1 Lane differentiation

In congested cities throughout the world, governments are looking for ways to make better use of existing infrastructure and to limit the impacts of congestion on freight and public transport. A subset of policy instruments employed to achieve these aims involves differentiating between traffic lanes either by restricting lanes to certain classes of users or charging vehicles for use of a less congested lane.

The most common example is high-occupancy vehicle (HOV) lanes whereby a lane is reserved for vehicles with either two or more, or three or more occupants. Buses are included. The hope is that HOV lanes will encourage ride-sharing so the same transport task can be achieved using fewer vehicles leading to gains in the efficiency of road use with savings in fuel consumption and emissions. Under-utilisation can be a major problem for HOV lanes. In US, a number of HOV lanes have been converted to high-occupancy toll (HOT) lanes, whereby other vehicles are permitted to travel in the HOV lane on payment of a toll. (BTRE 2002, pp 33-34). HOT lanes are also seen as a 'thin-edge-of-the-wedge' strategy to prepare the public to accept congestion pricing (Fielding and Klein 1993).

Another example of dedicated lanes is truck-only lanes. By separating trucks from other traffic, it is hoped that truck-only lanes will enhance safety and stabilise traffic flow. (Caltrans 2004). A few examples of truck-only lanes exist in the US. Truck-only lanes may be seen as a way to improve freight access to intermodal terminals and ports.

As these examples show, lane differentiation can be achieved either by regulation — decreeing which classes of road users have the right to travel in which lanes — or through pricing — charging different tolls for different lanes. The HOT-lane concept employs both methods.

The objectives of lane differentiation are varied — traffic flow, service quality of public transport, environment, access, safety, and energy. The present paper is concerned almost entirely with the objective of economic efficiency. The value to society of the impacts of lane differentiation is measured by summing the economic benefits and costs regardless of to whom they accrue. The sum total is referred to as 'economic welfare'. The paper also examines the distributional implications for different policies for traffic with different values of time.

2 Economics of traffic allocation to lanes

For the most economically efficient lane allocation outcome, no vehicle should be able to reduce the total cost of travel to all road users by changing lanes. Say the marginal social cost of a vehicle travelling a certain distance in lane 1 was \$3 and it was \$2 in lane 2. Then a change from lane 1 to lane 2 would save society as a whole \$1.

Drivers seeking to minimise their private costs will attempt to change lanes to engender equality of marginal **private** costs between lanes. In equilibrium, vehicles in all lanes will travel at identical speeds. The main difference between private and social costs of travel on a congested road is the externality a road user imposes on other road users by slowing them down. If the mix of vehicles having differing costs is the same for the different lanes, then the externality cost of congestion should be the same for all lanes. The gap between social and private costs should therefore be the same for all lanes. Equality of marginal private costs between lanes should be consistent with equality of marginal social costs. The equilibrium split of traffic produced by the free market appears to be the most economically efficient split.

Hence, on the face of it, lane differentiation seems to be an economically inefficient policy. But this need not be the case. The marginal cost of a vehicle, whether private or social, depends on both the value of time and the speed (which determines the time taken). In mathematical terms: cost = f(value of time, time taken). This leads to an interesting possibility. If the traffic could be split up so that the high value-of-time traffic travelled in a less congested lane or lanes at a high speed, and low value-of-time traffic travelled in a more congested lane or lanes at a lower speed, the marginal social costs could still be equalised. At the same time, by offering road users a choice of two price–quality combinations instead of one, the economic welfare of road users as a whole could be improved, and the road space used more efficiently.

For brevity, the rest of this paper is written, for the most part, in terms of two lanes, lane 1, the slow lane, and lane 2, the fast lane. It should be understood that, in practice, there are likely to be at least two slow lanes and at least one fast lane.

Say a road user incurs c_1 in generalised costs in the slow lane and c_2 in the fast lane, $c_1 > c_2$. By switching from the slow lane to the fast lane, he or she saves $c_1 - c_2$ in costs. At the same time, he or she pays $\pi_2 - \pi_1$, the symbol π representing the charges levied for travel in the respective lanes. The driver will stay in lane 1 if $c_1 - c_2 < \pi_2 - \pi_1$ and switch to lane 2 if $c_1 - c_2 > \pi_2 - \pi_1$. If cost is a linear function of value of time, we can write $c_1 - c_2 = v(t_1 - t_2)$ where v is the value of time for a vehicle, comprising both driver/passenger costs and vehicle operating costs. t is the time taken to travel the length of road in the lane indicated by the subscript. All vehicles having a value of time below $v^* = (\pi_2 - \pi_1)/(t_1 - t_2)$ will travel in lane 1 and all vehicles having a value of time above v^* will travel in lane 2. In line with Verhoef and Small (1999, p.5), we refer to v^* as the critical value of time.

The times taken in the respective lanes depend upon the total traffic in the lanes. Hence to ensure that traffic moves faster in lane 2 than in lane 1, there has to be less traffic in lane 2, that is, $q_1 > q_2$ to ensure that $t_1(q_1) > t_2(q_2)$ where q_1 and q_2 refers to the volumes of traffic in the respective lanes.

3 Optimal lane pricing

The theory of optimal pricing for congested roads with undifferentiated lanes is well known. If we set up an optimisation problem to maximise willingness-to-pay (the area under the demand curve between zero and the quantity consumed) minus total costs, we find that the total price paid by users ($p = c + \pi$) should equal the marginal social cost. However, average private cost falls short of marginal cost by an amount equal to the externality cost that the marginal road user imposes on other road users. A congestion charge equal to that externality cost has to be levied to ensure that the quantity of travel is at the economically efficient level. In mathematical terms:

$p - c = \bar{v}\frac{dt}{dq}q = \pi$ where $\bar{v}$ is the average value of time and, assuming that road-user cost is proportional to time taken, $c = vt$. An additional vehicle entering the traffic stream slows all other vehicles down increasing their time taken by dt/dq. Multiplying this by the average value of time for all users gives the additional cost imposed on the average user. Multiplying by the total number of users, q, results in the full externality cost.

To analyse the two-lane situation, the demand curve has to be split into a series of demand curves each having its own value of time. If we were to line up these demand curves side by side, in order of value time, we would have a demand surface $p(q, v)$ as illustrated in figure 1. With lane differentiation by price, the surface is split at the critical value of time, v^*. The time–quantity relationships for the two lanes, $t_1(q_1)$ and $t_2(q_2)$, are independent of user's values of time. However, the same travel time requirement will have different cost implications for different users, depending on their value of time. The cost surface for lane 1 is $c[t_1(q_1), v]$ or, with costs proportional to time, $vt_1(q_1)$. Similarly for lane 2, the cost surface is $c[t_2(q_2), v]]$ or $vt_2(q_2)$. For any given value of time, the equilibrium traffic quantity occurs at the intersection of the demand surface with the price surface, given by the cost surface plus the charge, $c[t_1(q_1), v] + \pi_1$ for lane 1 and $c[t_2(q_2), v] + \pi_2$ for lane 2. The total quantity for lane 1 is found by integrating the quantities for values of time from zero to *v**, and for lane 2, from *v** to infinity.

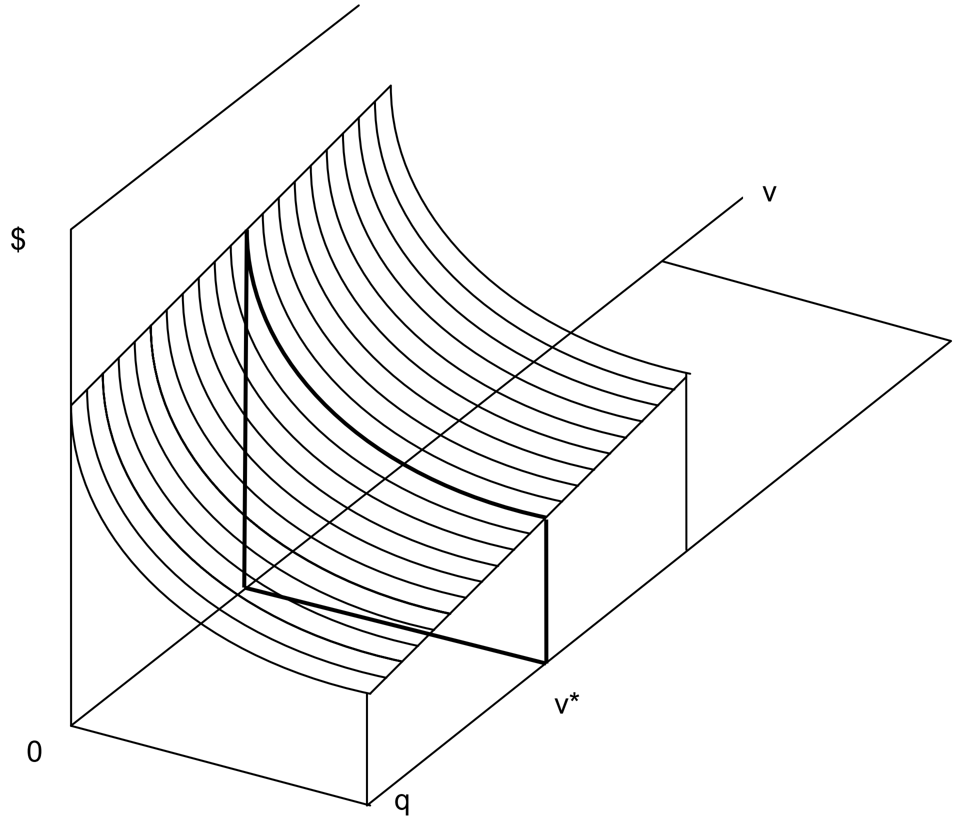

Figure 1 Demand surface with lane split

Total economic welfare is the sum of the total willingnesses-to-pay minus total costs for the two lanes. Willingness-to-pay is the volume under the demand surface bounded by quantities and *v**. The mathematical derivation of the optimal lane prices is provided in the appendix. The appendix is in terms of a general relationship between cost and time to allow for the fact that vehicle operating costs are not proportional to time taken. For the exposition, here, a proportional relationship, $c = vt$, is assumed. Setting the derivatives of economic welfare with respect to the quantities in equal to zero, the yields the normal optimal congestion pricing results for each lane.

$\bar{v}_1 \frac{dt_1}{dq_1} q_1 = \pi_1$ and $\bar{v}_2 \frac{dt_2}{dq_2} q_2 = \pi_2$

Setting the derivative with respect to v* equal to zero produces the result:

$v^* t_1 + \bar{v}_1 \frac{dt_1}{dq_1} q_1 = v^* t_2 + \bar{v}_2 \frac{dt_2}{dq_2} q_2$, that is, the marginal social cost for a vehicle with a value of time of v*, should be the same for both lanes. Since v* is the same regardless of lane, $v^* t_1 > v^* t_2$, so with an optimal traffic split $\bar{v}_1 \frac{dt_1}{dq_1} q_1 < \bar{v}_2 \frac{dt_2}{dq_2} q_2$. If the marginal vehicle (one with value of time *v**) switches from lane 1 to lane 2, it saves $v^* t_1 - v^* t_2$ in private costs, but, with $\bar{v}_2 > \bar{v}_1$, the cost imposed on other vehicles, $\bar{v}_2 \frac{dt_2}{dq_2} q_2 - \bar{v}_1 \frac{dt_1}{dq_1} q_1$, is exactly offsetting. At this optimal lane split, economic welfare cannot be improved by altering the allocation of vehicles to lanes.

Substituting the optimal congestion prices into the expression for the optimal lane split: $v^*(t_1 - t_2) = \pi_2 - \pi_1$, hence: $v^* = \frac{\pi_2 - \pi_1}{t_1 - t_2}$.

For any given traffic split there will be a pair of optimal congestion prices, one price for each lane. However, there with will be only one pair of congestion prices consistent with the level of v* needed to obtain the optimal split. Optimal lane pricing is therefore fully consistent with optimal congestion pricing. Just as optimal congestion pricing requires charges to vary with location and time of day, the optimal charges could also vary with lanes.

The appendix goes on to derive a formula for the second-best optimal lane charge for lane 2, when the charge for lane 1 is fixed. This is relevant for situations where it is not possible to levy any charge for lane 1, but lane 2 is a toll lane. It could also apply where there is a toll in lane 1 that is fixed for political or legal reasons. The appendix explains the rationale behind the formula in detail. Briefly, with the charge in lane 1 set below the optimal congestion price, an additional vehicle joining lane 1 causes a welfare loss equal to the difference between marginal social cost and price paid (average private generalised cost plus any charge, which is the consumer's valuation of the marginal unit). The second-best optimal price for lane 2 is that which splits the traffic in such a way that the marginal welfare losses in the two lanes are equal for a vehicle with value of time v* switching lanes (the gain from leaving the from-lane equals the loss in the to-lane). In the calculating these welfare losses, the formula takes account of all the secondary effects on other vehicles from a vehicle changing lanes: crowding out of existing vehicles when an extra vehicle moves into a lane; induced traffic in the lane from which the vehicle leaves; and the effect on quantity in lane 2 from the change in the charge necessary to shift the marginal vehicle.

Figure 2 compares the different charges. With undifferentiated lanes, there is a single optimal congestion charge. With optimal lane pricing, the charges for lanes 1 and 2 lie on either side of the undifferentiated optimal charge. In the second-best situation, say there was no charge at all levied for lane 1. If we retained the first-best charge for lane 2, the gap between the two charges would be too great — there would be too little traffic in lane 2 and too much in lane 1. So the second-best optimal change for lane 2 lies below the first-best optimal charge.

The mathematical appendix also derives the 'third-best' optimal charge where a class of vehicles is allowed to travel in the fast lane without charge. The fast lane would be classed as a HOT lane if the select class of vehicles were HOVs. It has been assumed throughout this paper that there is no substitutability between single- and high-occupancy vehicles. The extent to which for HOV lanes are effective at encouraging ride-sharing is debated. Fielding and Klein (1993) argued that 'Current HOV lanes are not very effective at reducing traffic; 43 per cent of car-poolers are members of the same household'. On the other hand, Gard reported a HOV lane scheme in San Francisco increasing carpool ridership by 65 per cent. (quoted in BTRE 2002). Extension of our model to allow for inter-relationships between demand curves for some classes of vehicles is a development to consider in the future. It would not change the result for first-best optimal lane pricing and would most likely have only limited effect on the second-best pricing results. The welfare results for HOV and HOT lanes, however, are likely to be more favourable.

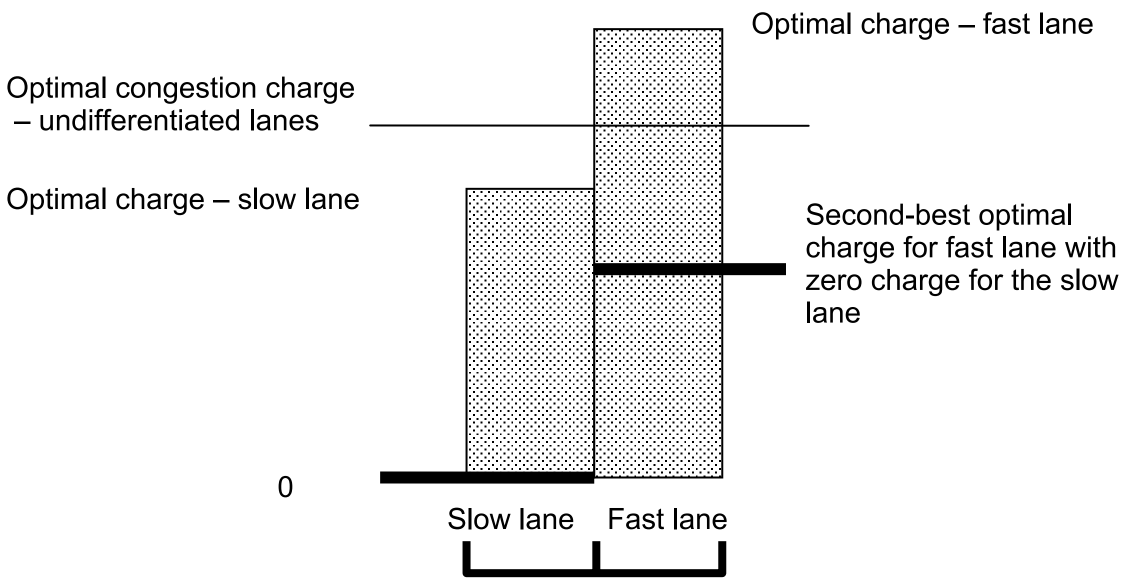

Figure 2 Relative sizes of optimal lane charges

The formula for the optimal HOT lane charge is similar to the second-best optimal charge in that the marginal welfare losses between the lanes are equated, allowing for secondary effects associated with a vehicle switching lanes. There is an additional secondary effect to take into account in the optimal HOT lane case. If the charge is decreased, causing traffic to shift from the slow to the fast lane, there is crowding out of HOV vehicles to consider as well as for the toll-paying vehicles in the fast lane. The HOV vehicles have their own marginal welfare loss distinct from the toll-paying vehicles in the fast lane.

4 Relationship to prior literature

Lane differentiation has been attracting increasing attention in the academic economic literature over the last several years. As the problem is mathematically complex, the approach taken is invariably to develop a mathematical model, make assumptions about orders of magnitude, and to undertake quantitative experiments with the model to compare the economic efficiency, revenue and other outcomes of different assumptions and policies. The present paper follows this same approach.

Often the models apply to two alternative routes but they are effectively the same as if the routes were different lanes on the same road.

Many of the papers assume homogeneous vehicles in terms of value of time. As Verhoef and Small (1999) observe, such an approach leads to under-estimation of the welfare gains from differential lane pricing. Small and Yan (2001) assuming homogeneity within lanes, but heterogeneity between lanes, found that the efficiency gains from lane pricing, whether first-best or second-best, increases with the assumed level of heterogeneity between lanes. Some models, such as Yang and Huang (1999), focus on the decision about whether or not to rideshare, but assume a single value of time.

The present paper adds value in that it:

- emphasises the distribution of road users across values of time, including deriving such a distribution in an original way;
- derives mathematical expressions for first-best and second-best optimal lane prices and explains the rationales behind them;
- develops an approach to modelling and evaluating lane differentiation proposals;

- provides some new insights into lane differentiation policies and discusses the policy implications.

5 Spreadsheet experiments

5.1 Model description

In order to gain further insights into lane pricing, spreadsheet experiments were undertaken using values for costs and quantities set at realistic levels. Finding equilibrium prices and quantities involves solving simultaneous equations. This was done using Excel's circular reference facility, which finds solutions through iteration.

The assumptions in table 1 were made in developing the demand system. Proportions of buses and cars with four or more occupants were assumed to be too small to warrant inclusion.

Table 1 Assumed values of time and coefficients of variation

	People $/h (%)	Freight $/h (%)	Vehicles $/h (%)	Total $/h (%)
Private car 1 person	9.23 (30)	0	8 (20)	17.23 (18.6)
Private car 2 people	9.23 (30)	0	8 (20)	26.46 (16.0)
Private car 3 people	9.23 (30)	0	8 (20)	35.69 (14.2)
Business car 1 person	29.52 (20)	0	8 (20)	37.52 (16.3)
Business car 2 people	29.52 (20)	0	8 (20)	67.04 (12.7)
Business car 3 people	29.52 (20)	0	8 (20)	96.56 (10.7)
Rigid truck	19.69 (5)	9.31 (30)	34 (30)	63 (16.9)
Articulated truck	20.94 (5)	27.57 (30)	40 (30)	88.51 (16.5)

Notes: Coefficients of variation are shown in brackets. Costs are resource costs, that is, they exclude taxes on non-labour inputs such as GST and fuel excise. Prices paid by road users were assumed to equal resource costs. Distinguishing between resource and financial costs was considered to be an unnecessary refinement. Coefficients of variation for the totals were obtained by summing variances of the components of total cost.
Sources: Austroads (2004) for values of time. BTRE Road Infrastructure Assessment Model with parameter values from Austroads (2004) for vehicle operating costs. Coefficients of variation are the author's assumptions.

The values of time and weighted average vehicle occupancy rates (see table 2) are those currently recommended by Austroads (2004) for urban traffic for undertaking cost–benefit analyses of road improvement projects. The vehicle occupancy rates appear high, but it was decided to keep with Austroads values.

Table 2 Assumed vehicle occupancy rates and proportions

Private car 1	Private car 2	Private car 3	Weighted average	Business car 1	Business car 2	Business car 3	Weighted average
70%	20%	10%	1.4 people	50%	40%	10%	1.6 people

Source: Austroads (2004) for the 1.4 and 1.6 vehicle occupancy rates. Proportions are the author's assumptions but set to be consistent with the Austroads vehicle occupancy rates.

The assumed vehicle mix was 90% cars, 5% rigid trucks and 5% articulated trucks. The assumed split of cars between private and business travel was 78:22 was based on Survey of Motor Vehicle Usage statistics. To make trucks comparable with cars in terms of use of road space, truck numbers had to be converted into passenger car units (PCUs). A rigid truck was assumed to account for 1.5 PCUs and an articulated truck, 3.0 PCUs. The resultant PCU split and values of time per PCU are shown in table 3.

It was assumed that values of time are normally distributed around the means shown in table 3 with the coefficients of variation from table 1. The distribution so-obtained was assumed to apply in the situation without congestion pricing. The tiny proportions of PCUs with values of time below $6 per hour and above $120 per hour were aggregated. The distributions of traffic in PCUs per hour, with and without congestion pricing, are shown in figure 3 with the traffic split into one-cent-wide bins for values of time. The distribution has two peaks caused by the dominance of private cars with one and two occupants, and has a long tail to the right.

Table 3 Assumed PCU split and mean values of time per PCU

	Private car 1 person	Private car 2 people	Private car 3 people	Business car 1 person	Business car 2 people	Business car 3 people	Rigid truck	Artic-ulated truck
Percent of PCUs	31.20	24.96	6.24	12.32	3.52	1.76	6.67	13.33
Value of time ($/PCU/h)	17.23	26.46	35.69	37.52	67.04	96.56	42.00	29.50

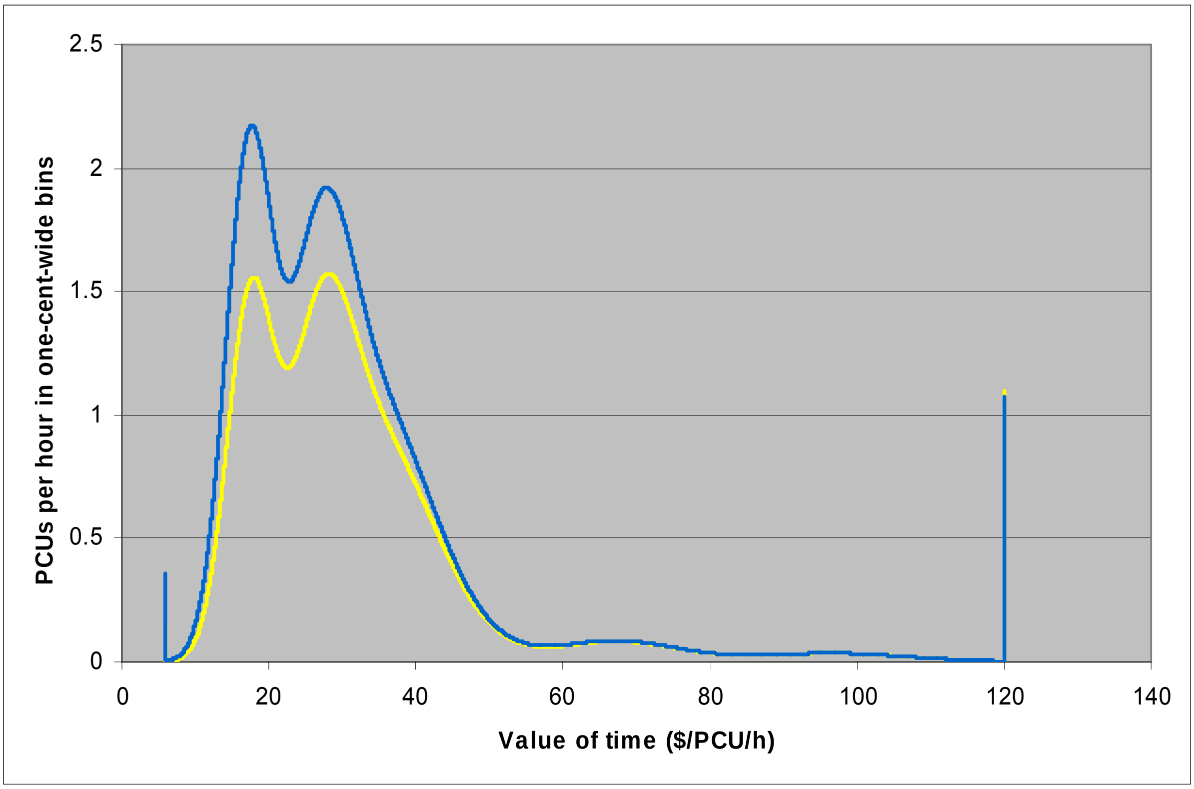

Figure 3 Distribution of PCUs across values of time without (dark curve) and with optimal congestion pricing (light curve)

To make the problem tractable for a spreadsheet, it was necessary treat the demand surface as a set of discrete demand curves. The market was segmented into 11,401 demand curves for values of time at 1 cent intervals from $6 to $120. Ideally, the charge would be set exogenously and the split of traffic between the two lanes derived therefrom. However, the discrete nature of the demand surface meant that the split of demand curves between the two lanes had to be set manually for each spreadsheet experiment involving lane differentiation by price. The implied charge was then calculated. The large number of demand curves was necessary to give the model the required level of sensitivity.

A constant-elasticity form of demand curve was was assumed — $p=aq^{-0.8}$ where *p* is the total price paid by a road user consisting of the cost incurred plus charge paid, *q* is quantity, *a*, a constant, and –0.8 the elasticity. In the model with no congestion pricing, the *a*'s were adjusted to obtain the desired frequency distribution of quantities with differing values of time.

The curve for time to traverse the length of road as a function of quantity was assumed to be $t = t_0 / (1 - VCR)^{0.3}$ where t_0 is the time taken travelling at free speed (absence of interference from other vehicles) and *VCR* is volume–capacity ratio. For the purposes of our spreadsheets, the length of road was taken to be one kilometre, the free speed 100 km/h and the lane-capacity 2400 PCUs per hour. The volume–capacity ratio (VCR) in the absence of congestion pricing was assumed to be 0.72. At higher ratios, Excel had increasing difficulty converging to solutions, especially with lane differentiation. The road was assumed to consist of three lanes per carriageway. Our model deals with one carriageway only. Thus, total capacity for a carriageway was 7200 PCUs per hour, total actual traffic was 5184 PCUs per hour, and speed 68.3 km/h.

Results of the experiments are summarised in table 4.

5.2 Optimal congestion pricing without lane differentiation

The optimal congestion price without lane differentiation was 16.4 cents, resulting in a reduction in VCR to 0.58 and an increase in speed to 77.2 km/h. Figure 3 shows that congestion pricing has a larger proportional impact on traffic with lower values of time. It has reduced the peak associated with single-occupant private cars to around the same level as that associated with dual-occupancy private cars. The average value of time has risen from $29.63 to $31.03. The reason is evident in figure 4, which shows the distribution of increases in total private generalised costs for PCUs (including charges) over the situation where there is no road charging (see the ‘cong p’ line for optimal congestion pricing without lane differentiation). The increase is smaller for traffic with higher values of time because the increase in speed is worth more to higher value-of-time traffic. Traffic with values of time above $96.61/PCU/h experience falls in generalised costs because the benefit of the time savings to them exceeds the cost imposed by the congestion charge.

Table 4 Policy options compared

	No pricing	Optimal congestion pricing	Optimal lane pricing	Second-best optimal lane pricing	HOV lane (3+)	Optimal HOT lane pricing (3+)
Charge (cents/km)	0.0	16.5	15.3, 18.6	0.0, 8.1	0.0, 0.0	0.0, 7.6
VCR	0.72	0.58	0.62, 0.51	0.76, 0.58	0.85, 0.22	0.76, 0.60
Speed (km/h)	68.3	77.2	75.0, 80.1	65.2, 77.1	56.3, 92.8	65.3, 76.0
Lane split (ratio)	1: 1: 1	1: 1: 1	1.2: 1.2: 1	1.3: 1.3: 1	3.9: 3.9: 1	1.3: 1.3: 1
Critical value of time (v*) ($/h)	na	na	34.56	34.06	na	35.13
Average value of time ($/h)	29.63	31.03	23.68, 49.00	23.00, 47.93	27.94, 49.07	23.11, 47.27
PCU quantity increase (index)	0.0	-100.0	-97.8	-14.4	-54.8	-10.2
Welfare increase (index)	0.0	100.0	106.8	32.6	-8429.4	-8223.0
Revenue increase (index)	0.0	100.0	99.3	16.5	0.0	10.9

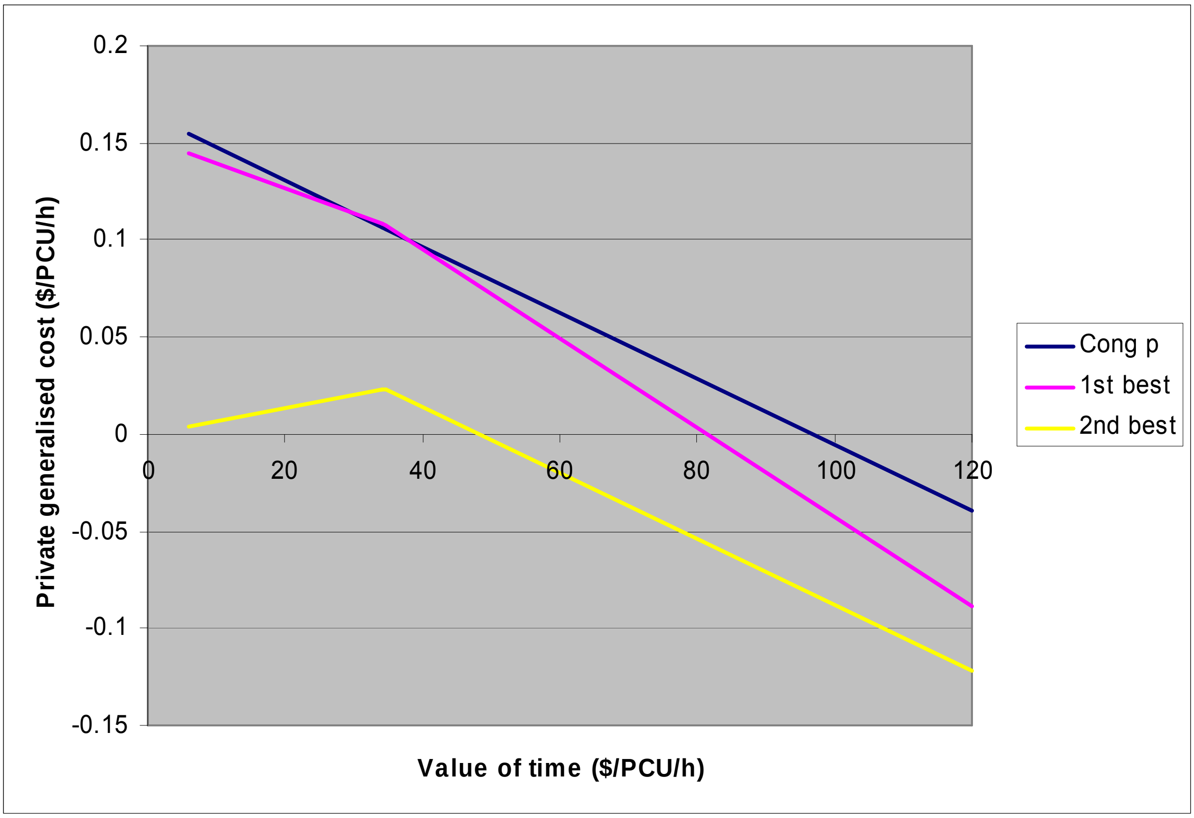

Figure 4 Increases in private generalised costs (including charges) over no charging case

Optimal congestion pricing leads to a quantity decrease of 1028 PCUs/hr, a welfare increase of $162.93 /h and revenue of $684 /hr. These are used as the standards of comparison for reporting the impacts of lane differentiation.

5.3 First-best optimal lane pricing

As expected, first-best optimal lane pricing leads to prices, VCR's and speeds on either side of the respective levels for optimal congestion pricing. There is a small increase in total traffic throughput and the welfare increase, at 6.8%, is small in comparison with optimal congestion pricing. This suggests that if optimal congestion pricing is feasible, there is not a great deal to be gained from adding the additional dimension of lane pricing. Indeed, given the margin for error in setting the optimum prices and the additional administrative costs, the modest welfare gains from optimal lane pricing compared to optimal congestion pricing could easily be negated. However, the same could be said about comparisons between congestion prices having differing degrees of variability with respect to time of day or location.

The same spreadsheet experiments were performed for the no-pricing situation with VCR's of 0.6 and 0.4. Although the gains from optimal congestion and lane pricing are smaller in absolute terms, as a percentage of the congestion pricing gain, the gain from optimal lane pricing is 11.2% for an initial VCR of 0.6, and 26.0% for an initial VCR of 0.4. At lower levels of congestion, the gains from congestion pricing are less compared with the no-pricing situation. However, the value of splitting the traffic stream tapers off more slowly as congestion falls. In the VCR of 0.4 case, the critical value of time is little changed ($34.39) and the reductions in the absolute differences between the speeds (86.4, 90.1 km/h) and VCRs (0.38, 0.29) between lanes are far from proportionate, suggesting that a significant amount of the gains from lane differentiation remain. The policy implication is that, in a road network where optimal congestion pricing is in place, gains from lane pricing may be more certain of being realised on moderately congested roads than on heavily congested roads.

Figure 4 shows that the impacts of optimal lane pricing on total private generalised costs of individual PCUs are not markedly different from with optimal congestion pricing without lane differentiation except for PCUs with high values of time. The kink in the curve occurs at the critical value of time. Traffic with the highest values of time in the slow lane and traffic with the lowest values of time in the fast lane do less well compared with congestion pricing. Traffic with the highest values of time benefit most. It is not surprising that a switch from one price–quality combination to two would disadvantage consumers in the middle and advantage consumers at the extremes. However, the loss to road users around the critical value of time compared with congestion pricing is practically negligible.

5.4 Second-best optimal lane pricing

When the charge for the slow lanes is constrained to zero, charging for use of the fast lane has more to offer. The VCR's and speeds for the lanes lie on either side of those for the no-pricing situation and the gaps between them are more pronounced than for optimal lane pricing. The welfare increase is 32.6% of the gain from optimal congestion pricing

By way of comparison, Verhoef and Small (1999, p. 10) obtained a welfare gain of 22.9% for second-best optimal pricing on one of their parallel links (effectively, the fast lane) expressed as a percentage of the welfare gain from first-best optimal lane pricing compared with no pricing. The comparable percentage in our case is 30.6%. Parry (2002, p. 347), also assuming heterogenous vehicles with respect to value of time, obtained results ranging from 3% to 33%, again expressed as percentages of the gain from first-best optimal lane pricing. The spread of values in Parry's results arises from varying the demand elasticity, the degree of heterogeneity of road users, and the level of congestion.

It is interesting to observe that for the three options shown in table 4 that involve differential lane pricing, the critical values of time, the lane splits, and average values of time in lanes are very similar. Many more experiments would be required, however, before one could generalise about this.

As with optimal lane pricing, the gain from second-best optimal lane pricing as a percentage of the gain from optimal congestion pricing is greater on less congested roads. With a VCR of 0.6 in the no pricing situation, the welfare gain is 35.7% of that for congestion pricing, and for a VCR of 0.4, the welfare gain is 49.7%. Such a relationship is also evident in Parry's (2002, p. 347) modelling results.

Figure 4 shows the impact of second-best optimal lane pricing on total private generalised costs. With less revenue being raised, all road users are better off compared with optimal congestion and lane pricing. Traffic in the slow lane is disadvantaged relative to the no-pricing situation because of the slower speed at which they are forced to travel. For traffic in the fast lane with values of time between the critical value and \$47.90/PCU/h, the toll exceeds the benefit of the higher speed. Traffic in the fast lane with values of time above \$47.90/PCU/h receives a net benefit.

5.5 Lane differentiation through regulation

In our spreadsheet model, if charging for road use were ruled out altogether, and the policy-maker had the power to declare which classes of vehicles were permitted to use the fast lane, a number of feasible combinations exist. The feasibility of combinations is limited by assumptions that it is not possible to separate private and business cars, and that vehicle occupancy in the fast lane must be either two and above, or three and above.

All feasible combinations were tested as well as a few non-feasible combinations. The best feasible combination on welfare grounds was to restrict the fast lane to HOVs with three people (HOV3+). Single- and dual-occupancy vehicles and rigid and articulated trucks must travel in the slow lanes. The welfare loss is 84 times the gain from congestion pricing. All other feasible combinations lead to even greater losses. Had it been possible to allow business cars with two or more occupants to travel in the fast lane and not private cars, the welfare loss could have been reduced to seven times the gain from congestion pricing.

The difference in VCRs and speeds between lanes is quite marked. Looking across the alternative combinations tested, a critical factor seems be the difference in weighted-average values of time between traffic in the slow and fast lanes. In the best feasible case, the average values of time are $28 and $49 per hour respectively. The most economically efficient case, where only business cars with two or more occupants use the fast lane, had average values of time of $27 and $77. A rule-of-thumb in deciding which groups to allow in the fast lane would be to seek the largest possible difference in the weighted average values of time between lanes, but without making the fast lane so exclusive that vehicles travel at close to free speed. This is contrary to the conventional wisdom that ensuring adequate utilisation of the fast lane is the prime determinant of economic efficiency — a point we return to in the next sub-section.

Figure 5 shows 95% confidence intervals around the assumed average values of time for each vehicle type. Business cars with two or three occupants are quite distinct from the other types in terms of the ranges covered. It points to the reason why the most economically efficient regulatory split would be to grant them exclusive use of the fast lane.

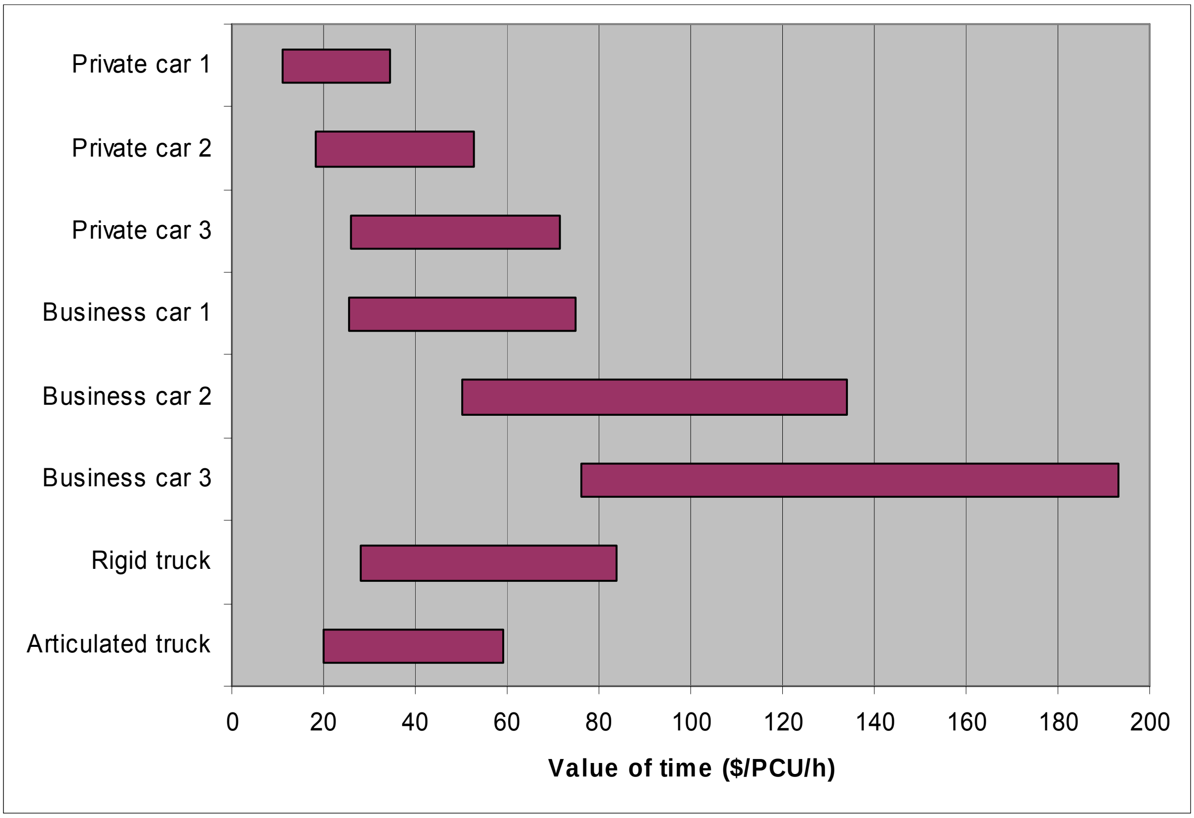

Figure 5 Confidence intervals at 95% around average values of time by vehicle type

While a whole rigid or articulated truck has a relatively high value average value of time compared with a private car, using Austroads values of time for crew and freight, their values per PCU are not much different from the values for high-occupancy private cars. Hence, granting trucks access to the fast lane, under our assumptions, reduces total economic welfare. In our model, a truck-only lane leads to a welfare loss of 177 times the gain from introducing congestion pricing.

An argument in favour of special treatment for trucks in lane allocation is that there are additional benefits from increased reliability that models, such as ours, do not recognise, and that reliability benefits are of greater significance for freight than for passengers. Small and Yan (2001, pp. 311-2) justified omitting reliability benefits on the grounds that: 'Value of time here proxies for value of reliability as well, since travel time and reliability are closely correlated in such corridors'.

Permitting HOVs with two or three people (HOV2+) to travel in the fast lane in our model is not feasible because the number of HOVs is sufficiently large to make the speed in the HOV lane slower than the speed in the other lanes. Although this result is an artefact our assumptions, it is not uncommon in practice. Poole and Orski (1999. part 3B and 2000, pp 17–18), presenting arguments in favour of HOT lanes, observe that: 'When HOV2+ lanes are converted to HOV3+ lanes because of severe congestion, the change usually results in unused capacity'. They cite the example of the Katy HOV lane in Houston, Texas where authorities have switched back and forth between HOV2+ and HOV3+ restrictions for the same lane because of 'heavy congestion' with the HOV2+ rule, and 'excess capacity' with the HOV3+ rule.

Our modelling demonstrates that pricing solutions are likely to be vastly superior to regulatory solutions on economic welfare grounds. First, as figure 5 shows, there is considerable overlap between the value of time distributions for different vehicle classes. Under any traffic split determined by regulation, there will be vehicles with lower values of time per-PCU travelling in the fast lane than for vehicles in the slow lane. An immediate improvement in welfare would accrue if these vehicles could swap lanes. Second, regulatory solutions cannot distinguish between private and business cars or between trucks having different values of time. Lane pricing cleanly segregates the traffic stream at a single value of time, solving the problems of overlapping value of time distributions and limited ability to discriminate between vehicles.

Third, the price difference between lanes can be finely adjusted to obtain the optimal split of traffic between lanes. With regulation, lane allocation can be undertaken only for large classes of vehicles.

A fourth point in favour of lane allocation though pricing rather than through regulation is that part or all of the benefits from congestion pricing can be realised. Optimal first-best lane pricing yields the full benefits of congestion pricing plus the benefits of splitting the traffic stream by value of time. Second-best optimal lane pricing, whereby a toll is charged for the fast lane and not for the slow lane, is able to realise a portion of benefit of congestion pricing because the charge for the fast lane and the higher generalised cost in the slow lane deter some traffic for which the marginal social cost is less than willingness-to-pay.

While welfare outcomes from lane differentiation through regulation must necessarily be inferior that from second-best optimal pricing, they need not be negative. A positive welfare gain is conceivable if there were a large class of vehicles with relatively high values of time and that class would be distinguished for reservation of the fast lane.

5.6 HOT lanes

In the final experiment, the HOV3+ lane was converted to a HOT lane. HOV3+ vehicles travel in the fast lane for free, while all other vehicles must pay a toll to use the fast lane. The toll was set at the optimal 'third-best' level. The optimal charge, VCRs, speeds, lane splits, critical and lane-average values of time, and quantity of PCUs all turn out to be similar to those for second-best optimal lane pricing. However, the welfare outcome is vastly different, with a loss of 82 times the gain from optimal congestion pricing. There is only a small improvement, in proportional terms, on the welfare outcome for the HOV2+ lane.

Although, the switch from HOV to HOT lane has lead to more balanced lane utilitisation, the welfare gains have been largely offset by losses from slowing down the high-occupancy business cars in the fast lane from 93 to 76km/h. The HOT lane option is vastly inferior on economic efficiency grounds to the second-best optimal lane pricing solution for two reasons. First, there are HOV3 private cars travelling in the fast lane with values of time below the critical value of $35.13 — 3.2% of total PCUs. Second, by reducing capacity in lane 2 for other vehicles, these HOV3 private cars drive up the critical value of time from $34.06 to $35.13, crowding out PCUs with values of time between $34.06 and $35.13 (2.0% of total PCUs) from the fast lane to the slow lane.

Under different sets of assumptions, a switch from a HOV to HOT lane would undoubtedly be more beneficial on welfare grounds. Nevertheless, these results sound a warning about the much-vaunted benefits from HOT lanes over HOV lanes in terms of achieving better lane utilisation (for example, Dahlgren (2002, p. 241) and Poole and Orski (2000, p. 15)). If HOV lanes are able let the highest-value-of-time road users through quickly, they may not be quite the ‘inefficient use of scarce road space’ that is claimed (Poole and Orski 1999). The costs of underutilised road space are highly visible, unlike the benefits of faster travel for high-value-of-time vehicles.

6 Policy implications

The model described in this paper gives greater emphasis to the heterogeneity of road users than most other models developed to examine lane differentiation. Consequently, it highlights the value of lane differentiation in separating out high-value-of-time traffic so it can travel at faster speeds. If properly applied, lane differentiation can result in a more efficient utilisation of road space. Lane differentiation becomes more worthwhile the greater the degree of diversity among road users in the value of time.

Optimal congestion pricing can be improved upon by differential lane pricing, but the benefits may not be large compared with the overall benefit from introducing congestion pricing. The percentage benefit is greater on less congested roads. In terms of relative attractiveness, second-best optimal pricing compared with no pricing is a better proposition.

Lane differentiation via regulation, such as HOV lanes and truck-only lanes, is vastly inferior on economic efficiency grounds to pure pricing solutions because regulation is a clumsy instrument for achieving segregation of high- and low-value-of-time traffic. It would require exceptional circumstances for a HOV or a truck-only lane to produce an economic welfare gain.

Although a truck has a high value of time for the vehicle as a whole, on a per PCU basis, its value of time is comparable to that of a private car. From an economic efficiency viewpoint the value of truck-only lanes is therefore highly questionable.

Conversion of under-utilised HOV lanes to HOT lanes may not be as beneficial on economic efficiency grounds as might be thought, because the gains from more balanced lane utilisation are, in part, offset by loses caused by slowing down high-occupancy vehicles in the fast lane having high values of time.

In considering the economic efficiency of different forms of lane differentiation, it is important not only to take account of lane utilisation, but also the degree of segregation achieved of traffic with different values of time. The amount of overlap between the ranges of values of time between the slow and fast lanes is a major indicator of economic efficiency as well as levels of lane utilitisation.

In developing lane differentiation policies, there are likely to be conflicts between economic efficiency and other objectives. Objectives such as encouraging ride-sharing and providing better access for public transport and freight are difficult to achieve without regulating lane usage. But from the point of view of the economic efficiency objective, pricing solutions usually perform much better than regulatory solutions. Where the economic efficiency objective conflicts with other objectives, it is highly desirable that decision makers be informed of the costs of economic efficiency sacrificed to achieve other objectives. This information can be gleaned only by undertaking cost–benefit analyses of policies.

Better evaluation of policies should also shed light on how well they achieve their stated objectives too. For example, BTRE (2002, p 33) took the view that 'the environmental gains from HOV lanes are often negative' as well as their efficiency gains.

7 Further research and development

There are some obvious lines of further research on this topic using the approach developed in this paper.

The modelling work undertaken here could be extended to cover a much greater variety of assumptions about value of time distributions and other traffic characteristics and lane configurations. The work would benefit from better information about traffic compositions and distributions of values of time.

A fuller consideration of HOV and HOT lane issues requires assumptions to be made about relationships between the demand functions for private cars with different occupancy rates, and possibly also between business cars with different occupancy rates.

The Excel spreadsheet models used in the quantitative work for this paper have the advantage of a high level of transparency in their calculations. However, they were unable to handle scenarios with very high volume–capacity ratios. In searching for solutions, Excel tends to push road use beyond the VCR of one where solutions are non-existent. Constraining VCR to an upper limit can send the spreadsheet into an infinite loop. Alternative modelling software is available that can handle large simultaneous equation problems.

In the long-term, a package might be developed to assist users to estimate the economic benefits and costs, revenues, traffic speeds, optimal prices, emissions levels and other variables to assist in evaluation of lane differentiation policies. It could develop a distribution of values of time based on data and assumptions inputted by the user along the lines shown in this paper. Allowance for time-of-day variation in demand would be desirable. Such a model could take account of the differences between private and social costs due to taxes and the non-linear relationship between vehicle operating costs and time.

8 Mathematical appendix

8.1 Optimal congestion price — simple case

The simple case of optimal congestion pricing without lane differentiation using our notation and approach is presented here first to facilitate understanding of the more complex derivations below.

q = quantity of traffic in passenger car units per hour
t(q) = time taken as a function of traffic volume $dt/dq \geq 0$
c(t) = is average generalised cost per vehicle as a function of time taken
MSC = marginal social cost
p(q) = inverse demand curve
π = congestion charge or toll
p = total price paid by a vehicle per passenger car unit $= c + \pi$
w = economic welfare

Maximise economic welfare = willingness-to-pay – total social costs

$$w = \int_0^q p(q)dq - c[t(q)] \;; \qquad \frac{\partial w}{\partial q} = p - c - \frac{dc}{dt}\frac{dt}{dq}q = p - MSC = 0 \;; \qquad p - c = \frac{dc}{dt}\frac{dt}{dq}q = \pi$$

If cost is assumed to be a linear function of time taken, that is, $c = \bar{v}t + k$, where $\bar{v}$ is the weighted average value of time and *k* a constant, then $\frac{dc}{dt} = \bar{v}$ and $\pi = \bar{v}\frac{dt}{dq}q$. Note that in practice, vehicle operating costs are not a linear function of time taken. However, it is a convenient approximation given that driver and passenger time costs outweigh vehicle operating costs.

8.2 First-best optimal lane pricing

All notation for the lane pricing derivations is the same as for the previous sub-section, except that subscript 1 indicates that the variable refers to lane 1, the slow lane, and subscript 2 indicates the variable refers to lane 2, the fast lane.

v = value of time, a continuous variable ranging from zero to infinity
c(t,v) = average generalised cost as a function of time taken and value of time. Here, value of time comprises both driver/passenger time and vehicle operating costs.
The two lanes have different t(q) functions, $t_1(q_1)$ and $t_2(q_2)$, to allow for the fact that lane 1 may represent two or more slow lanes and lane 2, one or more fast lanes.
q(q,v) = demand curve for a given value of time
p(p,v) = inverse demand curve at a given value of time
v* = critical value of time, the value of the time at which the traffic splits between the slow and fast lanes; the highest value for the slow lane and the lowest for the fast lane
c* and MSC* are the respective average and marginal social costs for traffic having a value of time v*.
q* = q(p,v*) is the quantity of traffic having a value time v*. It would switch lanes as a result of a one dollar change in v*.

$$q_1 = \int_0^{v^*} q(p,v)dv \text{ and } q_2 = \int_{v^*}^{\infty} q(p,v)dv$$

If we assume a linear relationship between average cost and time, $\frac{\partial c}{\partial t} = v$ and

$$\bar{v}_1 = \frac{\int_0^{v^*} \frac{\partial c}{\partial t(q_1,v)} q(p,v)dv}{\int_0^{v^*} q(p,v)dv} = \frac{1}{q_1}\int_0^{v^*} vq(p,v)dv = \text{weighted average value of time in lane 1}$$

$$\bar{v}_2 = \frac{\int_{v^*}^{\infty} \frac{\partial c}{\partial t(q_2,v)} q(p,v)dv}{\int_{v^*}^{\infty} q(p,v)dv} = \frac{1}{q_2}\int_{v^*}^{\infty} vq(p,v)dv = \text{weighted average value of time in lane 2}$$

Maximise economic welfare = willingness-to-pay for lane 1 + willingness-to-pay for lane 2 – total social costs for lane 1 – total social costs for lane 2

$$w = \int_0^{v^*}\int_0^q p(q,v)dqdv + \int_{v^*}^{\infty}\int_0^q p(q,v)dqdv -$$
$$\int_0^{v^*} c\left(t_1\left[\int_0^{v^*} q(p,v)dv\right], v\right)q(p,v)dv - \int_{v^*}^{\infty} c\left(t_2\left[\int_{v^*}^{\infty} q(p,v)dv\right], v\right)q(p,v)dv$$

For each individual value of v between zero and v*,

$$\frac{\partial w}{\partial q(p,v)} = p(q,v) - c\left(t_1\left[\int_0^{v^*} q(p,v)dv\right], v\right) - \int_0^{v^*} \frac{\partial c}{\partial t(q_1,v)}\frac{dt_1}{dq_1} q(p,v)dv = 0$$

$$p - c = \frac{dt_1}{dq_1}\int_0^{v^*}\frac{\partial c}{\partial t(q_1, v)}q(p,v)dv = \pi_1$$

For each individual value of v between v* and infinity

$$\frac{\partial w}{\partial q(p,v)} = p(q,v) - c\left(t_2\left[\int_{v^*}^{\infty}q(p,v)dv\right], v\right) - \int_{v^*}^{\infty}\frac{\partial c}{\partial t(q_2, v)}\frac{dt_2}{dq_2}q(p,v)dv = 0$$

$$p - c = \frac{dt_2}{dq_2}\int_{v^*}^{\infty}\frac{\partial c}{\partial t(q_2, v)}q(p,v)dv = \pi_2$$

If cost is a linear function of time:

$$\pi_1 = \frac{dt_1}{dq_1}\int_0^{v^*}vq(p,v)dv = \frac{dt_1}{dq_1}\bar{v}_1 q_1 \text{ and } \pi_2 = \frac{dt_2}{dq_2}\int_{v^*}^{\infty}vq(p,v)dv = \frac{dt_2}{dq_2}\bar{v}_2 q_2$$

Differentiating the welfare function with respect to v*:

$$\frac{\partial w}{\partial v^*} = \int_0^q p(q, v^*)dq - \int_0^q p(q, v^*)dq$$

$$-c\left(t_1\left[\int_0^{v^*}q(p,v)dv\right], v^*\right)q(p,v^*) - \int_0^{v^*}\frac{\partial c}{\partial t(q_1, v)}\frac{dt_1}{dq_1}q(p,v^*)q(p,v)dv$$

$$+c\left(t_2\left[\int_{v^*}^{\infty}q(p,v)dv\right], v^*\right)q(p,v^*) + \int_{v^*}^{\infty}\frac{\partial c}{\partial t(q_2, v)}\frac{dt_2}{dq_2}q(p,v^*)q(p,v)dv = 0$$

$$c_1^* + \frac{dt_1}{dq_1}\int_0^{v^*}\frac{\partial c}{\partial t(q_1, v)}q(p,v)dv = c_2^* + \frac{dt_2}{dq_2}\int_{v^*}^{\infty}\frac{\partial c}{\partial t(q_2, v)}q(p,v)dv$$

$$MSC_1^* = MSC_2^*$$

With optimal congestion prices charged for both lanes:

$c_1^* + \pi_1 = c_2^* + \pi_2$ or $c_1^* - c_2^* = \pi_2 - \pi_1$

If cost is a linear function of time: $v^* t_1 + \frac{dt_1}{dq_1}\bar{v}_1 q_1 = v^* t_2 + \frac{dt_2}{dq_2}\bar{v}_2 q_2$

$v^*(t_1 - t_2) = \pi_2 - \pi_1$ $\qquad v^* = \frac{\pi_2 - \pi_1}{t_1 - t_2}$

8.3 Second-best optimal charge for lane 2 with the charge for lane 1 fixed

π_1 = charge in lane 1, assumed to be fixed

π_2 = charge in lane 2, which we are seeking to optimise

G = the gap between marginal social cost and average private cost for a lane.

For any given v < v*:

$$G_1 = MSC_1 - c_1 = \frac{dt_1}{dq_1}\int_0^{v^*}\frac{\partial c}{\partial t(q_1, v)}q(p,v)dv = \frac{dt_1}{dq_1}\int_0^{v^*}vq(p,v)dv = \frac{dt_1}{dq_1}\bar{v}_1 q_1$$

For any given v>v*

$$G_2 = MSC_2 - c_2 = \frac{dt_2}{dq_2}\int_{v^*}^{\infty}\frac{\partial c}{\partial t(q_2, v)}q(p,v)dv = \frac{dt_2}{dq_2}\int_{v^*}^{\infty}vq(p,v)dv = \frac{dt_2}{dq_2}\bar{v}_2 q_2$$

From the derivation in the previous sub-section, for all values of v for lane 1:

$$\frac{\partial w}{\partial q(p,v)} = p(q,v) - c\left(t_1\left[\int_0^{v^*} q(p,v)dv\right], v\right) - \frac{dt_1}{dq_1}\int_0^{v^*}\frac{\partial c}{\partial t(q_1,v)}q(p,v)dv$$

The total price paid by a user of lane 1 having a given value of time is:
$p(q,v) = c\left(t_1\left[\int_0^{v^*} q(p,v)dv\right], v\right) + \pi_1$, which, when substituted into the previous expression gives:
$\frac{\partial w}{\partial q(p,v)} = \pi_1 - \frac{dt_1}{dq_1}\int_0^{v^*}\frac{\partial c}{\partial t(q_1,v)}q(p,v)dv = \pi_1 - G_1 = \frac{\partial w}{\partial q_1}$, since the charge and the gap are the same for all q(p,v) regardless of the value of v. Similarly, for lane 2, $\frac{\partial w}{\partial q_2} = \pi_2 - G_2$.

Consider economic welfare be a function of the quantities of PCUs in the two lanes: $w = w(q_1, q_2)$. There is no need to include v* because, with the charge for lane 1 fixed, for each value of π_2 there a unique pair of lane quantities and a unique value for v*.

For maximum welfare: $dw = \frac{\partial w}{\partial q_1}dq_1 + \frac{\partial w}{\partial q_2}dq_2 = 0$

Dividing through by dv*: $\frac{dw}{dv^*} = \frac{\partial w}{\partial q_1}\frac{dq_1}{dv^*} + \frac{\partial w}{\partial q_2}\frac{dq_2}{dv^*} = 0$

$$\frac{dw}{dv^*} = (\pi_1 - G_1)\frac{dq_1}{dv^*} + (\pi_2 - G_2)\frac{dq_2}{dv^*} = 0 \ ; \qquad \pi_2 = G_2 - (\pi_1 - G_1)\frac{dq_1}{dv^*}\bigg/\frac{dq_2}{dv^*}$$

This expression is easier to explain after rearranging it thus:
$(\pi_1 - G_1)\frac{dq_1}{dv^*} = -(\pi_2 - G_2)\frac{dq_2}{dv^*}$.

With first-best optimal pricing, $\pi_1 = G_1$ and $\pi_2 = G_2$, so both sides of the expression equal zero. In the second-best situation, π_1 is exogenously set at some level — usually zero. $\pi_1 - G_1$ is the welfare gain that results from an additional vehicle joining the traffic stream in lane 1. It is amount by which the price paid (marginal willingness-to-pay) exceeds marginal social cost. With $\pi_1 < G_1$, $\pi_1 - G_1$ is negative indicating that an additional vehicle generates a welfare loss. Likewise, $\pi_2 - G_2$ is the welfare gain (loss, if negative) that results from an additional vehicle joining the traffic stream in lane 2.

The terms $\frac{dq_1}{dv^*}$ and $\frac{dq_2}{dv^*}$ represent the full effects on the quantities of traffic in the respective lanes arising from a one dollar increase in v*. Since traffic shifts from lane 2 to lane 1, $\frac{dq_1}{dv^*} > 0$ and $\frac{dq_2}{dv^*}$<0. Hence $\pi_1 < G_1$ will cause $\pi_2 < G_2$. The second-best optimal charge for lane 2 is set such that if the vehicle with value to time v* changes lanes, the saving in deadweight loss to society in the lane it leaves is the same as the additional deadweight loss imposed in the lane it enters. Hence economic welfare cannot be improved by altering the traffic split. The deadweight losses are calculated taking account of all the secondary impacts caused by the changes in the charge itself and in the speeds for all other traffic.

The nature of these secondary impacts is explored further in the remainder of this sub-section in the course of deriving formulas for $\frac{dq_1}{dv^*}$ and $\frac{dq_2}{dv^*}$. To simplify the analysis, it is assumed that road user costs are a linear function of time taken.

A one dollar change in v* causes q* PCUs to switch lanes. Shifting of q* PCUs into lane 1 has **two** effects. First, q_1 is increased by q*. Second, with additional vehicles in the lane, VCR rises, causing time and costs to rise. The equilibrium points move up all the demand curves for traffic in lane 1 leading to a reduction in traffic. In other words, the additional q* PCUs crowd out some of the existing vehicles. The net increase in q_1 is therefore somewhat less than q*.

In the basic micro-economic model of a downward-sloping demand curve intersecting an upward-sloping supply curve to provide an equilibrium, a leftward shift of the supply curve of q* units, would result in a change in the equilibrium quantity of approximately $q^*\left(1-\frac{d'}{s'}\right)^{-1}$ where $d'<0$ is the slope of the demand curve and and $s'>0$, the slope of the supply curve. The change in the equilibrium quantity has the same sign as q*, but is of a smaller size because of the increase in price.

Combining the two effects:

$$\frac{dq_1}{dv^*}=q^*-q^*\left(1-\frac{1}{t_1'R_1}\right)^{-1}=\frac{q^*}{1-t_1'R_1} \text{ where } t_1'=\frac{dt_1}{dq_1} \text{ and } R_1=\int_0^{v^*} v\frac{dq(p,v)}{dp}dv.$$

A shift of q* PCUs out of lane 2 as a result of a one dollar increase in v* has **three** effects on the traffic in lane 2. First, there is the loss of q* PCUs. Second, the reduction in vehicles in the lane causes VCR, time and costs to fall, inducing some extra traffic. These two effects are the mirror images of the two effects in lane 1 and are estimated in the same way. A negative sign must proceed q* and the traffic-inducement effect is added.

The third effect is the impact of the increase in π_2 raising prices, causing movement up the demand curves for the traffic remaining in lane 2. This third effect works against the traffic inducement effect. The perspective adopted here is that v* undergoes a one dollar increase. So we need to know the increase in π_2 associated with a one dollar increase in v^*. Since $\pi_2-\pi_1=v^*(t_1-t_2)$, and π_1 is fixed, $d\pi_2=(t_1-t_2)dv^*+(dt_1-dt_2)v^*$, from which:

$$\frac{d\pi_2}{dv^*}=(t_1-t_2)+\left(\frac{dt_1}{dq_1}\frac{dq_1}{dv^*}-\frac{dt_2}{dq_2}\frac{dq_2}{dv^*}\right)v^*.$$

In the basic demand–supply model, an upward movement in the supply curve caused by increasing a tax π_2 by $\Delta\pi_2$ dollars causes the equilibrium quantity to change by:

$$\Delta q_2=-\Delta\pi_2(s'-d')^{-1}=\Delta\pi_2\frac{1}{d'}\left(1-\frac{s'}{d'}\right)^{-1}$$

Combining the three effects:

$$\frac{dq_2}{dv^*}=-q^*+q^*\left(1-\frac{1}{t_2'R_2}\right)^{-1}+D_2\left(1-\frac{1}{t_2'R_2}\right)^{-1}\left[t_1-t_2+\left(t_1'\frac{dq_1}{dv^*}-t_2'\frac{dq_2}{dv^*}\right)v^*\right]$$

where, $t_2'=\frac{dt_2}{dq_2}$, $D_2=\int_{v^*}^{\infty}\frac{dq(p,v)}{dp}dv$ and $R_2=\int_{v^*}^{\infty}v\frac{dq(p,v)}{dp}dv$.

We already have an expression for $\frac{dq_1}{dv*}$. Rearranging to combine the two occurrences of $\frac{dq_2}{dv*}$ and simplifying:

$$\frac{dq_2}{dv*} = \frac{D_2\left[t_1 - t_2 + t_1' \frac{dq_1}{dv*} v*\right] - q*}{1 - t_2'\left(R_2 - D_2 v*\right)}$$

8.4 Third-best optimal charge for a HOT lane

The subscript 3 is used to refer to the select group of vehicles allowed to travel in lane 2 for free. Hence, π_3, the charge levied on the select group, is zero. It is assumed that traffic in lane 2 travels faster than in lane 1 so that the entire select group of vehicles will chose to travel in lane 2. The PCUs in the select untolled group of vehicles has to be removed from the function q(p,v). It is described by a separate function $q_3(p,v)$. The subscript 2 on the supply side refers to the cost curve for lane 2, and on the demand side, for tolled traffic using the fast lane.

The gaps (G's) between marginal social and average private costs for lane 2 need to be redefined because an increase in travel time caused by an additional PCU (t_2') in lane 2 increases costs for all vehicles in the lane, both tolled and untolled.

$$G_2 = G_3 = t_2'\left[\int_{v*}^{\infty} v q_2(p,v)dv + \int_0^{\infty} v q_3(p,v)dv\right]$$

The definite integral for q_3 is from **zero** to infinity because **all** vehicles in the select group of untolled vehicles travel in lane 2 regardless of value of time.

$$\frac{dw}{dv*} = \left(\pi_1 - G_1\right)\frac{dq_1}{dv*} + \left(\pi_2 - G_2\right)\frac{dq_2}{dv*} - G_3\frac{dq_3}{dv*} = 0$$

$$\pi_2 = G_2 - \left[\left(\pi_1 - G_1\right)\frac{dq_1}{dv*} - G_3\frac{dq_3}{dv*}\right] \Big/ \frac{dq_2}{dv*}$$

The formula for $\frac{dq_1}{dv*}$ is unchanged from that derived the previous sub-section.

The three effects of a change in v* on q_2 are the same as for the previous sub-section, except that the demand system is more complex. The formula here was derived by solving a system of linear equations and differentiating to discover the effects on q_2 and q_3 of a rightward shift in the $t(q_2)$ curve and an increase in π_2.

$$\frac{dq_2}{dv*} = \frac{D_2\left[t_1 - t_2 + t_1'\frac{dq_1}{dv*}v*\right] - q*}{\frac{1 - t_2'\left(R_2 + R_3\right)}{1 - t_2' R_3} + t_2' D_2 v*} \quad \text{where } R_3 = \int_0^{\infty} v \frac{dq_3(p,v)}{dp} dv.$$

The effects on q_3 of the shift of vehicles from lane 2 to lane 1 and the altered charge for tolled traffic in lane 2 associated with a change in v* both occur via the effect on travelling time in lane 2. Hence the effect on q_3 can be estimated simply as:

$$\frac{dq_3}{dv*} = \frac{dq_2}{dv*}\left(1 - \frac{1}{t_2' R_3}\right)^{-1}$$

Note that $\frac{dq_2}{dv*}$ has been multiplied by –1 here because $\frac{dq_2}{dv*}$ is negative but, from the point of view of q_3, there is an increase in available road capacity.

8.5 Verification of formulas

All the formulas for economic welfare-maximising charges developed in this mathematical appendix were verified using the spreadsheet model, treating the definite integrals as summations over the values of time. Values of v^* were varied manually to confirm that the optimal charges predicted by the formulas were indeed optimal, and that the formulas for derivatives of quantities with respect to v^* gave correct results.

9 References

Austroads (2004) *Guide to Project Evaluation — Part 4: Project Evaluation Data*, AP-G82-04, Austroads, Sydney, http://www.austroads.com.au.

BTRE (2002) *Greenhouse Policy Options for Transport*, Report 105, BTRE, Canberra.

Dahlgren, J. (2002) High-Occupancy/Toll lanes: Where should they be implemented, *Transportation Research Part A*, 36, pp. 239-255.

California Department of Transportation (Caltrans) Traffic operations program, Office of truck services (2004) *Truck-only Lanes*, Fact sheet, http://www.dot.ca.gov/hq/traffops/trucks/trucksize/fs-trucklanes.htm

Fielding, G.J. and Klein, D.B. (1993) High Occupancy Toll Lanes: Phasing in Congestion Pricing, *Reason Public Policy Institute*, http://www.rppi.org/transportation/ps170.html

Parry, I.W.H. (2002) Comparing the Efficiency of Alternative Policies for Reducing Traffic Congestion, *Journal of Public Economics*, 85, pp. 333–362.

Poole Jr., R.W. and Orski, C.K. (1999) Building a Case for HOT Lanes: A New Approach to Reducing Urban Highway Congestion, *Reason Public Policy Institute*, http://www.rppi.org/257.html

Poole Jr., R.W. and Orski, C.K. (2000) HOT Lanes: A Better Way to Attack Urban Highway Congestion, *Regulation*, 23(1), pp. 15–20.

Small, K.A. and Yan, J. (2001) The Value of 'Value Pricing' for Roads: Second-best Pricing and Product Differentiation, *Journal of Urban Economics*, 49(2), pp. 310–336..

Verhoef, E.T. and Small, K.A (1999) Product Differentiation on Roads: Second-Best Congestion Pricing with Heterogeneity under Public and Private Ownership, *Tinbergen Institute Discussion Papers* number 99-066/3, http://www.tinbergen.nl/discussionpapers/99066.pdf

Yang, H and Huang, H (1999) Carpooling and Congestion Pricing in a Multilane Highway with High-Occupancy Vehicle Lanes, *Transportation Research Part A*, 33, pp. 139-155.