Chapter 2

On revenue recycling and the welfare effects of second-best congestion pricing in a monocentric city

2.1. Introduction

This chapter examines the design of congestion taxes in a monocentric city with pre-existing labor taxes. Labor taxation is highly relevant in this context, since it reduces labor supply, commuting flows and, therefore, the level of congestion externalities in the transport system. With a pre-existing tax on labor income, the traffic level in an untolled equilibrium might already lie below its optimal level, even when the absence of road charges makes the generalized price of a commuting trip fall short of its marginal social cost (Parry and Bento, 2001). Then, a policy intervention that introduces road tolls but leaves the labor tax unaffected (for example because the revenues are returned lump-sum) may be expected to produce a decline in social welfare rather than the increase that is hoped for.

This fundamental issue has received only limited attention in the transport economics literature, although similar questions have spawned several contributions in the literature of environmental economics (see, for example, Bovenberg and De Mooij, 1994; Parry, 1995; Goulder, 1995a; Bovenberg and Goulder, 1996; Parry and Bento, 2000). An exception is the paper by Parry and Bento (2001). They conclude that, in order to increase welfare, road toll revenues must be used to reduce the distortionary tax. In that paper, as in a later contribution aiming to investigate other critical distortions within the transport system (Parry and Bento, 2002), commuting distance is assumed to be exogenous, since the model has no spatial dimension. Therefore, commuters react to the introduction of a road tax by adjusting their labor supply, but not their commuting distance. While this assumption might be realistic in the short run, one might question to what extent the above policy recommendation is valid in the long run, i.e. as commuters are able to relocate.

More importantly, the lack of a spatial dimension implies that questions involving the differentiation of road taxes over space cannot be studied. This prevents the evaluation of second-best congestion pricing schemes such as the cordon toll (Mun et al. 2003; 2005) or a flat kilometer tax (Sullivan, 1983) which are more realistic in practice, but deviate from the first-

---

15 The chapter is based on joint work with Erik T. Verhoe and Jos van Ommeren published in the Journal of Urban Economics (Tikoudis et al., 2015b). The authors would like to thank the two anonymous referees for valuable comments and suggestions in the previous versions of the paper. Also, participants in the 6th Kuhmo-Nectar conference (Berlin, July 2012), in the Urban Economics session of NARSC conference (Ottawa, November 2012), the attendants of the seminar in Tinbergen Institute and the participants in the meetings of i-PriSM project for useful comments and feedback. We are indebted to the Netherlands Organization for Scientific Research (NWO) and to the ERC (AdG Grant #246969 OPTION) for financial support.

16 For a more general investigation of marginal tax reforms, see Mayeres and Proost (2001).
best. Furthermore, it prevents one from analyzing optimal spatial differentiation of a revenue-raising labor tax. This issue is relevant since one expects that the optimal labor tax is not space-invariant; it varies over space because the labor supply elasticity, as well as the marginal utility of income vary over space. In addition, suboptimal pricing in public transport is ignored in Parry and Bento (2001). Thus, their model may understate the welfare gains of policies that use the road toll revenue to subsidize the providers of public transport when operating under increasing returns.

In this chapter, we develop a monocentric city model in which household location is endogenous and both residential density and labor supply vary over space.\(^{17}\) We combine the insights of the non-spatial labor supply model by Parry and Bento (2001) with the monocentric city model by Verhoef (2005), which allows city size and commuting distance to be endogenous. Our aim is to identify the optimal policy, \textit{i.e.} a combination of a road toll scheme and a revenue recycling program, taking into account the equilibrium impacts of transport policies, as well as (in an extended model) the presence of a suboptimally priced public transport alternative.\(^{18}\) To some extent, the model also resembles Bento et al. (2011), which is the only application of a double-dividend tax reform in a monocentric city that we are aware of. In that paper, an environmental policy motivated by sprawl-style externalities (a spatially-uniform development tax that increases open space) generates revenues in order to reduce the level of a preexisting spatially uniform property tax. However, in contrast to a spatially uniform development tax, this chapter considers a continuous-in-space road tax, something that prevents us from deriving an analytic expression for the welfare change of a revenue-neutral tax swap in the way Bento et al. (2011) do.

To facilitate comparability of our numerical results with earlier work, we calibrate the model’s parameters in line with Parry and Bento (2001) and Verhoef (2005). We compare the welfare gains (or losses) for a range of road toll schemes and revenue recycling programs. The welfare changes are reported in a relative manner, for instance as in Parry (2002), as well as in monetary terms, by computing the compensating variations for the representative household. Since the numerical results are likely to be sensitive to the model’s parameterization, we perform sensitivity analyses with respect to the initial level of labor tax and the elasticity of substitution in consumption.

One of our main conclusions is that a \textit{space-varying road tax} is not desirable as a congestion management policy only: with a tax-distorted labor market, it may be welfare improving to spatially differentiate taxes even in the absence of congestion. There are two reasons for this. The first is that the elasticity of labor supply exhibits variation over space, while the labor tax is independent of residential location. Bento \textit{et al.} (2011) discusses a similar

\(^{17}\) Throughout the chapter, we assume that the labor market is competitive. For the impacts of congestion tolls in a wage bargaining model, see De Borger (2009).

\(^{18}\) The present chapter focuses on the efficiency gains of road pricing for a representative household in a monocentric city. Although revenue recycling is closely related to the various equity considerations of road pricing (Langmyhr, 1997), the distribution of the total gains among heterogeneous households in the context of a monocentric city is another future research challenge. See Ramjerdi \textit{et al.} (2008) for an empirical approach of this issue.
mechanism when deriving the marginal excess burden of a property tax in a monocentric city model. Thus, a space-varying road tax might improve the performance of an inefficient labor tax system by functioning as a spatial correction of a suboptimal labor tax, allowing it to vary with the elasticity of labor supply. This finding is in line with Parry and Bento (2000) and Bento and Jacobsen (2007), where an environmental tax is shown to be part of the optimal tax system, because it corrects failures of the existing tax system. Consequently, the optimal tax may be above the Pigouvian level.

The spatial differentiation of taxes is thus in line with the standard Ramsey rule for minimizing the distortionary impacts of taxes used to raise revenues. Our spatial setting, however, reveals a second reason to differentiate taxes by residential location. This concerns the Mirrlees rule, which states that taxes should be lower where the marginal utility of income is higher (Mirrlees, 1972; Wildasin, 1986). In our model (and probably in other spatial configurations as well), these two arguments appear to be working in opposite directions. Labor supply elasticity falls with commuting distance, as equilibrium labor supply falls; while the marginal utility of monetary income rises, as monetary income falls as well.19 As a result, the Ramsey-Mirrlees component, which we define as the deviation of the optimal road tax from the marginal external cost of congestion (conditional on a given labor tax), can portray complex, even non-monotonic patterns over space.

This chapter is connected to two streams of literature. The first stream concerns the presence of various constraints (e.g. tolls on a subset of road links) or other sources of inefficiency within the transport system.20 One such inefficiency regards suboptimal pricing in public transport, for instance average cost pricing combined with substantial fixed costs. To juxtapose labor tax cuts against public transport subsidies, we expand the base model to account for a public transit alternative, which operates with fixed costs. The numerical results show that the optimal type of revenue recycling depends on the degree of inefficiency in public transport pricing. For cities served by unsubsidized operators with significant fixed costs, it is optimal to return the road toll revenue in the form of a public transport subsidy, contrary to the results by Parry and Bento (2001).

At the same time, the chapter contributes to the double-dividend stream of literature by focusing at a pre-existing source of inefficiency outside the transport system. In particular, this stream focuses on the occurrence of failures, most often tax-induced distortions accompanied by significant marginal excess burdens in markets that interact with the transport market to be regulated. The labor market is one such market: when it does not operate efficiently, there is a divergence between the (inverse) demand for commuting trips and the marginal social benefits.21

---

19 Note that the latter pattern may be reversed when income heterogeneity is introduced, with higher incomes locating further from the CBD (Brueckner et al., 1999).

20 For instance, optimal tolls under unpriced alternative routes in the network have been investigated by Verhoef et al. (1996) and Small and Yan (2001). Kidokoro (2010) expanded the above work by considering revenue recycling within the transport system, i.e. capacity expansion and public transport subsidies.

21 The other primary factors of production exhibit significant interactions with the transport market, and can play the role of labor as well. Furthermore, inefficiencies outside the transport system may include various regulations in the housing market: rent controls and density regulations (e.g. height restrictions and zoning), which may result in a
In line with this, our results will show that optimal road charges fall short of the Pigouvian tolls when road toll revenues are recycled lump-sum. However, in our numerical model they remain strictly positive at least at some parts of the city, because the tax-interaction effect (i.e. the efficiency loss in the labor market caused by a marginal increase in the road toll) diminishes with distance from CBD and is outweighed by the Pigouvian effect (i.e. the welfare benefits from a marginal reduction of externality) for residents living at the more distant parts of the city. Therefore, even with lump-sum revenue recycling, welfare gains are possible through road pricing at the suburbs of a monocentric city, a result that can only be obtained by a spatial model.

The double-dividend stream of literature sheds light on the interplay between an externality-correcting tax (e.g. a road toll) and a pre-existing distortionary tax (e.g. a tax on labor) by investigating revenue-neutral and marginal swaps between them. Apart from the Pigouvian and tax interaction effects (just defined above), these swaps produce a revenue recycling effect, i.e. a welfare gain in the initially distorted (labor) market, due to the lowering of the (labor) tax. This lowering is financed by the additional revenue the marginal increase of the externality tax (on road use) generates. If the revenue recycling effect dominates the tax interaction effect, a so-called strong double dividend emerges, and this is a case for setting the externality tax above its Pigouvian level. However, most of the earlier double-dividend literature suggests that this will not occur (Bovenberg and Mooij, 1994; Parry, 1995; Goulder, 1995b; Bovenberg and Goulder, 1996).

Regarding this stream, the chapter contributes in highlighting a series of complications arising when applying the standard double-dividend framework in a less abstract context, like the one considered in this study. The first complication arises when the tax bases of the two taxes involved in the revenue-neutral swap overlap perfectly. In particular, following the literature by assuming that daily labor supply is constant and that adjustments occur only through the number of days worked, an actor’s payments on the labor tax and on the road tax both vary in proportion with labor supply, given the residence location. The consequence is that the optimal tax schedule can be identified uniquely only up to the level of an optimal spatially differentiated tax: the sum of the common spatially undifferentiated labor tax and the road tax paid by an individual at that location. Adding one dollar to the labor tax and subtracting it from all road taxes would leave the equilibrium unaltered. Given that it is therefore not possible to uniquely identify the levels of the road taxes and the level of the labor tax in the optimum, it is also not possible to unambiguously determine whether or not the road tax exceeds the Pigouvian level, as given by the marginal external congestion cost. Moreover, it is not possible to unambiguously compute the Pigouvian, tax-interaction, and revenue-recycling effects of marginal tax changes, to assess the influence of the pre-existing distortionary tax upon the optimal design of an externality-correcting tax, as is

suboptimal allocation of space across economic agents and activities (Glaeser and Luttmer, 2003). Road pricing affects the private and social benefits and costs of land use. Therefore, it can indirectly alter the magnitude of welfare losses in a distorted land/housing market.

22 Recent papers have demonstrated that it is possible to get a double dividend if the tax system is not optimal to begin with. Two key references here are Parry and Bento (2001), who consider tax deductions, and Bento and Jacobsen (2007), on Ricardian rents. As a consequence, the optimal level of the environmental tax is actually above its Pigouvian level.
common in the double-dividend literature. The analytical expressions for these different effects contain both the labor tax and road tax as arguments, and not only the sum of these taxes, whereas it is the sum of the taxes which determines the equilibrium. Hence, the size of the different effects would be based on an arbitrary decomposition of the spatial tax into a labor and road tax component.

The other type of complication highlighted in this chapter is that we do not deal with one tax instrument addressing an externality, but rather a (spatial) continuum of local taxes. Even if these local taxes could be unambiguously decomposed into an externality tax and a labor tax, there would be tax interactions not only between those two types of taxes as such, but also between taxes at different locations, which would further complicate the unambiguous definition and interpretation of the three effects normally distinguished in the double-dividend literature.

Due to the above complications, the chapter discusses the impact of the pre-existing tax (and the underlying revenue target) on second-best road pricing in a different and — we believe — more intuitive way.\(^{23}\) We focus on the difference between the spatial variation of the second-best optimal tax (under the constraint implied by the revenue objective) on the one hand, and the spatial variation of marginal external congestion cost on the other. So we are interested to decompose the spatial variation in this optimal tax into a Pigouvian component and a Ramsey-Mirrlees component.\(^{24}\)

The sign of the latter component is easy to interpret: positive values of the Ramsey-Mirrlees component reflect that the local tax is higher than what would be justified by the spatial variation on marginal external congestion costs alone. Likewise, negative values reflect that the local tax is lower than what would be justified by spatial variation on marginal external congestion costs alone.

Dealing with the above methodological barriers, our numerical model demonstrates that there is no guarantee that the optimal space-varying tax is monotonic in distance from CBD, unlike the total marginal external congestion cost as a function of residential location. We find that, at the optimum, the tax scheme is slightly non-monotonic over space (97% of its gains can be captured by a monotonic tax), reflecting the non-monotonic pattern of the Ramsey-Mirrlees component. That is, the tax peaks at a level 2% higher than the level at the fringe, at an interior location 75% of the distance between the CBD and the city fringe.\(^{25}\) The relative efficiencies of

\(^{23}\) One can rewrite the conventional expressions from the double-dividend literature to account for the fact that in our framework the labor tax and the road tax are strictly additive at a given location, and compute space-varying indicators of the Pigouvian, revenue-recycling and tax-interaction effects. These are presented in Appendix 2.B.

\(^{24}\) The Pigouvian component we are interested in is closely related to the Pigouvian effect as used in the double dividend literature. Similarly, the Ramsey-Mirrlees component is closely related to the combination of revenue recycling and tax interaction effects. However, there is an important difference in interpretation. Here, both terms refer to components in the second-best optimal tax, whereas in the double dividend literature the effects are expressed as marginal changes in welfare from marginal changes in tax levels. As a consequence, the spatial pattern of the Pigouvian effect expressed in terms of the double dividend terminology (see Appendix 2.B) differs qualitatively from the spatial pattern of the Pigouvian term that we are interested in.

\(^{25}\) Roughly, this non-monotonicity occurs from a Ramsey part of the component suggesting that the tax should be higher with distance from CBD (assuming that the elasticity of labor supply falls with distance from CBD), and a
archetype second-best road pricing schemes – under the same overall revenue constraint – are high: 84% for cordon toll and 70% for flat kilometer tax. Still, these relative gains are somewhat lower than those reported by Verhoef (2005) for a first-best environment without labor taxation (where relative efficiencies are 90% for the cordon toll and 91% for the flat kilometer tax).

The structure of the chapter is as follows. Section 2.2 introduces the analytical model in its general form. Section 2.3 presents the calibration, policy implications and sensitivity analysis in the unimodal framework. Section 2.4 expands the analysis in a bimodal setting. Section 2.5 summarizes and concludes.

2.2. Model

2.2.1. Households

Consumption and labor supply are derived from the utility maximization problem of a representative household, which can be located anywhere within a linear, closed monocentric city, i.e. a city with a single central business district (CBD) and exogenous population $N$.26 Household location is, therefore, a point $z$ which belongs to the open interval $(0, \bar{z})$, where $\bar{z}$ is the distance between the CBD and the endogenously determined city fringe. For a given residential location, $z$, the household maximizes the CES utility function:

$$U_z = \left[ (\delta_y y_z)^\rho + (\delta_s s_z)^\rho + (\delta_f T_F z)^\rho \right]^{\frac{1}{\rho}},$$  \hspace{1cm} (2.1)

where the parameters $\delta_y$, $\delta_s$ and $\delta_f$ correspond to the consumption of a composite good, $y_z$, housing space, $s_z$, and leisure time, $T_{Fz}$, and $\rho$ is the scale parameter that determines the elasticity of substitution, $\sigma$, via $\sigma = 1/(1 - \rho)$.27 The total time endowment, $T$, is spent on commuting, $T_{cz}$, working, $T_{Lz}$, and leisure, $T_{Fz}$:

$$T = T_{cz} + T_{Lz} + T_{Fz}.$$  \hspace{1cm} (2.2)

Labor supply is inelastic throughout a working day, thus the working day is of fixed duration, $t_L$. Every trip to work requires $t_z$ units of time. For $D_{Wz}$ working days the time constraint becomes:

$$T = D_{Wz} t_z + D_{Wz} t_L + T_{Fz}.$$  \hspace{1cm} (2.3)

Normalizing the duration of the working day, $t_L$, to one, the above constraint becomes:

---

26 The model and its exposition in Section 2.2 closely follows, and draws from, that in Verhoef (2005). The linear monocentric city model can be derived as the limiting case of the circular city as the fraction of land available for development approaches zero. The amount of land available for housing between any two arbitrary points depends only on the distance between them and not their exact locations (as it is the case with the circular city). In other words, the linear city can be seen as a corridor of fixed width, but endogenous length.

27 We use a Constant Elasticity of Substitution function to avoid behavioral restrictions imposed by a Cobb-Douglas specification, which implies a unitary elasticity of substitution.
\[ T = D_{Wz} (1 + t_z) + T_{Fz} \iff D_{Wz} = \frac{T - T_{Fz}}{1 + t_z}. \]  

The net wage per working day is defined as the difference between wage, \( w \), the labor tax, \( \tau_L \), and the road toll for the full commuting trip from \( z \) to the CBD, \( \tau_{Rz} \). The full income of the household at distance \( z \), \( M_z \), is the maximum income which can be realized when leisure time is zero. That is:

\[ M_z = B + \frac{(w - \tau_L - \tau_{Rz})}{1 + t_z} T, \]  

where \( B \) denotes a lump-sum transfer from the government to the household, independent of its location. This full income can be used to buy back leisure at its shadow price, \( (w - \tau_L - \tau_{Rz})/(1 + t_z) \), and for the consumption of the composite good and residential space.

The budget constraint can then be written as:

\[ B + \frac{(w - \tau_L - \tau_{Rz})}{1 + t_z} T = \frac{(w - \tau_L - \tau_{Rz})}{1 + t_z} T_{Fz} + p y_z + r_z s_z, \]  

where \( p \) is the price of the composite good and \( r_z \) the rental price per unit of space at location \( z \).

Maximizing (2.1) subject to (2.6), and defining \( \chi = \rho/\rho - 1 \) yields the Marshallian demand functions for the composite good, space and leisure time respectively, at location \( z \):

\[ y_z^* = M_z \frac{(p/\delta_y)^{-\sigma}}{(p/\delta_y)^{\chi} + (r_z/\delta_s)^{\chi} + \left[ w - \tau_L - \tau_{Rz} \right]^{\chi}}, \]  

\[ s_z^* = M_z \frac{\left( r_z/\delta_s \right)^{-\sigma}}{\left( p/\delta_y \right)^{\chi} + \left( r_z/\delta_s \right)^{\chi} + \left[ w - \tau_L - \tau_{Rz} \right]^{\chi}}, \]  

\[ T_{Fz}^* = M_z \frac{\left[ w - \tau_L - \tau_{Rz} \right]^{-\sigma}}{\left( p/\delta_y \right)^{\chi} + \left( r_z/\delta_s \right)^{\chi} + \left[ w - \tau_L - \tau_{Rz} \right]^{\chi}}. \]

Finally, substituting (2.7), (2.8) and (2.9) into the objective function yields the indirect utility as a function of \( z \):

---

\(^{28}\) The labor tax is assumed to be distortionary, i.e. its optimal value is zero. The implicit assumption that the actual labor tax exceeds its optimal level captures the underlying view that marginal income taxes are too high in many countries. This assumption is critical for the results derived in Sections 2.3.2 and 2.3.3.
\[ V_z = \left[ B + \left( \frac{w - \tau_L - \tau_{Rz}}{1 + t_z} \right) T \right] \left[ \left( \frac{p}{\delta_y} \right)^x + \left( \frac{\tau_z}{\delta_\delta} \right)^x \right]^{x^{-1/x}} \]

(2.10)

2.2.2. Firms

The competitive, representative firm is located in the CBD and produces a homogenous output, \( Q \), under constant returns to scale, with labor as the only input:

\[ Q = AL, \quad (2.11) \]

where \( A \) denotes the marginal productivity of labor. The zero profit condition restricts the wage to be proportional to the price of the composite good, so:

\[ w = Ap. \quad (2.12) \]

2.2.3. Road technology and externalities

Commuting from any given location \( z \) to the CBD is taking place through a single road, used by all households. Letting \( n_z \) denote the residential density at location \( z \), we can write the travel time per unit of distance at location \( z \) as:

\[ t_0 + t_1 \int_{z}^{\bar{z}} n_\zeta D_{W\zeta} d\zeta, \quad (2.13) \]

i.e. the sum of the free-flow travel time per unit of distance, \( t_0 \), and a term that represents the congestion delay caused by the aggregate traffic flow at \( z \).

This flow contains all commuting trips by households located between \( z \) and the city limit, \( \bar{z} \). Multiplying the aggregate flow with a sensitivity parameter, \( t_1 \), yields the time delay per unit of distance at each point \( z \). Integrating (2.13) over the interval \((0, z)\) yields the commuting time for the household at \( z \):

\[ t_z = z t_0 + t_1 \int_{0}^{\zeta} min\{z, \zeta\} n_\zeta D_{W\zeta} d\zeta. \quad (2.14) \]

An additional commuting trip generated by a household at distance \( z \) increases the travel time of a commuter located at \( \zeta \geq z \) by \( t_1 z \), and the travel time of a commuter located at \( \zeta \leq z \) by \( t_1 \zeta \).

Multiplying this delay with the shadow value of time at \( \zeta \), \( (w - \tau_L - \tau_{Rz})/(1 + t_\zeta) \), and labor

\[ \text{Therefore, the essence of congestion is captured by assuming that the travel time per unit of distance increases with the number of vehicles passing that point; the linear form is chosen for analytical convenience.} \]

\[ \text{Note that in (2.13), } z \text{ refers to a location along the road and } \zeta \text{ to household locations, while in (2.14) and (2.15) } z \text{ refers to household location.} \]
supply at \( \zeta \), \( n_\zeta D_{W\zeta} \), provides a measure for the marginal external cost of an additional unit of labor imposed by the commuter at \( z \) to the commuter at \( \zeta \).\(^{31}\) Integrating over the interval \((0, \bar{z})\) yields the *marginal external cost of congestion*, generated by the household at \( z \):

\[
\text{mecc}(z) = t_1 \int_0^\bar{z} \min\{z, \zeta\} n_\zeta D_{W\zeta} \left( \frac{w - \tau_L - \tau_R \zeta}{1 + t_\zeta} \right) d\zeta.
\] (2.15)

2.2.4. Public budget

We assume that the government runs a balanced budget by imposing road pricing schemes (discussed below) and by returning the revenue from labor and road tax taxes, as well as the aggregate excess land rents (discussed below). The total rent paid by the households is equal to the sum:

\[
R = \bar{z} r_{\bar{z}} + \int_0^\bar{z} (r_\zeta - r_z) \, d\zeta,
\] (2.16)

where \( R_a \) denotes the cost of acquiring land from an absentee land owner at the opportunity cost \( r_\bar{z} \) and \( R_e \) the *aggregate excess land rent*, which is returned to the consumers in a lump-sum manner. This assumption ensures that: i) the city cannot expand without cost, ii) there is no additional failure in the land market and iii) the excess rents remain within the urban economy.\(^{32}\) Similarly, total tax revenue (labor tax revenue, \( G_L \), plus toll revenue, \( G_R \)) is:

---

\(^{31}\) The volume delay function is the same used in Verhoef (2005), and in line with those used in other monocentric city models, for instance by Mun et al. (2003) and Brueckner (2007).

\(^{32}\) The absentee landlord is a conceptual device used to support the assumption of the exogenous opportunity cost of land. The role of the absentee landlord is similar, but not identical (see below why) to the role that any industry located in the *rest of the world* (hereafter, \( ROW \)) would play in a general equilibrium model. That is, the industry in \( ROW \) would produce a good that would be supplied in a perfectly elastic manner, i.e. at an exogenous price (to all price-taker zones/countries modelled explicitly). Here, the behavior and welfare of the supplier(s) of land are not modelled explicitly, exactly as the behavior of an industry located in \( ROW \) (i.e. its cost and conditional factor demand functions) would not be modelled explicitly. In line with general equilibrium settings where closure (total value preservation) conditions impose that the total value of imports from \( ROW \) should be equal to the total value of exports to \( ROW \), we here impose the macroclosure condition in (2.29), which states that the total opportunity cost of land (the imported entity) should be equal to the total value of the foregone consumption of composite good produced by the residents of the city.

However, the above parallel is not complete: the imported factor (in this case land) is heterogeneous in that every unit of land is unique: it encapsulates a unique level of accessibility (generalized commuting cost) to the CBD. Therefore, to mimic a competitive outcome on the land market, the government acquires land from \( ROW \) at its world price (the agricultural rent representing the opportunity cost of land), sells every additional unit in a price that equals the marginal willingness to pay for it, and returns the total surplus (i.e. the aggregate differential/excess land rents) to the (exogenous) population of the city in a lump-sum manner. This ensures that the aggregate differential (or excess) land rents do not leak out of the urban economy.
\begin{equation}
G = G_L + G_R = \int_0^{\hat{z}} n_\zeta D_{W_\zeta} \tau_L d\zeta + \int_0^{\hat{z}} n_\zeta D_{W_\zeta} \tau_{R_\zeta} d\zeta . \tag{2.17}
\end{equation}

Note that \( \tau_{R_\zeta} \) is the toll per trip (not per unit of distance) paid by a household located at \( z \). In this chapter, every toll policy consists of two components: i) a functional form for \( \tau_{R_\zeta} \), which assigns a unique road charge (full trip) to each location \( z \), and ii) a description of how \( G \) is returned to society. From now on, we are going to refer to those policy components as \textit{road toll scheme} and \textit{revenue return scheme} respectively.

\subsection*{2.2.4.1. Road toll schemes}

A first possible tolling scheme is the \textit{Pigouvian toll}, which equates tolls to the marginal external costs of congestion, as defined in (2.15):

\begin{equation}
\tau_{R_\zeta} = \text{mecc}(z) = t_1 \int_0^{\hat{z}} \min\{z, \zeta\} \ n_\zeta \ D_{W_\zeta} \left( \frac{w - \tau_L - \tau_{R_\zeta}}{1 + t_\zeta} \right) d\zeta . \tag{2.18}
\end{equation}

The road user located in \( z \) therefore pays exactly the marginal external cost she imposes upon the rest of the road users when increasing labor supply by one working day. In absence of other distortions (\textit{i.e.} when \( \tau_L = 0 \)) such a Pigouvian toll constitutes the \textit{first-best policy}. We assume that the implementation of a Pigouvian toll is feasible: the government can perfectly monitor commuting activity and households accept the implementation of a sophisticated road toll scheme, such as the one in (2.18).\textsuperscript{33}

For \( \tau_L > 0 \) the pricing scheme in (2.18) would no longer be optimal: while the generalized price would then be equal to the marginal costs on the road, the same would not be true for the labor market. The optimal toll scheme takes this into account. Such a scheme would allow \( \tau_{R_\zeta} \) to vary in space, but the optimal road charge would deviate from the Pigouvian charge in (2.18), to account for the distortion in labor market. We refer to this scheme as a \textit{varying-kilometer tax}. Practical second-best schemes, however, may have additional constraints. One such scheme we consider is the \textit{cordon toll}, under which the charging function becomes:

\begin{equation}
\tau_{R_\zeta} = \begin{cases} 
c & \text{if } z \geq \hat{z} \\
0 & \text{otherwise.} \end{cases} \tag{2.19}
\end{equation}

Under cordon pricing, therefore, those located at any distance \( z \leq \hat{z} \) face a zero toll and, given that an optimal toll would be positive, are underpriced. The level of charge, \( c \), and the location of the toll, \( \hat{z} \), determine which shares of those located at \( z \geq \hat{z} \) are overpriced or underpriced.

The other scheme we consider involves a \textit{flat kilometer tax}, with charging function:

\textsuperscript{33} In fact, as emphasized in the second-best literature, the cost of attaining this information is very high and public acceptance of any scheme that exhibits space variation as in (2.18) may be limited.
\[ \tau_{Rz} = \bar{\tau}z, \]  

where \( \bar{\tau} \) is the charge per unit of distance, and \( \bar{\tau} \geq 0 \).

### 2.2.4.2. Revenue return schemes

In absence of road tolls, i.e. in the base equilibrium, the lump sum transfer received by each household is:

\[ B_0 = \frac{R_0 + G_{L0}}{N}, \]  

where \( N \) denotes population, \( R_0 \) the excess land rents and \( G_{L0} \) the labor tax revenue.\(^{34} \) Note that the road tax revenue in the base equilibrium is zero, i.e. \( G_{R0} = 0 \). Given the presence of a toll, two cases are considered. In the first, public revenue is returned as a lump-sum transfer (LST). For each household, this is:

\[ B_1 = \frac{R_1 + G_{L1} + G_{R1}}{N}. \]  

The alternative scenario involves labor tax cuts (LTC), so \( \tau_L \) is reduced to keep total tax revenue at its base equilibrium level, i.e. \( G_{L1} + G_{R1} = G_{L0} \). The latter case is the situation that we will consider when optimizing policies under a public revenue constraint. It may seem odd to assume that the government aims to collect tax revenues to redistribute them in a lump-sum fashion. One might therefore interpret the model as one pertaining to a government that needs the revenues to (indirectly) finance some public good, while avoids introducing the additional market failures that a pure public good (explicitly modelled) would bring into the analysis. This formulation at the same time prevents tax revenues from leaking away from the general equilibrium setting. A consequence is that the marginal utility of the public good (if the above interpretation is followed) is equal to the marginal utility of lump-sum recycled tax revenues.

### 2.2.5. Equilibrium

In equilibrium, households must be indifferent on where to locate.\(^{35} \) Thus, the partial derivative of the indirect utility function in (2.10) with respect to \( z \) must be zero. This implies a first order differential equation which defines the rent gradient:

\(^{34} \) Although recycling the labor tax revenues in a (non-spatial) lump-sum manner would be a peculiar policy in reality, we choose to model the public budget this way in order to consider the inefficiencies from labor taxation without introducing further inefficiencies from public good provision. The latter would complicate the analysis while being peripheral to what we are interested in: the interactions between labor and transport taxes.

\(^{35} \) The model's specification is fully aligned with other monocentric city models. Through (2.10) we can express the equilibrium rent, \( r_z \), as a function of the equilibrium utility, \( V \), i.e. \( r(z, w - \tau_L - \tau_{Rz}, t_p, p, V) \). The differential equation in (2.23) can be made redundant, because \( r_z \) can be replaced \( r(z, w - \tau_L - \tau_{Rz}, t_p, p, V) \) in the Marshallian demand functions, in equations (2.7) to (2.9). This implies that the model is equivalent, for instance, to
\[
\frac{d}{dz} r_z = r_z^{1-x} \left\{ \left( \frac{\delta_s}{\delta_f} \right)^x \left( \frac{w - \tau_L - \tau_R z}{1 + t_f} \right)^{x-1} D_z \right. \\
\left. - T \delta_s^x D_z \left[ \left( \frac{p}{\delta_y} \right)^x + \left( \frac{r_z}{\delta_s} \right)^x + \left[ \frac{w - \tau_L - \tau_R z}{\delta_f(1 + t_f)} \right]^x \right] \right\},
\]

(2.23)

with terminal condition:

\[
r_z = r_A,
\]

(2.24)

where \( r_A \) is the exogenous agricultural rent and:

\[
D_z = - \frac{d}{dz} \left( \frac{w - \tau_L - \tau_R z}{1 + t_f} \right).
\]

(2.25)

The aggregate demand for the composite good in the equilibrium is:

\[
Y = \int_0^z n_\zeta \left[ M_z \left( \frac{p/\delta_y}{\delta_f} \right)^{-\sigma} \left( \frac{p/\delta_y}{\delta_f} \right)^x + \left( \frac{r_\zeta}{\delta_s} \right)^x + \left[ \frac{w - \tau_L - \tau_R \zeta}{\delta_f(1 + t_\zeta)} \right]^x \right] d\zeta.
\]

(2.26)

Using (2.4) and (2.9) to derive the labor supply, and aggregating over all households yields the aggregate labor supply:

\[
L = \int_0^z n_\zeta \left( \frac{1}{1 + t_\zeta} \right) \left[ T - M_z \left( \frac{w - \tau_L - \tau_R \zeta}{\delta_f(1 + t_\zeta)} \right)^{-\sigma} \left( \frac{p/\delta_y}{\delta_f} \right)^x + \left( \frac{r_\zeta}{\delta_s} \right)^x + \left[ \frac{w - \tau_L - \tau_R \zeta}{\delta_f(1 + t_\zeta)} \right]^x \right] d\zeta.
\]

(2.27)

The assumption of exogenous population, \( N \), implies that:

\[
N = \int_0^z n_\zeta d\zeta.
\]

(2.28)

Finally, the monetary value transferred to the absentee land owner must be equal to the value of the difference between production and consumption. This condition takes the form:

\[
p(Q - Y) = \bar{z} r_A.
\]

(2.29)

a restricted version of Bertaud and Brueckner (2005), in which floor-to-area ratio would be kept constant across space.
Note that this implies that urban land is, from the city’s perspective, purchased at a unit cost, \( r_A \), while excess land rents (see (2.16)) are redistributed as in (2.21). The urban equilibrium is characterized by the vector of endogenous variables \((Q,Y,G_L,G_R,R,B)\), the price vector \((p,w)\), the land rent function (solution of (2.23) and (2.24)), the travel time as a function of \( z \), and the equilibrium city limit, \( Z \).

2.3. Unimodal framework

2.3.1. Calibration and base equilibrium

The parameters of CES utility function \((\delta_Y, \delta_s, \delta_f, \rho)\), road technology \((t_0, t_1)\) and the policy instruments \((\tau_L, \tau_R)\) codetermine the level of congestion and a series of endogenous variables and elasticities in general equilibrium. These parameters are calibrated in order for the model to fit a series of stylized facts in the base equilibrium (i.e. the equilibrium without road pricing). The consumption share of income is equal to 0.52 (thus, the expenditure share on rents, mortgages and other types of expenditure related to the residence is equal to 0.48). The labor tax is set to 0.3, in line with Parry and Bento (2001). The (general equilibrium) elasticity of labor supply with respect to the labor tax ranges between -0.26, close to CBD, and -0.08, in the city fringe. The median household is characterized by a labor supply elasticity of -0.13. The parameter \( \rho \) gives rise to an elasticity of substitution of 0.72.

Households are working on average 50% of their disposable time, which translates to an average labor supply of 6 hours per day in the framework of a 12-hour time endowment. This framework excludes a minimum time requirement for sleep and the essential household activities (shopping, child care etc.) from the full 24-hour day. Since daily labor supply is fixed, the above result can be translated to an annual average labor supply of 182 days. This average encompasses weekends, holidays, sick leaves, part time jobs and temporary unemployment. The commute-to-work time ratio is 0.2, implying that the average one-way commuting time is 36 minutes. The car speed in the CBD relative to the city fringe is 0.2. The free flow travel time parameter, \( t_0 \),

---

36 We use Newton’s method for multivariate equations combined with line search to compute the equilibrium vector. We have employed heuristic uniqueness tests involving repeated computations over a set of initial price-quantity vectors.

37 The base equilibrium parameter vector values are: \( \delta_Y = 0.007904, \delta_s = 0.202399, \delta_f = 0.031456, \rho = -0.389, t_0 = 0.911348 \times 10^{-3}, t_1 = 0.782732 \times 10^{-5} \).

38 The model therewith applies to economies with a substantial labor tax. The labor tax functions both as a labor income tax rate and a tax per unit of labor supply. To see why, note that there are two basic normalizations in the chapter: i) the price of the numeraire (\( p \)) is fixed to one, and ii) through firm’s cost minimization the equilibrium wage (\( w \)) is equal to the price of the numeraire (thus, \( w = p = 1 \)). Then, the after-tax household income when the tax based on labor hours worked, i.e. \((w - \tau_L)D_W\), coincides with the after-tax household income when the tax is based on gross income, i.e. \((1 - \tau_L)W_DW\).

39 All elasticities of labor supply are computed with respect to the uniform labor tax, \( \tau_L \). We use finite differences (see equation 2.B.2 in Appendix 2.B) to compute the derivative of labor supply with respect to the labor tax. The reported elasticities can be re-expressed with respect to net wage, \((1 - \tau_L)w\), through:

\[
\frac{d \log \bar{D}_W}{\tau_L} = - \frac{d \log \bar{D}_W}{(1 - \tau_L)w} \left( \frac{\tau_L}{1 - \tau_L} \right).
\]
suggests that the travel time from the city fringe to CBD for an uncongested road is approximately 56 minutes (one-way).

2.3.2. Policy analysis

Table 2.1 shows the numerical results for the various policy scenarios. These results provide several insights. We begin with those insights that reposition Parry and Bento (2001) key findings in our spatial framework.

Table 2.1. The impact of first-best and second-best congestion pricing.

<table>
<thead>
<tr>
<th>Revenue recycling</th>
<th>Lump-sum transfers</th>
<th>Labor tax-cuts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road toll scheme</td>
<td>Base</td>
<td>Optimal</td>
</tr>
<tr>
<td>Policy instruments</td>
<td>First best</td>
<td>Monotonic</td>
</tr>
<tr>
<td>Labor tax, τ_L</td>
<td>Naive</td>
<td>Cordon</td>
</tr>
<tr>
<td>Charge per z</td>
<td>see</td>
<td>Flat km</td>
</tr>
<tr>
<td>Cordon location, ẑ</td>
<td>see (2.18)</td>
<td>37.56</td>
</tr>
<tr>
<td>Cordon charge, c</td>
<td>see text</td>
<td>0.086</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Endogenous variables</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>495.7</td>
<td>497.9</td>
</tr>
<tr>
<td>Y</td>
<td>365.8</td>
<td>372.8</td>
</tr>
<tr>
<td>z̄</td>
<td>129.8</td>
<td>125.1</td>
</tr>
<tr>
<td>Welfare measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ω</td>
<td>1.0 -0.458</td>
<td>0.129 0.125 0.109</td>
</tr>
<tr>
<td>ô</td>
<td>1.0 0.968 0.843 0.698</td>
<td></td>
</tr>
<tr>
<td>CV</td>
<td>1230 -563</td>
<td>159 154 134 111</td>
</tr>
</tbody>
</table>

Notes: L: aggregate labor supply, Y: aggregate output, z̄: city limit, CV: annual compensating variations in 2005 €, ω: relative efficiency (benchmark: first best), ô: relative efficiency (benchmark: varying kilometer tax). The value of ô can be computed only when revenue is recycled in the form of tax cuts.

Four results stand out for the case where toll revenues are returned in a lump-sum fashion. First, let us see in which parts of the city the tax interaction effect, i.e. the marginal welfare loss from the aggravation of the labor market distortion, dominates the revenue recycling effect, i.e. the marginal beneficial effect from the return of revenues. Evaluated at the margin of the Pigouvian equilibrium where road tolls are equal to marginal external congestion costs, such dominance implies that a marginal increase of that road tax is welfare decreasing; consequently, the optimal road tax lies below its Pigouvian level. In the context of lump-sum revenue returns, the revenue recycling effect is absent and a negative tax interaction effect automatically implies that the optimal tax lies below its Pigouvian level (and vice versa).40

40 Note that in this general equilibrium setting a lump-sum transfer would induce changes in households’ economic behavior, including adjustments in labor supply. The statement that the revenue recycling effect is zero reflects its definition, which ties it exclusively to a change in labor tax induced by a change in the externality tax, in a context of a revenue-neutral tax swap.
A second, natural follow-up question is whether the tax interaction effect dominates the (positive) Pigouvian effect, i.e. the welfare gain for correcting part of the congestion externality, at the margin of the base equilibrium. In the same context of lump-sum revenue return, such dominance implies that the optimal toll is negative (and vice versa). Thus, any positive toll would cause welfare losses.

Figure 2.1. Upper panels: optimal tax (dashed) and marginal external congestion cost (solid) under lump-sum revenue return across three different labor tax rates: 38% (left panel), 25% (middle panel) and 10% (right panel). Lower panels: optimal cordon toll (dashed) and flat kilometer tax (solid) for the corresponding levels of labor tax.

The upper panels of Figure 2.1 provide insights related to the above questions (computational details follow below and in Appendix 2.C). The panels display the optimal, non-negative tax schedules against the marginal external congestion cost (i.e. the Pigouvian levels of the road toll) by residential location and for three different labor tax rates: 38% (left panel), 25% (middle panel) and 10% (right panel). The second-best optimal road tax is strictly positive for residents at some distant locations, who generate higher marginal external congestion costs, but supply labor less elastically. Even when the labor tax is as high as 38%, this holds for roughly 10% of the distance from fringe to CBD. Therefore, even with lump-sum revenue return (i.e. in total absence of a revenue-recycling effect) road pricing for residents located at the suburbs of a monocentric city could yield welfare gains. A lower labor tax rate implies a smaller initial distortion of labor supply and thus a lower initial spatially undifferentiated tax on commuting. Because the tax interaction effect weakens and the Pigouvian effect strengthens, the critical

41 Throughout the chapter we impose a non-negativity constraint on road taxes.
location from which road pricing becomes efficient shifts inwards. For example, when the tax rate reduces to 0.25 the threshold moves from 10% to roughly 50% of the distance from fringe to CBD.

Third, we investigate what the pattern of the optimal tax implies for the (relative) efficiency of archetype pricing schemes. The lower left panel of Figure 2.1 shows that the optimal (non-negative) flat kilometer charge is zero. Thus, even a minimal positive flat kilometer tax with revenues returned lump sum ($FkmLST$) produces welfare losses. On the other hand, a distant, low priced cordon toll ($CoLST$) can capture some of the minimal welfare gains generated by the optimal tax scheme. As the level of labor tax becomes lower (middle and right panels), the optimal kilometer tax becomes positive and the relative efficiency of flat kilometer tax increases, restoring gradually its second-best superiority over the cordon toll, in accordance with what was found in Verhoef (2005) for the case without labor taxes.

In the above analysis we have highlighted how partial equilibrium models, which fail to take tax interactions into account, would erroneously support the introduction of a Pigouvian pricing scheme with a lump-sum redistribution of revenues. We complete the analysis of policies including lump-sum revenue return by examining the consequences of introducing such a scheme. That is, we consider the naive equilibrium, i.e. the equilibrium that prevails when the Pigouvian toll (see equation (2.18)) is charged with revenues returned lump-sum. Our results show that such policies can lead to very inefficient outcomes: assuming a disposable income of €52000, the monetized extra burden of this policy is €563 per year, i.e. slightly more than 1% of income (see the last line of Table 2.1, where monetized benefits of each policy are provided in terms of compensating variations).

The second set of insights regards the emergence of a weak double dividend (Goulder, 1995b). So far, the analysis has been restricted on the efficiency of pricing schemes across different levels of labor taxation, assuming that revenue is returned lump-sum. We now focus on the base scenario (labor tax 38%) and examine the performance of road pricing combined with labor tax cuts. In contrast to policies which return revenue lump-sum, the efficiency gains of the two imperfect pricing instruments are now far from negligible (€134 for the cordon toll and €111 for the flat kilometer tax). The equilibrium values of the policy instruments and endogenous variables associated with these policies are shown in the last two columns of Table 2.1.

The efficiency gains of the discussed policies are computed relative to the efficiency gains of two benchmark policies. The first one is a natural choice from the conceptual viewpoint but has only limited policy relevance, because it involves a complete elimination of the labor tax and the introduction of a Pigouvian toll (see the charging function in (2.18)) to commuters, with all revenue returned as a lump sum transfer. In this hypothetical first-best scenario, the road tax varies optimally over space. We denote the compensating variation under this scenario ($\tau_{RZ}$ as in

---

42 The compensating variation is the amount of money that the representative household has to forego (after the implementation of each policy) in order to return to the utility level of the base equilibrium.

43 Labor supply is below its social optimum in the base equilibrium (tax rate 38%), thus policies that return toll revenue in the form of labor tax cuts are at least as good as policies that return public revenue in the form of a lump sum transfer.
(2.18) and \(\tau_L = 0\) relative to the base scenario \((\tau_R = 0 \text{ and } \tau_L = 0.38)\) by \(CV_1\); the respective compensating variation under any second-best scenario is denoted by \(CV_2\). Then, the relative efficiency of an arbitrary second-best policy can be expressed as \(\omega = CV_2/CV_1\). The revenues of the cordon toll allow the labor tax to drop from 0.380 to 0.327 and yield a relative efficiency gain of 11\%. In line with this result, a flat kilometer tax reduces the labor tax to 0.340 with a relative efficiency gain of 9\%. Hence, both policies obtain similar gains in this context.

**Figure 2.2. Total tax per unit of supplied labor, rent, labor supply and compensating variations.**

One may argue that \(\omega\) is not a very relevant indicator of efficiency, as \(CV_1\) refers to a case in which tax revenue, \(G\), is allowed to be much lower than its base level.\textsuperscript{44} Therefore, it seems more appropriate to replace the denominator of \(\omega\) with a measure of the maximum welfare gain that can be realized with a *revenue-neutral tax swap*, i.e. when road tax revenue is used to

\textsuperscript{44} This would be inconsistent with \(G\) being taken to represent the money needed to finance some not explicitly modelled public good, as described above. Note that \(G = 188.07\) in the base equilibrium and only 48.81 in the first-best equilibrium.
finance labor tax cuts. This second benchmark policy combines a *varying kilometer tax*, \( \tau_R \), with a labor tax, \( \tau_L \), such that, in equilibrium, \( G \) will not differ from its base level. Denoting the compensating variation of this *marginal tax reform* with \( CV_{1G} \), we define a new measure of relative efficiency: \( \hat{\omega} = CV_2 / CV_{1G} \).

The optimal varying kilometer tax cannot be solved for analytically, but can be approximated numerically with an \( n \)-th degree polynomial. We do this by dividing the entire road interval \((0, \bar{z})\) from the CBD to the city fringe in \( n \) road segments. The optimal road charge can then be computed for each of the \( n + 1 \) points that separate these segments, given that Lagrangian interpolation (see Judd, 1998) is used to compute the road charge for every location \( z \). The associated labor tax with this policy turns out to be 0.28, *i.e.* 0.10 below the base labor tax. The profile of the optimal total local tax (*i.e.* sum of the labor and road tax) under a varying kilometer tax is shown by the solid curve in the upper left panel of Figure 2.

As explained in introduction, the optimal tax schedule is defined only up to the level of the total local tax, because the number of commuting trips is proportional to labor supply. Adding one unit to the total road tax paid at each location and subtracting it from the labor tax would leave the equilibrium and welfare unaltered. Still, it is possible to decompose the local tax into three components.

The first concerns the *Pigouvian* component of the local tax, equal to the marginal external cost caused by a resident at that location, as given by (2.18). The other two components jointly make sure that the revenue target is realized. For this, we first define the *spatially undifferentiated tax* (per unit of labor supply), \( \bar{\tau} \), which corresponds to the average labor tax that is needed to achieve the revenue target after accounting for the revenues that Pigouvian pricing \( (G^p_R) \) would bring:

\[
\bar{\tau} = \frac{G - G^p_R}{L}.
\]  

(3.30)

Letting \( n^*_\zeta \), \( D^*_W\zeta \) and \( mecc^*_\zeta \) denote residential density, labor supply and marginal external congestion costs at location \( \zeta \) at this equilibrium, \( G^p_R \) can be written as:

\[
G^p_R = \int_0^\bar{z} n^*_\zeta D^*_W\zeta mecc^*_\zeta \, d\zeta.
\]  

(3.31)

The final part of the tax is the *Ramsey-Mirrlees* (RM) component, *i.e.* a spatially differentiated tax (expressed as a deviation from \( \bar{\tau} \) and the marginal external cost of congestion), aiming to raise the predetermined revenue, \( G \), as efficiently as possible. Summarizing the above:

\[\text{We have used a division of the road interval} \ (0, \bar{z}) \ \text{in five segments of equal length. Letting the vector} \ \delta \ \text{denote the distance of those points from CBD as percentages of the total distance between the CBD and the city limit,} \ \bar{z}, \ \text{the vector} \ \bar{z} \ \delta' \ \text{expresses the exact location of each selected point in the sequence. In the present chapter we choose} \ \delta' = (0, 0.2, 0.4, 0.6, 0.8, 1.0). \ \text{See Appendix 2.C for optimization details.}\]
\[ \tau_z^* = \tau_L + \tau_{Rz} \Rightarrow \tau_z^* = \tau_L + mecc_z + \frac{(\tau_{Rz} - mecc_z)}{\text{road overcharge}} \]
\[ \Rightarrow \tau_z^* = \bar{\tau} + mecc_z + \left\{ (\tau_{Rz} + \tau_L) - (\bar{\tau} + mecc_z) \right\}. \]

It is important to note that we have defined the Ramsey-Mirrlees (RM) component in such a way that the average Ramsey-Mirrlees (RM) component is equal to zero (it can be shown by integrating both sides of (2.32), and using (2.30)). Arguably, taking \( \bar{\tau} \) as the reference value for the undifferentiated labor tax in the optimum is arbitrary. There is however a strong argument to use \( \bar{\tau} \), since it is the average labor tax that makes the budget constraint to hold after accounting for the revenues from Pigouvian road pricing.\(^{46}\)

The Ramsey-Mirrlees (RM) component reflects the constraints under which the optimal values of policy instruments have been computed: Ramsey refers directly to induced distortions on the labor market, and implies that the tax should (generally) be higher where labor supply with respect to the labor tax is less elastic. Since the elasticity of labor supply decreases (in absolute value) with distance from CBD, this would suggest an increasing (in \( z \)) tax.\(^{47}\) Mirrlees refers to the observation that the marginal utility of income rises with distance from CBD, since pecuniary commuting costs increase. Because the initial labor tax does not reflect differences in marginal utility of income, the optimal road tax will be adjusted to partially correct for this inefficient treatment (Wildasin, 1986). This effect alone would make the tax fall in \( z \).

The two forces underlying the Ramsey-Mirrlees component thus work in opposite direction. The combined effect can therefore portray a complex pattern, which for our base calibration turns out to be non-monotonic over space. Translating the location-based tax into a road link-based tax would then mean that the charge in the most remote links would be negative

\(^{46}\) Note that we will examine whether our results hold for other values of the undifferentiated labor tax. In order for the Ramsey-Mirrlees component to provide a satisfactory indication on where the second dividend emerges, it has to be robust across all combinations of policy instruments that give rise to the same equilibrium. For instance, adding a constant \( \theta \) to the road tax \( \tau_R \) at each location leaves the equilibrium intact, when the level of tax revenue is predetermined: from the budget constraint in (6) and the governmental budget constraint in (17) it can be inferred that \( \tau_L \) will be reduced by the same amount, \( \theta \). The RM component in (2.32) is unique to every set of policy instruments \( (\tau_R, \tau_L) \) that give rise to a unique equilibrium. In contrast, the road overcharge function, \( \tau_{Rz} - mecc_z \), is not.

\(^{47}\) It is currently unknown to what extent labor supply elasticity with respect to the wage rate varies over space. This question is hard to answer because of the econometric issues involved related to the endogeneity of wages with respect to hours worked. Empirical studies which analyse the effect of fixed costs on labor supply have focused particularly on childcare (Blau and Robins, 1988). The overall elasticity of labour force participation of mothers with regard to child-care prices is between \(-0.05\) and \(-0.35\). Importantly, a recent stream of empirical literature analyses the effects of commuting on number of hours worked (Gershenson, 2013; Fu and Viard, 2015). This stream establishes the main underlying assumption of our approach, \( i.e. \) that labour supply is reduced when workers live further away from their residence location. These latter studies improve on the approach introduced by Gutiérrez-i-Puigarnau and van Ommeren (2010) by examining labour supply of workers who are confronted with an exogenous change in commuting distance due to a relocation of their employer. Similarly, Gutiérrez-i-Puigarnau and van Ommeren (2010) show that the number of days worked is a decreasing function of commuting distance. However, the latter also suggests that commuting distance increases labor supply in the intensive margin (hours of work per day) such that overall supply is hardly sensitive for workers who do not change employer.
(in line with Van Dender, 2004). This non-monotonicity of the Ramsey-Mirrlees component is reflected in a modest reduction of the relative efficiencies of archetype schemes in the case of a revenue-neutral tax swap (84% for cordon toll and 70% for the flat kilometer tax). To explain why these schemes still perform well relative to a non-monotonic pricing scheme we need a measure over the efficiency loss caused entirely by the non-monotonicity of the optimal pricing scheme. For this calculation, we have used an additional policy which introduces a varying kilometer tax restricted to be monotonic over space. The results of Table 2.1 show that the non-monotonicity of the optimal scheme is almost negligible from a welfare perspective: 97% of the welfare gains generated by the optimal non-monotonic scheme can be captured with a monotonic space varying tax. In the sensitivity analyses following in Section 2.3.3 we argue that the value of \( \hat{\omega} \) of the monotonic scheme is a satisfactory indicator of the respective \( \hat{\omega} \) of the archetype schemes.

Another insight regards the welfare effects of a space-varying tax in absence of congestion. A natural way to acquire a measure over the magnitude of the Ramsey-Mirrlees component is to compare its space variation to this of the total road tax.\(^{48}\) This expresses how much of the total tax is dedicated to purposes that are not directly related to congestion. From the patterns displayed in the upper left panel of Figure 2.3, we can derive that this is approximately 35% of the total road tax, implying that a space-varying road tax might be welfare improving even without congestion.\(^{49}\) The second and third panel in the left column show the spatial variation of the labor supply elasticity and the marginal utility of income, where the flattening of the former towards the city’s fringe and the continued increase of the latter make the observed pattern of the second-best tax plausible.

The road tax’s feature of allowing spatial differentiation of the road tax remains relevant also in absence of congestion, the case considered in the right panels in Figure 2.3. The upper right panel of Figure 2.3 suggests that, in absence of congestion, the Ramsey-Mirrlees component retains most of its spatial variation. The middle and lower panels of the same figure show the corresponding elasticity of labor supply (with respect to the labor tax) and the marginal utility of income respectively.

Overall, the sign of Ramsey-Mirrlees component at each location is related to the concept of a strong double dividend (Goulder, 1995b). Recall that a positive sign means that the local tax is higher than what would be justified by spatial variation on marginal external congestion costs alone. However, as discussed above, the sign of this component depends on the level of the undifferentiated labor tax \( \bar{\tau} \).\(^{50}\) One may argue that given another – also arbitrary – value of the labor tax, results may substantially differ. Note however that this is not the case: even when we choose a much lower level of undifferentiated labor tax, for example the minimum of local taxes

\(^{48}\) Note that, if marginal utility of income were constant, i.e. if utility function were quasi-linear in consumption, Mirrlees component would disappear and only a Ramsey component would remain (see Appendix 2.A for an example).

\(^{49}\) We use the following measure: \( \int_0^\xi (RM_\zeta - RM_0) d\zeta / \int_0^\xi (RM_\xi + mecc_\zeta - RM_0) d\zeta \).

\(^{50}\) Given our definition, the average RM component is zero, so there are always locations with a positive RM component and locations with a negative one.
as shown in Figure 2.2 (around 0.28), the implied road tax \((0.42 - 0.28 = 0.14)\) is still substantially (about 40%) above the maximum of marginal external costs shown in Figure 2.3 (0.10). So, also then, a strong double dividend would be found.

**Figure 2.3. Marginal external cost of congestion and the Ramsey-Mirrlees (RM) component (upper panels), labor supply elasticity (middle panels) and the marginal utility of income (lower panels) across space, with (left panels) or without (right panels) congestion.**

2.3.3. Sensitivity analysis

To verify the sensitivity of the above numerical results with respect to the values of some key parameters, we perform one-way sensitivity analyses with respect to the base labor tax rate, \(\tau_L\), and the elasticity of substitution, \(\sigma\). Figure 2.4 summarizes the findings. We set the range of the tax rate between 0.20 and 0.50 to account for the fact that the majority of workers in Western Europe and North America face a labor tax within this range. We set \(\sigma\) to vary between 0.64 and 0.82.

Figure 2.5 plots the Ramsey-Mirrlees component from the optimal tax at the extreme values of the parameters considered in the sensitivity analysis. It is shown that a lower income
tax, as well as a lower elasticity of substitution, gives rise to a flatter Ramsey-Mirrlees component. For the lower income tax, the interpretation is that the lower revenue requirement (which is adjusted accordingly), and the lower variation in labor supply elasticity and in marginal utility of income justify a smaller spatial differentiation of taxes on top of the differentiation motivated by the Pigouvian component. For the lower elasticity of substitution, the flatter pattern of the Ramsey-Mirrlees component occurs especially near the city fringe, because the city is smaller with a lower elasticity of substitution (thus, there is less room for the lowering of the labor tax in response to differences in marginal utility of income).

**Figure 2.4. Sensitivity analysis of the relative efficiency, \( \hat{\omega} \), with respect to \( \sigma \) and \( \tau_L \).**

This flattening favors the relative performance of a flat kilometer tax, which approaches that of the cordon toll. In fact, when the labor tax is completely removed it may even exceed it, as in Verhoef (2005), who abstracted from distortionary taxation.

**Figure 2.5. The Ramsey-Mirrlees component at the bound values of sensitivity analysis.**
2.4. Bimodal framework

In this section we expand the model to account for the existence of a mass transit provider which, in the base equilibrium, operates with fixed costs and applies average cost pricing. We subsequently juxtapose policies which return the road tax revenue in the form of labor tax cuts against policies which use this revenue to subsidize the public transport provider. It is shown that the optimal policy depends heavily on the size of fixed costs in public transport.

2.4.1. Expanded model

Now let the same household choose between commuting with car or public transport. Then, the demand functions in (2.7), (2.8), (2.9) and the indirect utility in (2.10) are not only conditional to the commuting distance, $z$, but are also mode-specific, representing the car mode. If the household chooses public transport, its time constraint will be:

$$d_{wz}^p = \frac{T - T_{Fz}^p}{1 + t_{z}^p},$$

(2.33)

where the superscript $P$ denotes the choice of the public mode. Similarly, the conditional budget constraint will be:

$$B + \frac{(w - \tau_L - p_{pz})}{1 + t_{z}^p} T = \frac{(w - \tau_L - p_{pz})}{1 + t_{z}^p} l_z + p y_z^p + r_z s_z^p,$$

(2.34)

where $p_{pz}$ denotes the ticket fare from $z$ to the CBD and $t_{z}^p$ the respective commuting time. The Marshallian demand functions for the composite good, space and leisure time, and the indirect utility function are respectively:

$$y_{z}^{P^*} = M_z^P \left( \frac{(p/\delta_y)^{\sigma}}{(p/\delta_y)^{\chi} + (r_z/\delta_s)^{\chi} + \frac{w - \tau_L - p_{pz}}{\delta_f(1 + t_{z}^p)}} \right)^{\frac{1}{\chi}},$$

(2.35)

$$s_{z}^{P^*} = M_z^P \left( \frac{(r_z/\delta_s)^{\sigma}}{(p/\delta_y)^{\chi} + (r_z/\delta_s)^{\chi} + \frac{w - \tau_L - p_{pz}}{\delta_f(1 + t_{z}^p)}} \right)^{\frac{1}{\chi}},$$

(2.36)

$$T_{Fz}^{P^*} = M_z^P \left( \frac{\delta_f^p (1 + t_{z}^p)}{\delta_f (1 + t_{z}^p)} \right)^{\sigma} \left( \frac{w - \tau_L - p_{pz}}{\delta_f (1 + t_{z}^p)} \right)^{\frac{1}{\chi}},$$

(2.37)
\[ V^p_z = \left[ B + \left( \frac{w - \tau - p_{pz}}{1 + t^p_z} \right) T \right] \left[ \left( \frac{p}{\delta y} \right)^x + (r_z/\delta s)^x \right] \\
+ \left[ \frac{w - \tau - p_{pz}}{\delta y (1 + t^p_z)} \right] ^{-1/x} \]  
(2.38)

The indirect utility functions in (2.10) and (2.38) are the systematic utilities for the choice of car and public transport respectively. To prevent cases in which residents choose exclusively car in some areas and public transport in others, a stochastic term, \( \varepsilon_{nm} \), which is i.i.d. extreme value type I across households (denoted by the subscript \( n \)) and transport modes (denoted by the subscript \( m \)) is added to these functions. The stochastic term captures daily variations in the household preferences regarding the commuting mode. This assumption implies that the same household will some days commute by car and some others using the public transport mode, with a logit choice probability. Letting \( y^c_z, s^c_z, T^c_z \) and \( V^c_z \) denote the respective demand functions and the systematic utility when choosing car, the logit choice probability for car is then:

\[ P_C = \frac{\exp(V^c_z)}{\exp(V^c_z} + \exp(V^p_z). \]  
(2.39)

Furthermore, it can be shown (Small and Rosen, 1981) that the expected maximum utility in this choice situation (car versus public transport mode given the commuting distance \( z \)) is:

\[ E[\max\{V^c_z + \varepsilon^c, V^p_z + \varepsilon^p\}] = \ln[\exp(V^c_z) + \exp(V^p_z)]. \]  
(2.40)

In equilibrium, the derivative of (2.40) with respect to the land rent, \( r_z \), should be equal to zero. The expected equilibrium consumption and labor supply are calculated as probability-weighted sums of the conditional expressions given above.

We assume that the public transit operator that produces passenger kilometers, \( q \), is subject to a cost function:

\[ C(q) = F + \varphi_0 q + \varphi_1 q^2. \]  
(2.41)

In equilibrium, the degree of scale economies, defined as the ratio of average to marginal cost, is:

\[ s(q) = \frac{AC(q)}{MC(q)} = \frac{(F/q) + \varphi_0 + \varphi_1 q}{\varphi_0 + 2 \varphi_1 q}. \]  
(2.42)

The operator incurs the fixed cost \( F \), which might include construction, maintenance and organization, and which might lie in a wide range, depending on the mode and the idiosyncratic characteristics of the city and its transport system. We ignore congestion effects in public transport, as for example in Basso and Jara-Díaz (2012). Therefore, the commuting time with the public mode is simply proportional to distance:

\[ t^p_z = t^p z, \]  
(2.43)
where the technology parameter $t_2$ denotes the travel time per unit of distance.

The operator is assumed to run a balanced budget, as for example in Japan (Priemus and Konings, 2001), so the public transport kilometer price is equal to the average cost, i.e. $p_p(q) = (F/q) + \varphi_0 + \varphi_1 q$. This introduces an inefficiency in the transport market, since the resulting price deviates from the marginal social cost of a passenger kilometer. However, the revenue of the road toll can be used to subsidize the operator, rather than to reduce the income tax. In this case the resulting price is: $p_p(q) = ((F - G_R)/q) + \varphi_0 + \varphi_1 q$. With this policy intervention, a part of the inefficiency in the transport market is restored, but the labor market benefit will be smaller, compared to the case of labor tax cuts. The only changes that occur stem from labor supply change due to cheaper public transport. The purpose of the next section is to compare the welfare implications of these two revenue recycling programs.

2.4.2. Results

In order to calibrate the parameters of the extended model we have assumed that, in the base equilibrium of the bimodal framework, 1% of the city's output is used to pay for the fixed costs of public transport. This number is only speculative; a sensitivity analysis in a substantial range around this level follows below. The public transit technology parameter, $t_2$, is set to 0.003. This implies that the speed of the public transport mode is 30% of the car speed on the freeway. The parameters of the public transit cost function are such that, in the base equilibrium, $s(q)$ in (2.42) is equal to 1.4 and $s(2q)$ is equal to 1. We therefore initiate our sensitivity analysis from a base equilibrium characterized by significant economies of scale and a level of output (passenger kilometers) which lies halfway to its break-even level.51 Otherwise, the equilibrium stylized facts (consumption share of expenditure, CBD-to-fringe speed ratio, labor supply elasticity, commute-to-work time ratio) resemble those reported in the unimodal case.52

Figure 2.6 summarizes the results of sensitivity analysis with respect to the level of fixed costs, $F$. The figure shows the welfare gain from using the revenues for labor tax cuts relative to using it for public transport subsidies, when the pricing scheme is the cordon toll (dashed line) or the flat kilometer tax (solid line). Each point on the horizontal axis represents an initial equilibrium, in which a different fraction of the city's total output, $Y$, is directed to cover the transport system's fixed costs, $F$, while the rest of the parameters in the model are fixed. Thus, the extent of scale economies, $s(q)$, varies significantly across the range of fixed costs (0-2% of total output); from approximately 0.87 (i.e decreasing economies of scale) when $F$ is close to zero, to 1.8 when $F$ is 2% of the output value (strongly increasing economies).53 For the equilibria within this range, we report the relative compensating variations of the two archetype schemes, i.e. the compensating variations of a cordon toll combined with labor tax cuts over the compensating variations generated by a cordon toll combined with public transport subsidies

51 Bimodal base equilibrium: $F = 5.097, \varphi_0 = 0.583 \times 10^{-3}, \varphi_1 = 0.8827 \times 10^{-8}$.
52 Consumption share: 0.50, commute-to-work ratio: 0.134, labor supply elasticity: -0.154, CBD-to-fringe speed ratio: 0.2.
53 This range of scale economies is roughly consistent with the empirical findings by Viton (1981).
(\(CoLTC\) over \(CoPTS\) in the dashed line) and the respective fraction for the case of a flat kilometer tax (\(FktLTC\) over \(FktPTS\) in the solid line).

Figure 2.6. Compensating variations generated by labor tax cuts expressed relative to those generated by public transport subsidies.

As the fixed costs of public transport decrease, the gains from policies that return the revenue from the toll road in the form of a subsidy to the operator (\(CoPTS\), \(FktPTS\)) decrease when compared to the gains generated by policies that reduce the labor tax (\(CoLTC\), \(FktLTC\)). The reason is that with lower fixed costs, the average-cost pricing provider operates much closer to the break-even point and average cost pricing is much closer to the (optimal) marginal cost pricing. Thus, the marginal social benefit of spending one euro in the labor market (through a decrease in the labor tax) is much higher than the respective marginal benefit of using one euro to subsidize the public transport operator. The opposite holds for high values of \(F\). Therefore, this chapter advocates for revenue recycling strategies that complement those by Parry and Bento (2001), who exclude public transport fixed costs from their model and derive a policy recommendation which universally rejects public transport subsidies in favor of labor tax cuts. Contrary to this, our results indicate a threshold value of fixed costs, above which public transport subsidies become the most efficient form of revenue recycling in a monocentric city. This threshold value lies between 0.6 and 1.0\% of the total output in our numerical example, and probably even lower in a model that accounts for the excess environmental externalities generated by the use of private modes.

2.5. Concluding remarks

This chapter studied the design and impacts of road pricing schemes and associated revenue recycling programs, in the framework of a monocentric city with a distorted labor market. We showed that, when road tax revenue is used to reduce the labor tax, only a part of the spatial variation of the optimal road tax corresponds to congestion externalities. For our numerical example, this part is approximately 65\%. The remaining 35\% of the variation concerns what we called the \textit{Ramsey-Mirrlees} component, which aims to minimize the distortions in the labor
market from revenue-raising labor taxation. Specifically, the Ramsey part of the component aims to minimize the social cost of raising a given tax revenue by setting taxes higher where labor supply is less elastic. The Mirrlees part aims to set taxes higher where the marginal utility of monetary income is lower.

There are three important findings associated with the Ramsey-Mirrlees component. First, it retains its spatial profile even when road externalities are absent. This motivates a space-varying tax even in contexts without traffic congestion. Second, it may display a non-monotonic profile over space, resulting in a road tax which, for some parts of the city, is decreasing with distance from CBD. Third, the sign of the Ramsey-Mirrlees component is related to the local emergence of the strong double dividend, despite the overlap of the labor and road tax bases. This overlap renders the two taxes indistinguishable, obstructing the use of conventional definitions of tax interaction and revenue recycling effects. Still, for two plausible definitions of what part of the local taxes constitutes the labor tax, locations with a strong double dividend (i.e., a road tax exceeding the marginal external congestion cost) were found.

Independent of the emergence of a strong second dividend, a weak double dividend (Goulder, 1995b; Bovenberg, 1999) is generated: all else equal, it is always more efficient to return the revenue in the form of a labor tax cut than via a lump-sum transfer. A varying kilometer charge which uses the road toll revenues to reduce the preexisting labor tax generates annual welfare gains which account for € 160 per household. In absence of the revenue recycling effect, i.e. when the toll revenue is recycled lump-sum, tax interactions might be strong enough to generate very large annual welfare losses. The usual textbook prescription of a Pigouvian toll on top of the labor tax, in combination with a lump-sum recycling of the revenues, produces losses of € 560 per household.

This implies that the optimal tax under lump-sum recycling lies much lower than the marginal external congestion cost. Our spatial framework allows us to identify where it is strictly positive. In the base calibration of the model, this only holds for residents living at the most remote locations, where the Pigouvian term is relatively high. Subsequently, the relative efficiency of a (non-negative) flat kilometer tax, which implies positive taxes for residents at all locations, is essentially zero. At the same time, the cordon charge lends itself well for keeping tolls zero for residents living closer to the CBD and being positive for more remote residents. Consequently, with lump-sum recycling, a part of the welfare gains from the second-best space-varying tax can be captured by a distant cordon toll. This result deviates substantially from Verhoef (2005), who highlights the ability of these two schemes to capture a very large portion of the welfare gains generated by the optimal tax scheme in a city without distortionary labor taxation. It also suggests, somewhat surprisingly, that if only lump-sum recycling is possible, cordon charging may be preferred on efficiency grounds over flat kilometer charges in urban applications. Consistent with the findings in Verhoef (2005), the relative supremacy of the flat kilometer tax over the cordon toll is gradually restored as the initial labor tax is lowered.

The numerical model furthermore suggested moderate efficiency gains of the two archetype second-best pricing schemes (i.e. a flat kilometer tax and a cordon charge) when
revenues are instead used to lower the labor tax. The accompanying sensitivity analyses show that the efficiency gain of the cordon toll remains modest in a wide range of the assumed parameters and the base labor tax, but this is less so the case for the flat kilometer tax.

In order to juxtapose labor tax cuts against public transport subsidies, we have furthermore considered a bimodal setting where commuters may choose between car and public transport. Assuming fixed costs and average cost pricing in the provision of public transport, we find that public transport subsidies may be superior to labor tax cuts, if fixed costs in the supply of public transport are substantial. Therefore, this chapter allows for the possibility of welfare-improving revenue recycling strategies that complement those by Parry and Bento (2001), who exclude fixed costs from public transport and, subsequently, universally reject public transport subsidies in favor of labor tax cuts.

This chapter has modified the monocentric model to account for road externalities, labor market distortions and non-marginal cost pricing in public transport. However, a series of other relevant inefficiencies in urban markets, notably those stemming from housing and land market regulations that affect the urban density have still been ignored and will receive attention in the next chapter. Furthermore, we have exclusively focused on a monocentric city. Tax reforms that combine labor taxation with road tolls in a polycentric setting are examined in Chapter 5.

Appendix 2.A: Constant marginal utility of income

The marginal utility of income is constant if the CES utility function is replaced by a quasi-linear utility in consumption. For instance, consider:

\[
u = \ln \left\{ \left[ (\delta_s s z)^\rho + (\delta_f T_{FZ})^{\rho} \right]^{\frac{1}{\rho}} \right\} + y_z, \tag{2.1A.1}\]

which is subject to the same full income constraint, given in (2.6). Setting

\[ D = \left[ (\delta_s s z)^\rho + (\delta_f T_{FZ})^{\rho} \right]^{\frac{1}{\rho}}, \tag{2.1A.2}\]

as a subutility function and minimizing the expenditure \( \frac{(w-\tau_L-\tau_{RZ})}{1+t_z} T_{FZ} + \tau_z s_z \) for a given level of \( D \) yields the CES expenditure function:

\[
c(D) = D \left[ \frac{r_z}{\delta_s} \right]^{\frac{\rho-1}{\rho}} + \left( \frac{w-\tau_L-\tau_{RZ}}{\delta_f(1+t_z)} \right)^{\frac{\rho-1}{\rho}} = D p_D, \tag{2.1A.3}\]
where $p_D$ is the expenditure per unit of (sub)utility. The problem transforms to the maximization of $u = \ln(D) + y_z$ subject to the transformed constraint:

$$y_z = B + \frac{(w - \tau_L - \tau_Rz)}{1 + t_z} T - Dp_D. \quad (2.A.4)$$

The corresponding indirect utility is:

$$V = \ln\left\{\frac{1}{p_D}\right\} + B + \frac{(w - \tau_L - \tau_Rz)}{1 + t_z} T - 1. \quad (2.A.5)$$

The associated marginal utility of income, $\partial V / \partial B$, is constant.

### Appendix 2.B: Marginal tax reform and double dividend

As stated in the introduction, the problem studied in this chapter naturally bears close resemblance to problems studied in the double dividend literature, but the perfect additivity of the labor tax and the road tax (with both varying in proportion with an individual’s labor supply due to the assumed fixed relation between labor supply and commuting trips and hence affecting the same behavioral margin in a perfectly additive way) renders a straightforward application of the approach followed in the conventional double dividend literature problematic. This appendix sets out why this is the case, and presents what we believe is a meaningful way to assess the effects of the pre-existing labor tax on the efficiency of marginal tax reforms in terms of the concepts used in the double-dividend literature.

Let us first summarize the conventional approach, as presented also for example in Goulder et al. (1997), which we will follow while adjusting notation in order to stay as close as possible to ours. The standard analysis ignores commuting time and assumes that a consumer consumes leisure time $T_F$ out of a total time endowment ($T$), trading it off against working hours $T_L$, and facing a net wage, $w - \tau_L$. The resulting income can then be spent on consumption goods, one of which (R) gives rise to marginal external costs (mec), and which can be taxed at a rate $\tau_R$. Goulder et al. (1997) then present the following decomposition of the welfare effect of a marginal change in $\tau_R$ (expressed in monetary terms due to the premultiplication by the inverse of the marginal utility of income, $\lambda$):

$$\frac{1}{\lambda} \frac{dU}{d\tau_R} = (\tau_R - mec) \frac{dR}{d\tau_R} + MEB \left( R + \tau_R \frac{dR}{d\tau_R} \right) - (1 + MEB) \tau_L \frac{\partial T_F}{\partial \tau_R}, \quad (2.B.1)$$

where:
\[ MEB = \frac{\tau_L \frac{\partial T_F}{\partial \tau_L}}{(T - T_F) - \tau_L \frac{\partial T_F}{\partial \tau_L}} \]  

(2.B.2)

gives the marginal excess burden: the social shadow cost of raising an additional unit of tax revenues through a tax in the labor market. The numerator in (2.B.2) can be understood by imagining the (Harberger) welfare triangle representing the welfare loss in the consumption of leisure due to the initial tax. The tax \( \tau_L \) gives the height of this triangle, and a marginal change in that tax makes the triangle increase by that height multiplied by the induced increase in leisure consumption. The denominator is the marginal revenue of an increase in the labor tax (note that without commuting time, \( T - T_F \) equals \( T_L \)).

Equation (2.B.1) then decomposes the full marginal welfare effect of a revenue-neutral tax swap in three terms. The first is the Pigouvian effect: again the change in the size of the Harberger triangle, but this time in the market for the polluting consumption good (it multiplies the marginal wedge with the marginal effect of the tax change on the consumed quantity). For a tax below mec, both sub-terms are negative, and the Pigouvian effect is therefore positive. The second is the revenue recycling effect. It multiplies the marginal excess burden with the marginal revenue of the externality tax, and thus represents the social benefit that the additional revenues bring through allowing a reduction in the labor tax. The third term is the tax interaction effect. Without the marginal excess burden included, this term would give the change in the Harberger triangle of the labor market due to the induced change in leisure consumption following a change in the externality tax. The term \( MEB \) is added to properly weight the change in labor tax revenues that follows from this change in leisure consumption.

Now let us consider the case where the consumption of good \( R \) is perfectly complementary with the supply of a primary factor, in this case \( T_L \). Now, it is the sum of \( \tau_L \) and \( \tau_R \) that determines the consumer’s simultaneous choice of \( R \) and \( T_L \). This is the situation that applies at every location in our spatial model, where the consumption of road trips and the supply of labor occur in perfect proportion, but we can treat the essence of the problem also for the non-spatial model introduced above. As in the main text, define the sum of taxes as \( \tau^* = \tau_L + \tau_R \). Any given value of \( \tau^* \), although corresponding to the same equilibrium, could then be decomposed in an infinite number of ways into different values of \( \tau_L \) and \( \tau_R \). This would in turn mean that different values of the marginal excess burden, the Pigouvian effect, the revenue recycling effect and the tax interaction effect could be obtained for the same equilibrium. Among others, that reflects the simple observation that if we declare, for a given equilibrium, the labor tax to be zero and hence the equilibrium tax to be entirely a tax on consumption of the polluting good \( R \), there will be a computed zero marginal excess burden from that labor tax according to equation (2.B.2) and hence a zero revenue recycling effect, as well as a zero tax interaction effect.

A further problematic aspect of applying the expressions (2.B.1) and (2.B.2) in this context is that, with the taxes affecting the same single margin of behavior in the same (perfectly
additive) way, also for the computation of Harberger triangles, one should look at the sum of
taxes rather than at an arbitrarily chosen level of the one or the other tax. This makes the
conventional framework ambiguous in the context of a non-spatial model with perfectly
complementary taxes. And this ambiguity remains when moving to the spatial setting, where also
any continuous schedule of local taxes can be decomposed in an endless number of ways into a
common labor tax and space-varying local road taxes.

Still, there is insight to be gained from expressing the welfare effects in our model in
terms of concepts used in the double dividend literature. The closest someone may get to
presenting the results in terms of that terminology, taking into account the above considerations,
is to abandon the idea of two wedges and instead to consider for every location a single “wedge”.
That wedge applies simultaneously in the labor market and in the transport market, and is given
by the difference in the local tax and the mec at that location. Furthermore, in specifying the
equivalents of the three effects mentioned above, one would have to take into account spatial
interactions, between residents located at different distances from the CBD. Indeed, instead of
having a separate labor tax and externality tax we merge these into one local tax, but at the same
time we now have a continuum of such local taxes, that will interact across space. Hence, even
though at every location the labor tax \( \tau_L \) and externality tax \( \tau_{RZ} \) have been combined to a single
local tax \( \tau_z \), the spatial continuity implies that we have an infinite number of tax levels to
optimize.

Accounting for all these considerations, one may adjust the definitions in equations
\((2.2.1)\) and \((2.2.2)\) and define the local Pigouvian effect \( W_{Pz} \), the local revenue recycling effect,
\( W_{Rz} \), and the local tax interaction effect, \( W_{Iz} \), by simultaneously changing a common space
invariant tax component (that we will keep referring to as \( \tau_L \)) for all locations. Using \( MEB \) as the
marginal excess burden from a marginal tax increase of a space-invariant component \( \tau_L \), we
arrive at the following expressions for the effects in the present context:

\[
MEB = \int_0^z (\tau_L + \tau_{RZ} - mec_z) \frac{\partial n_z T_{EZ}}{\partial \tau_L} dz \frac{\partial}{\partial \tau_L} \left\{ \int_0^z n_z T_{LZ} (\tau_L + \tau_{RZ}) dz \right\},
\]

\[ (2.2.3) \]

\[
W_{Px} = \int_0^z (\tau_L + \tau_{RZ} - mec_z) \frac{\partial n_z T_{LZ}}{\partial \tau_{Rx}} dz,
\]

\[ (2.2.4) \]

\[
W_{Rx} = MEB \frac{d}{d \tau_{Rx}} \left\{ \int_0^z (\tau_L + \tau_{RZ}) n_z T_{LZ} dz \right\},
\]

\[ (2.2.5) \]
\[ W_{lx} = \int_0^z (\tau_L + \tau_{RZ} - mec_z) \frac{\partial n_z T_{FZ}}{\partial \tau_{Rx}} \, dz - \frac{d}{d\tau_{Rx}} \left\{ \int_0^z \text{MEB}(\tau_L + \tau_{RZ}) n_z T_{LZ} \, dz \right\}. \] (2.B.6)

The fact that we now have a single tax at each location translates into the perfect correspondence between the revenue recycling effect and the second term in the tax interaction effect. Similarly, there is an equivalence between the Pigouvian effect and the first term of the tax interaction effect (the equivalence becomes perfect if the change in commuting time from a marginal tax change approaches zero so that the changes in \( T_{LZ} \) and \( T_{FZ} \) are equal in absolute size). These equivalences reflect that for a single tax, the concept of tax interaction between two perfectly additive components of that tax is quite abstract. It is only for comparability with the double dividend literature that we compute these terms numerically. Especially for the Pigouvian motivation of the road tax, it is interesting to compare the spatial pattern of the Pigouvian effect as defined above to that of \( mec \) shown in Figure 2.3 in the main text – the Pigouvian component of the optimal tax.

The measures from equations (2.B.3)-(2.B.6) were approximated numerically, and the results are shown in Figure 2.B.1 below. The lowest curve, giving the Pigouvian effect, has a maximum near \( z = 110 \), while the Pigouvian component of the tax depicted in Figure 2.3 in the main text naturally increases monotonically with distance. The difference in qualitative patterns stems from the fact that the latter is defined as one of multiple components making up the optimal tax level, whereas the former is defined as one of multiple effects that a marginal change in the tax level would bring. The Pigouvian component of the tax increases monotonously with distance as more distant travelers cause congestion over a longer trip, fully covering the trips made by travelers with a sorter trip. In contrast, the Pigouvian effect of a marginal change of the tax peaks in an interior point of the residential area. This is the result of the balance of various factors determining the size of this effect, each having a different pattern over space.

A first is the fact that at a greater distance, an increase of the tax will bring congestion benefits to other travelers over longer distances. A second is that the sensitivity per individual, as well as the number of individuals, decrease with distance. The first factor would make the Pigouvian effect rise with distance, the second would make it decline, and the combined result is the peaked pattern shown in Figure 2.B.1. Next, the revenue recycling effect (as defined above) falls monotonously over space. Apparently, the volume effect of having a higher density closer to the CBD dominates the sensitivity effect of greater volume adjustments in response to tax changes closer to the CBD, throughout the city. Finally, because of the additivity of the three components, the pattern of the tax interaction effect computed in this way is simply the combined result of the two patterns just described.
Figure 2.B.1. Numerical approximations of the tax interaction, revenue recycling, and Pigouvian effect.

Appendix 2.C: Technical details (numerical optimization).

For optimizing, we have used a combination of direction set methods (Newton method and BFGS) coupled with an adaptive search algorithm, which is described below. The direction set methods reformulate the constrained optimization problem of maximizing (2.10) subject to the equilibrium conditions enumerated in Sections 2.2.1-2.2.5 and the accompanying non-negativity constraints for all prices, quantities, times and land rent differentials \((r_x - r_A)\). That is, for any arbitrary road tax vector, \(\tau\), the interpolation method described in Section 2.3.2 is used to provide the function \(\tau_{Rz}\). The system of equations of Section 2.2.5 is solved whenever the resulting equilibrium utility \(u(\tau)\), \textit{i.e.} the objective value of the reformulated problem, is to be computed. Finite differences are used to approximate the gradient vector \(\nabla u(\tau)\) (where \(u_t(\tau) \approx \frac{u(\tau + \Delta \tau_t) - u(\tau)}{\Delta \tau_t}\)) and the Hessian matrix \(H(\tau)\) for the case of Newton method.

However, in the optimum of the reformulated problem it is not guaranteed that the above non-negativity constraints are respected. When this is the case, a heuristic, derivative-free algorithm which generates random, candidate vectors in a ball \(B_d(\tau)\) is called forth. Solutions from candidate vectors \(\tau'\) in which any non-negativity constraint is violated are discarded. The road tax vector that yields the highest utility in iteration \(k\), \(\tau_k^*\), is used in iteration \(k+1\), in order to draw policy vectors in ball \(B_d(\tau_k^*)\). The radius of the ball, \(d\), the number of draws, the size of \(\tau\), as well as the corresponding points used to interpolate \(\tau\) are updated across iterations, depending on how satisfactory the progress is.