the effect from the modal-shift and the effect on labor supply for the remaining car commuters. The effect is illustrated in Figure 1 where the change in labor supply as a function of the wage before and after the introduction of a road toll is shown. To isolate the direct effect from the road toll we assume that the revenues are not recycled back to the commuters.
Figure 1: Change in labor supply as a function of daily wage for a road toll in a population with continuous wage distribution where the revenues are recycled elsewhere in the economy. The road toll has a negative effect on labor supply for the car commuters who switch from car to public transport, but has a positive effect for the commuters that continue to commute by car.

The increased monetary cost of car commuting shifts the modal-split wage upwards, reducing the number of car commuters. For the commuters that change travel mode, the road toll has a negative effect on labor supply. For the remaining car commuters, the positive effect on labor supply from the reduced congestion more than offset the negative effect from the increased monetary cost of commuting. The road toll has therefore a positive effect on labor supply for the commuters that continue to commute by car.

By introducing user heterogeneity into the analysis we see that the distributional impact of the policy is important for explaining the full effect on labor supply. Increasing the monetary cost of car commuting will make some people to work less while others will work more. The effect on total labor supply is hence ambiguous. When the revenues are used elsewhere in the economy, aggregate labor supply is positive in this example.

The difference between this and previous models is that this model explicitly takes into account that people have different value of time and explicitly model how their choice of travel mode is affected by their daily wages. The mode-choice self-selection mechanism reduces the negative impact on labor supply from a road toll since it makes the commuters with the lowest willingness to pay for car transport switch transport mode so that the remaining car commuters, with a higher willingness to pay, can work more.
2.2 Revenue recycling and the welfare effect of road pricing

When analyzing a transport policy in a distorted economy, spillover effects from the transport market into other distorted markets can have a large effect on the total welfare of the policy. In an undistorted labor market, the welfare effect from a marginal decrease in labor supply would be negligible. With pre-existing distortions in the labor market, the labor supply effect becomes more important. This can be illustrated in a simple example.

We consider a situation where a working population commutes on a congested road. To reduce congestion, a road toll is introduced. If the toll increases the cost of commuting, it decreases the net wage after tax and transportation in a similar way as a raised income tax. The size of the effect depends on what alternative transport modes the commuters can use to get to work. Given that the labor market already is distorted by an income tax, the increased cost will be added to the existing distortion, creating an additional deadweight loss in the labor market. The effect is illustrated in Figure 2.

Figure 2: The welfare effect in the labor market from a road toll that increases the commuting cost in a labor market with a pre-existing distortionary income tax.

We assume that labor supply is directly proportional to the net wage after taxes and transportation, i.e. $L(w) = w$. To simplify the calculations, we normalize the gross wage to one. The income tax $t$ gives rise to a loss in the labor market equal to $\frac{t^2}{2}$. We further assume that the road toll increases the cost of commuting with $\Delta \tau$. From the figure we see that the road toll increases the pre-existing deadweight loss from the income tax, reducing the net wage from $1 - t$ to $1 - t - \Delta \tau$ and labor supply from $L_t = 1 - t$ to $L_{t+\Delta \tau} = 1 - t - \Delta \tau$. The resulting welfare loss is given by the dashed area and is equal to $t\Delta \tau + \frac{\Delta \tau^2}{2}$. Without the pre-existing distortion, the loss from the road toll in the labor market would only have been $\frac{\Delta \tau^2}{2}$. Interactions between the transport market and other distorted markets can therefore have a significant effect on both the total welfare effect of a road toll and the optimal toll level.
Parry and Bento (2001) argue that the resulting loss in the labor market from a road toll can be of the same magnitude as the welfare gain in the transport market from the reduced congestion externality. However, if the collected toll revenues are used to cut the distortionary income tax in the labor market, welfare will increase both in the transport market and in the labor market since the reduced income tax \( \Delta t \) decreases the initial deadweight loss.\(^1\) They therefore argue that the chosen form of revenue recycling is crucial for the overall welfare effect of a road toll.

### 2.2.1 Revenue recycling revisited

To analyze the welfare effect of a proposed road toll, we need to specify a social welfare function, demand functions for the agents in the economy, production functions for the firms, and budgetary and behavioral constraints for the agents. A behavioral constraint can for example be that an agent treats an endogenous parameter as exogenous, such as the marginal impact on travel time from traveling on a congested road. We also need to choose the available policy instruments and define restrictions for what the government is allowed to do. The restrictions can be motivated on different grounds; a political restriction can be that the government cannot use lump-sum taxes, and a technological restriction can be that the road toll cannot price discriminate between different travelers based on their trip purpose.

In addition to this, we need to calibrate the model by specifying the model parameters and set initial values for the policy instruments (taxes, tolls and transfers) in the model. How the policy instruments are chosen is important, especially when we want to compare the effect of different revenue recycling policies in a distorted economy. The reason for this is that when we model the economy, we also implicitly determine an optimal combination of policy instruments that maximizes social welfare given the constraints placed on the government.

We can therefore make a distinction between initial situations where the policy instruments are chosen to maximize social welfare, and initial situations where they are not optimally chosen. In a distorted economy, this is the same as asking whether the distorted markets are balanced or unbalanced. By balanced we mean that the marginal costs (and benefits) of all non-constrained policy instruments are equal and that the government cannot increase social welfare without violating any constraint. Observe that distortionary taxes still can be used in a balanced tax system if they are motivated by for instance welfare from public spending and the absence of lump-sum taxes.

Analyzing the welfare effect of a new road toll can in many situations be interpreted as an analysis of the effect of lifting a constraint on a previously restricted policy instrument. That is, we study a situation where the government prior to the policy was restricted from using road tolls (by technical, political or other reasons).

---

\(^1\) Note however that while a marginal reduction of the income tax \( \Delta t \) reduces the pre-existing deadweight loss from the income tax, it has a negligible effect on the additional deadweight loss caused by the road toll. The reduction in the pre-existing deadweight loss is \( t \Delta t - \frac{\Delta t^2}{2} \) and the reduction in the additional loss from the road toll only is \( \Delta \Delta r \). If we look at marginal changes, all squared delta terms are negligible.
If we use a model with an unbalanced tax system in the analysis, such as the model in Parry and Bento (2001) where the initial income tax is above the optimal level, we can always increase welfare by adjusting the initial tax system, regardless of whether the road toll is used or not. This makes it difficult to separate the welfare effect of the road toll from the welfare effect of the general adjustment of the underlying tax system. The approach can be reasonable if we want to study the composition of the tax system as a whole, but becomes problematic if we want isolate the welfare effect of a specific policy instrument. The point is illustrated in Figure 3.

In the figure we see social welfare as a function of two policy instruments. The grey lines show iso-curves for social welfare and we assume that welfare increases in both instruments. Prior to the road toll, the government’s budget constraint is given by the solid black line. In the initial situation we are in point A. The tax instruments are unbalanced in this situation since the government can increase welfare just by adjusting the policy instruments along the existing budget line.

Assume now that the government wants to know the welfare effect of introducing a new road toll. Without loss of generality, we assume that the only effect of the road toll is to provide the government with additional revenues. The road toll hence shifts the budget constraint outwards to the dashed line. This allows the government to use the revenues to change the two policy instruments. Since the tax system is unbalanced, recycling the revenues through policy instrument 2 (A to B) increases welfare more than if the revenues are recycled through policy instrument 1 (A to C). But going from A to B is from a welfare perspective identical to first making a budget-neutral adjustment of the tax system (by moving from A to D) and then recycling the revenues from the road toll through policy instrument 1 (D to B).

Analyzing the welfare effect of a road toll in an unbalanced tax system is problematic since we compare two things at the same time; the welfare effect of the road toll and the welfare effect of an adjustment of the underlying tax system. Because why should the analysis be limited to only
evaluate policies where merely the toll revenues are used for adjusting the tax system, when we, without violating any constraints can improve welfare even more by making an even larger adjustment part of the recycling policy, (A to E) for example.

An argument for analyzing a situation with an unbalanced tax system is that the real tax system we want to model is likely to be unbalanced. A new policy instrument can also provide the government with an opportunity to rebalance the tax system. However, there can still be a good idea to separate the welfare effect of the general tax adjustment from the welfare effect of the road toll, not least because it is difficult to measure (and agree upon) all the costs and benefits in the tax system and hence to know how balanced the tax system in the initial situation actually is. By analyzing the system “as if” the policy instruments in the initial constrained no-toll situation are chosen on welfare maximizing principles, we can isolate the welfare effect of the specific policy instrument we are interested in from the effect of a general adjustment of the tax system.

The whole revenue recycling approach can also be criticized from a much broader perspective since it disregards the effect a chosen recycling policy may have on subsequent policies and decisions by assuming that the recycling policy is additional to the business-as-usual situation. To determine the marginal welfare effect of a revenue recycling policy where we choose between recycling the revenues either on project A or on project B, we need to know whether the non-selected project will be implemented anyway or not. If both projects will be carried out regardless of what we choose, then it becomes problematic to say that one of the recycling policies improves welfare more than the other. Without this knowledge, it may be better to value the collected toll revenues more in terms of an “average” marginal cost of public funds, instead of the marginal cost (or benefit) of a particular policy instrument or project that the revenues are earmarked for.

An illustration of this problem is the Stockholm congestion charging trial where the collected toll revenues were earmarked for road investments (Eliasson et al. 2009). Because the road investments in Stockholm also are funded by from a national level, the earmarked toll revenues may reduce the regular funding from the national government. In this sense revenue recycling is just swings and roundabouts. How the system boundary for the analysis is drawn therefore has a large effect on the relative efficiency of different revenue recycling polices.

2.2.2 The difference between marginal and non-marginal policies

If we analyze a situation where we relax a constraint on a policy instrument starting from a constrained optimum, then there exists a subset of budget neutral recycling policies with the same marginal benefit. This result does however only hold for marginal policies. For non-marginal policies we also need to consider how the interaction between the road toll and the remaining system affects the optimal levels of the other policy instruments. An illustration of this is shown in Figure 4 where the welfare effect of a road toll under five different recycling policies are analyzed as a function of the chosen toll level.
For marginal toll levels, the road toll $\tau$ has the same effect on welfare, regardless of whether the revenues are used to decrease the income tax, to increase the public transport subsidy, to increase the consumption of a public good, or if all remaining policy instruments are re-balanced to maximize welfare as a function of the road toll. Since the distortions in the initial situation is balanced, the marginal cost of public funds of the income tax is equal to both the marginal benefit of the public good and the marginal benefit of the public transport subsidy. Only the lump-sum recycling policy performs worse than the other recycling policies. Since the model does not provide any motivation for the use of lump-sum transfers (such as inequality aversion) this implies that the government, if allowed to, would use a lump-sum tax rather than an income tax to finance public spending. Because the lump-sum transfer is constrained to zero in the initial no-toll situation, the marginal benefit of increasing the transfer is lower than for the other non-constrained policy instruments.

For non-marginal toll levels, the relative efficiency of the different recycling policies starts to deviate from each other. When the toll increases, social welfare increases more if the revenues are recycled through a lower income tax or an increased consumption of the public good, compared to when the revenues are spend on increasing the public transport subsidy. The explanation for this is that the benefit of using the public transport subsidy in this model primarily comes from its ability to reduce congestion by attracting commuters to change from car to public transport. Since the road toll reduces congestion, it also reduces the need for (and potential benefit of) the public transport subsidy. Recycling the revenues through an increased public transport subsidy therefore leads to an over-subsidized public transport system, given that the subsidy was set at its optimal level in the initial no-toll situation. The result crucially depends on the motivation for using the public transport subsidy in the initial situation. If the subsidy for example is motivated by a Mohring effect or an environmental externality rather than by reduced congestion, the results can
change. Higher-order interactions between the policy instruments therefore complicate the analysis for non-marginal policies by shifting the optimal levels of the policy instruments.

2.2.3 The distributional impact of a road toll

Even if the choice of revenue recycling instrument has no or little effect on welfare in a balanced tax system, the distributional impact of the policy can still depend on how the collected toll revenues are recycled. Using the model developed in paper II, the distributional impact of a road toll under two different revenue recycling policies with the same effect on aggregate welfare is analyzed.

Figure 5 shows welfare as a function of the daily wage in a population of commuters with a heterogeneous wage distribution. To clarify the discussion, the population is divided into four different groups (I,II,III,IV); group I consists of people with the lowest wage that do not commute; group II contains public transport commuters; group III consists of car commuters in the no-toll situation that switch to public transport because of the road toll; and group IV consists of the car commuters that remain to drive car after the road toll is introduced.

![Figure 5: Welfare as a function of daily wage for a road toll of 5€ under two different revenue recycling schemes, a lump-sum transfer and an income tax cut.](image)

As can be seen from the figure, the distributional impact of the road toll strongly depends on the chosen form of revenue recycling even though the effect on total welfare is the same for both recycling policies. In the example, people on a low income (group I and II) benefit more from increased public spending (through a lump-sum transfer in this example) compared to a labor tax cut, while people on a high income (group II and IV) have opposite preferences. We also see that although the commuters with the highest income pay most of the charges, they also gain most from the road toll regardless of revenue use since they have the highest value of time.
2.3 Economic efficiency in a utilitarian perspective

An important question when analyzing a transport policy is whether, and to what extent, the policy improves social welfare. The question is often framed as a cost-benefit analysis where the costs of a policy are weighted against the benefits. Since it is not possible to measure welfare in utility terms directly, the welfare analysis often requires that the costs, benefits and the resulting social surplus can be expressed in monetary or financial terms. A monetary measure that is often used is the net present value (NPV) which is an indicator of how much value an investment adds to the society. This value is calculated as the discounted sum of all benefits (measured in monetary terms) minus the costs of the investment. A simple decision rule is to accept projects with a positive NPV or if several alternative policy alternatives are present, choose the one with the highest NPV.

Economic welfare analysis rests on a utilitarian framework where a social welfare function is used to measure and aggregate the utility changes for different individuals into a single number, often measured in monetary terms. This methodology is however not without its complications.

In this section we discuss the how the utility from different individuals are aggregated in a standard cost-benefit-analysis and what demands this places on the utility and social welfare functions. The main focus is on the Kaldor-Hicks criterion which, in combination with a money metric utility measure, is used to compare different allocations.

2.3.1 Welfare and utility – two approaches to welfare economics

From a utilitarian perspective we can think of each individual as carrying a “jar of utility”. A policy such as a road toll or a transport investment will in general increase the utility of some individuals and decrease the utility of other individuals. According to Wolff (1996) “[t]he fundamental idea of utilitarianism is that the morally correct action in any situation is that which brings about the highest possible total sum of utility”. In order to maximize utility we therefore need to measure the utility of all individuals. Total utility can therefore be calculated as $U(u_1, ..., u_n) = \sum_{i=1}^{n} u_i$ where $u_i$ is utility level of individual $i$ belonging to the affected population $i = 1, ..., n$. From a utilitarian perspective we can therefore define social welfare as the total utility in a population.

A common assumption in economic theory is that utility is expressed in our preferences, and that our preferences are expressed in our choices. Hence we can study utility by studying the choices people make. A problem with this notion is that the choices only reveal utility differences at an ordinal scale. The underlying cardinal utility (assuming the existence of such a thing) is not revealed. The lack of cardinality will in general not cause any problem, as long as we only are interested in making comparisons of different policy alternatives (including the null alternative) and only consider the effect on a single individual. Since the ordinal scale let us order the individuals preferences over any number of policy alternatives, we can always find the policy alternative that maximizes the individual’s utility even though we cannot measure the exact level of utility associated with that policy. In this discussion we also disregard from the difficulty of measuring the ordinal scale that for instance can occur of the preferences are not transitive.

The lack of cardinality does however create problems when a policy affects more than one individual, since measuring social welfare require us to take the sum over something that is only
measurable on an ordinal scale. This problem is often referred to as the problem of interpersonal comparisons of utility (Wolff 1996). Because it is not possible to measure individual utility on other than an ordinal scale, it is not possible to measure social welfare unless we impose sufficiently strong assumptions on the structure of the underlying individual utility functions. Without these assumptions, welfare may only be measured at an ordinal scale in terms of Pareto efficiency.

The approach taken by Pareto, Hicks and Kaldor among others to deal with this problem is to differentiate between efficiency and distribution. Efficiency is evaluated using the Pareto efficiency criterion and the Kaldor-Hicks compensation test, both of which only require ordinal measures of utility. For questions related to the income distribution part, this approach still requires cardinal utility functions since it relies on a social welfare function to transform the individuals’ utility into a single measure of the overall welfare or utility in the society. The social welfare function is however not restricted to just summing the individual utilities. Instead almost any function that for each combination of individual utility gives a measure of the social welfare can be used, i.e. \( f : \mathbb{R}^n \rightarrow \mathbb{R} \).

Social indifference curves for two such functions are illustrated in Figure 6 and Figure 7 above. The utilitarian (or Benthamite) welfare function computes social welfare as the sum of utility of all individuals in the society. If an individual increases his or her utility with one unit of utility, social welfare is increases with the same amount, regardless of how the utility is distributed in the society. An opposite position is taken in the maximin (or Rawlsian) welfare function that measure social welfare as the utility of the member in society being worst off. This means that social welfare only can increase if it improves the utility of the individual with the lowest utility. In practice some combination of the two functions are used, forming a convex social utility function that incorporate egalitarian aspects of how the welfare is distributed in the society with the total utility level in a strict utilitarian sense.

### 2.3.2 Pareto improvements and Kaldor-Hicks efficiency

Pareto efficiency implies it is impossible to make one person better off without making someone else worse off. The notion can be applied for evaluating the effect of a proposed policy against the status quo, or for comparing two policies against each other. If a policy, if implemented, would make no one worse of and at least one individual better off we call it a “Pareto improvement” compared to status quo. Likewise is a policy “Pareto efficient” if no further Pareto improvements can be made, that is, no one can be better off without someone else being worse off.
A problem with the Pareto improvement criterion is that the set of policy alternatives that is Pareto improving in many situations can be very restrictive since it requires that no one is made worse off. Instead the Kaldor-Hicks criterion is used in the traditional cost-benefit-analysis. With the Kaldor-Hicks criterion an outcome is defined to be more efficient if those who are better off in theory could compensate those that are worse off, and this would result in a Pareto improving outcome. The Kaldor-Hicks does not require that the compensation is actually carried out in practice. In policy making the Kaldor-Hicks test is usually used for evaluating whether or not a policy increases social welfare, i.e. whether the net present value is positive or negative. The criterion does not say anything about the distributional impact of the policy. These are instead often “considered to be a question for the politicians” (SIKA, 2005, p. 12).

The Kaldor-Hicks criterion means that we should prefer $x$ before $y$ if we can find a state $x'$ that can be reached by a series of reallocations of the resources in $x$, but not in $y$. The dotted line in Figure 8 corresponds to the set of Pareto improving states, and the solid line defines the set of possible Kaldor-Hicks improvements. The set of possible Pareto improvements is therefore a subset of the set of Kaldor-Hicks improvements.

The approach is however problematic. The core of the problem is the idea that social welfare is independent of how the welfare is distributed in the society. This is seen as a question of equity or fairness, not one of overall welfare. If we for instance think of social welfare as something that incorporates egalitarian aspects, the monetary Kaldor-Hicks criterion (without compensation) cannot guarantee that the social welfare do not decrease, since a policy that produces a social surplus in monetary terms can decrease social welfare if it increases the inequalities in the society.

A counterargument to this critique is that a cost-benefit-analysis deliberately should disregard from egalitarian aspects and analyze social welfare from a utilitarian perspective, i.e. the overall level of utility in the society. Questions regarding equity could then be performed in a separate analysis. The counterargument is however not unproblematic, even from a utilitarian perspective, since it imposes strong restrictions on the underlying structure of the individual utility functions and preferences. We use a small example to prove the point.

Assume that two individual have identical utility functions, $u_i = m_i^{1/\lambda}$ where $\lambda = 2$. Assume further that the individuals have to choose between two income allocations, $y = (10,10)$ and $x = (21,1)$. From Figure 8 and the Kaldor-Hicks criterion we see that $x$ is a Kaldor-Hicks improvement of $y$ since the first individual can compensate the second individual so that the new theoretical outcome $x'$ is a Pareto improvement. From a fast glance at the money metric utility functions in Figure 8 we could think that this also means that going from the state $y$ to the state $x$ also would increase social welfare in a utilitarian sense. To falsify this statement we only need to calculate and compare the total utility of $x$ and $y$. Since $U(10,10) = \sqrt{10} + \sqrt{10} > \sqrt{21} + \sqrt{1} = U(21,1)$ we see that the Kaldor-Hicks criterion (without compensation) does not necessarily imply that also total utility increases.

The main reason for this problem is that we in the example have assumed that the individuals have diminishing marginal utility of money. Since the utility variations are weighted by the marginal utility of money to get a money metric utility measure, this implies that the social welfare function represented by the Kaldor-Hicks criterion (without compensation) is anti-egalitarian in a utilitarian
perspective with concave indifference curves. This means that the utility of individuals with a low income is valued less than the utility of individuals with a high income. This can result in a tendency for the cost-benefit-analysis to overvalue policies targeted at high income earners since their utility is worth more in the analysis. The economic intuition behind this is because it often can be cheaper to compensate individuals with a low income than to compensate individuals with a high income. The problem is that since the compensation is only theoretical, this means that if it is less expensive to compensate a person with money than to do a project that benefit this person, the person will neither get the benefit of the project nor the compensation. If we let $\lambda$ go to infinity, corresponding to a situation where the marginal utility of money goes to zero, we get the social welfare function shown by the solid line in Figure 9.

![Figure 8: Indifference curves measured in monetary terms for a social welfare function based on Kaldor-Hicks efficiency.](image1)

![Figure 9: Indifference curves measured in utility terms for a social welfare function based on the Kaldor-Hicks criterion when $\lambda \to \infty$.](image2)

Figure 9 also shows that a Kaldor-Hicks improvement can result in a state with a lower total utility. Comparing the two figures we see that when welfare is measured on a monetary scale, $x$ improves social welfare more than $y$, but when welfare is measured as total utility, $y$ is better than $x$.

### 2.3.3 Consequences for the policy making process

From a utilitarian perspective, this means that the distinction between economic efficiency and distribution in welfare analysis is problematic since the Kaldor-Hicks criterion implicitly defines a social welfare function that transforms the individuals’ utility into a single measurement that takes into account both the effect on total utility and how the utility is distributed in the population. With diminishing marginal utility of income, the associated social welfare function is also anti-egalitarian from a utilitarian perspective.

The distributional impact cannot therefore be treated as something only for the politicians (i.e. as a matter of equity or fairness). If the compensations are not realized, the cost-benefit-analysis run the risk of misjudging the “true” social welfare of the proposed policy, if we by true refer to social welfare in a utilitarian meaning. From the opposite perspective one can argue that the use of the Kaldor-Hicks criterion in policy making increases the need for a redistributive policy.
A way of mitigating some of the problems associated with a traditional cost-benefit-analysis, is to complement the welfare analysis with a study of the acceptability of the different policies that is evaluated. A simple measure of acceptability is to estimate the share of individuals that benefits from the policy and compare this with the share of individuals that loses from the same policy. Especially if the analysis is included in a larger package of many different policies, it can be important to assess the winners and losers of the whole package.

The restrictions imposed on the utility functions in a standard cost-benefit-analysis can be especially problematic when evaluating the welfare effect of a policy in a heterogeneous population. An example of this is given in paper II that analyzes the welfare effect of a road toll in a population with heterogeneous value of time. Empirically we know that people make different trade-offs between time and money. From a utilitarian perspective this implies that the ration between the marginal utility of time and the marginal utility of income differ between people. In order to enable interpersonal comparisons of utility, a common assumption in a standard cost-benefit-analysis is that the whole population has the same marginal utility of income. This simplifies interpersonal comparisons of utility since it allows us to measure welfare changes directly using willingness to pay without applying distributional weights to compensate for marginal utility differences. However, this also implies that the marginal utility of time will differ in a population with heterogeneous value of time.

The chosen unit for comparison (money, time or utility etc.) can have large policy implications for a road pricing policy in a population with heterogeneous value of time. Mayet and Hansen (2000) show that the optimal toll level depends on how the aggregated social welfare is calculated. When aggregating the utility of individuals with different value of time, different outcomes is reached depending on if all individuals are assumed to have equal marginal utility of income or equal marginal utility of time. "If money is used, we implicitly ascribe the same marginal utility of income to all individuals, and thus set tolls at higher levels to accommodate those with high values of time. If welfare is expressed in time units, then we assume that all individuals have the same marginal utility of time (and thus that those with a high value of time have a lower marginal utility of income), leading to lower tolls." (Mayet and Hansen 2000).

This is an argument for not using differentiated values of time distributions in welfare analyzes that compare effects on different travelers groups or different geographical regions. If we assume that commuters in one region has higher value of time than commuters in another region, this can (from a utilitarian perspective) introduce a bias in the welfare analysis given that the commuters have diminishing marginal utility of income. However, not using differentiated values of time can introduce problems when analyzing the effect of road tolls in congested areas since one of the features of a road toll is that it sorts travelers according to their value of time. It is nevertheless important not to ignore the moral philosophical assumptions and value judgments that the welfare analysis rests upon.
3 Political acceptability of road pricing policies

Urban road pricing policies are characterized by conflicting interests between different stakeholders and different geographical areas. This can both make road tolls unpopular and lead to inefficient tolling. The difficulty of achieving political support for transport pricing has led some authors to discuss the tension between acceptability and economic efficiency in terms of a paradox, where efficient instruments in the transport sector are not acceptable while acceptable policies in general are less efficient (Steg 2003). Other authors have approached the problem using political equilibrium models and tried to measure the cost of acceptability in terms of reduced efficiency (Westin et al. 2012) in a similar way as the traditional equity-efficiency trade off (Mayeres and Proost 2001). A common assumption in the approach is that a reform only will be accepted if a sufficiently large majority of the voters gain (or do not lose) on the policy compared to the initial situation (De Borger and Proost 2011). These models are often based on the assumption that people are primarily concerned about their own well-being and not the well-being of the society as a whole.

To illustrate how political constraints can make policy makers propose inefficient tolling strategies we consider a situation where a regional highway passes through the city center of a small city. The traffic in the city center gives rise to a local environmental externality for the inhabitants of the city. To solve this problem, the local decision makers consider introducing road tolls on the city road or building a tolled bypass or both.

In the analysis, we distinguish between two groups of travelers: local travelers who use the road to get to the city center, and transit travelers who use the road only to pass through the city. The transit travelers can therefore choose between the bypass (if built) and the city road whereas the local travelers only can choose the city road. We assume that Wardrop’s principle holds, that is, we assume that Wardrop’s principle holds, that is, we assume that all transit travelers only choose the road with the lowest generalized cost. We further assume that the generalized user cost before the toll of travel on the bypass $g_B$ is lower than the generalized cost before the toll of travel on the city road $g_C$ and that there is no congestion on either of the roads.

An efficient policy would be to use marginal cost pricing on all roads. For the bypass to increase total welfare, the annualized construction cost must be lower than the value of the time gain for the transit travelers with efficient pricing. In the following example we assume that the bypass, with efficient tolling, increases total welfare. Assuming that the marginal external cost of the city road is constant and equal to $MEC$ and that the bypass is built outside populated areas and hence has no external cost, an efficient tolling policy would be to set the toll on the city road equal to its marginal external cost and no toll on the bypass, i.e. $t_C = MEC$ and $t_B = 0$.

Economic efficiency in this situation therefore implies that the travelers on the city road pay for the construction of the bypass even though they do not directly benefit from it, whereas the travelers on the bypass benefit from the shorter travel time for free. This optimal tax result can therefore be counterintuitive for many people.
When analyzing the political acceptability of road pricing empirically, redistributive concerns are often found to be important for the public opinion on different road pricing policies (Verhoef, Nijkamp and Rietveld 1997). We assume that an individual accepts a policy if he or she does not lose on it compared to the initial situation. A measure of political acceptability is then the share of citizens that accepts a given policy. A problem with this definition is that it in general requires detailed information about each individual's preferences and how the net revenues from the policy are recycled.

A simpler approach to study political acceptability is to assume that each citizen belongs to one or more interest groups where each interest group is modeled as a representative agent. The interest groups represent different “hats” that an individual citizen can have. An interest group is assumed to accept a policy in the same way as an individual. The idea is that if an influential interest group (car drivers, environmentalist or tax payers) is worse off from a policy, they can oppose the policy and try to stop the decision. We can then study political acceptability by analyzing which interest groups that accepts a given policy. An important assumption in the analysis is that the policy makers do not have access to any compensatory transfers between the interest groups. If able to fully compensate the losers, the trade-off between efficiency and acceptability would disappear since the policy maker could redistribute welfare to make every welfare improving policy a Pareto improvement.

We consider four different interest groups. The first two interest groups are local travelers and transit travelers. The third interest group consists of residents living near the city center that is affected by the negative externality from the city road. The last interest groups are the tax payers that pay for the construction of the bypass and receive the toll revenues in return.

In the example, the efficient policy was to set a road toll on the city road to internalize the marginal external cost of driving in the city center and construct an un-tolled bypass. This means that welfare decreases for the local travelers who pay the toll on the city road whereas welfare increases for transit travelers who benefit from the lower generalized cost on the bypass. Since both they bypass and the road toll decrease traffic on the city road, local residents also benefit from the policy. The effect on the tax payers finally depend on whether the toll revenues from the city road can cover the cost of constructing the bypass. The efficient policy is therefore likely to raise opposition from local travelers who lose on the policy.

A cautious decision maker, concerned about the political acceptability of the proposed policy would therefore hesitate to propose efficient tolling as a way of reducing the environmental externality in the city center. Instead a cautious decision maker could try to find a policy that maximizes political acceptability or a policy that is guaranteed to make no one worse off? Since the individuals can belong to one or more interest groups and the composition of the individuals is unknown, the only way the decision maker can guarantee that no individual is worse off is to find a policy that does not reduce welfare for any interest group. Can we find such a policy?

Since a toll on the city road reduces welfare for the local travelers that cannot use the bypass, a Pareto-improving solution requires that $t_c = 0$. From the transit travelers’ point of view, any toll level on the bypass can be accepted since they can continue to use the city road if the toll on the
bypass is too high, i.e. $t_B > 0$. As long as the transit travelers use the bypass, the city environment is improved and the residents living near the city center are better off even without a toll on the city road. Hence the toll on the bypass must satisfy $t_B \leq g_C - g_B$. Finally the taxpayers are better off if the collected revenues cover the cost of the bypass. If this is true, then there exists a solution that does not make any interest group worse off.

### 3.1.1 The cost of acceptability

In previous section we saw that political acceptability and economic efficiency produced very different results regarding the use of road user charges. The economically efficient solution was to put a toll on the city road to internalize the marginal external cost where we do not want people to drive and no toll on the bypass where we want people to drive. From an acceptability point of view (or the view of a local politician wanting to avoid making influential interest groups angry), the tolls should instead be placed the other way around: no toll in the city to keep local travelers happy and a positive toll on the bypass to finance the construction of the bypass. The result is in line with the road pricing literature where a trade-off between efficiency and acceptability often is found (Verhoef, Nijkamp and Rietveld 1996).

Acceptability in this context therefore has a cost in terms of lower efficiency. The cost of acceptability measured as the welfare loss compared to efficient tolling is illustrated in Figure 10.

![Figure 10: The cost of acceptability measured as efficiency loss compared to efficient toll levels.](image)

The cost of acceptability in terms of lower efficiency is shown by the two triangles. Since the road toll on the city road is zero, only the negative environmental externality from the transit travelers is reduced. This means that even though the environment is improved compared to the initial situation, the welfare gain is not as large as it could have been. Since a toll on the bypass is needed to finance the construction cost in order not to hurt the taxpayers, this creates an additional efficiency loss among the transit travelers.
An important assumption for the existence of a Pareto improving policy is that there are no rural residents living near the bypass. Adding a marginal external cost of bypass traffic to the model would motivate a toll on the bypass to internalize the negative externality. This would also imply that by just building a bypass, residents living near the bypass are worse off. Instead of trying to find a Pareto improving policy we could study political acceptability as the outcome of a majority voting process.

4 Conclusions
How should we analyze the welfare effect of a road pricing policy or a transport investment in a distorted economy? Since a transport market often has large feedback externality effects into other distorted markets, the total welfare effect of a road pricing policy both depends on effects in the transport market and on effects in other interconnected markets. This makes it important to also consider wider economic benefits and losses in the analysis.

A market that has been given special attention in the road pricing literature is the labor market. Since urban congestion often is associated with morning and afternoon peak-hour commuting, concerns have been raised that a road toll that increases the monetary cost of commuting can have a negative effect on labor supply in a similar way as a raised income tax. When explicitly taking into account that commuters have different value of time, this result can change. Since a road toll primarily price out travelers with a low willingness to pay, this implies that the remaining commuters with a higher willingness can drive (and work) more. Disregarding equity considerations, the road toll therefore leads to a more efficient use of the available road space. The effect on total labor supply is therefore ambiguous and depends on, among other things: the flexibility in the labor market, availability of alternative travel modes, the share of non-work-related traffic, traveler heterogeneity and the commuters’ ability to adapt to the new situation.

An argument often found in the road pricing literature is that the welfare effect of a road toll in a distorted economy critically depends on revenue use. The argument is related to the double dividend debate, i.e. if the toll revenues are used to cut distortionary income taxes rather than returned in a lump-sum, a double dividend can arise where welfare increases both in the transport market and in the labor market. The argument is however difficult to generalize since the relative efficiency of different revenue recycling policies strongly depends on both model assumptions regarding the efficiency of the initial tax system and that the recycling policy is additive to the business-as-usual situation. Without a thorough analysis of the initial tax system and the political situation surrounding the decision, it might therefore make more sense to analyze the effect of a road toll “as if” the policy instruments in the initial constrained no-toll situation are chosen on welfare maximizing principles. By starting from a constrained optimum, we can isolate the welfare effect of the road toll from the effect of a general adjustment of the tax system.

Urban road pricing is characterized by conflicting interests between stakeholders and constituencies from different geographical areas and regions. Road pricing policies are therefore not only likely to raise considerations about fairness and equity at a structural or principal level, they are also likely to trigger opposition from unfavored groups. Since the welfare gains from a road toll to a large extent come in the form of revenues, revenue use can be important for both the
distributional impact and the acceptability of a road pricing policy since it enable the decision makers to compensate the losers. For the political process to result in an efficient pricing policy, the decision making process must balance different interests against each other. If unable to do this, conflicting interests and acceptability constraints protecting certain interest groups can therefore lead to inefficient tolling.

Instead of focusing on political acceptability for explaining why efficient transport pricing policies are so seldom used, it might therefore be more fruitful to analyze political feasibility and investigate how the political process can resolve the inherent conflicting interests associated with efficient transport pricing and revenue recycling.

That is, however, a topic for another thesis.
References


To Ida, Stella and Konrad.