***Victoria Transport Policy Institute***
1250 Rudlin Street, Victoria, BC, V8V 3R7, CANADA
www.vtpi.org    info@vtpi.org
Phone & Fax 250-360-1560
*"Efficiency - Equity - Clarity"*

# Reassessing "The Social Desirability of Urban Rail Transit Systems"

*Critique of Winston and Maheshri*

18 May 2007

By Jay Warner
Principal Scientist
Warner Consulting, Racine, WI, USA


## Abstract

A recent article, "The Social Desirability of Urban Rail Transit Systems," (Winston and Maheshri, 2006) presents an econometric analysis of urban rail transit. It concluded that, except for San Francisco's BART, all U.S. urban rail systems have a negative social desirability. This paper evaluates that article.

## The Article's Objective

The intent of Winston and Maheshri's article is to estimate social welfare of urban transit based on demand for and cost of service. "Social welfare" is defined as benefits of urban transit less the cost of operating the trains, denominated in constant dollars. If the authors can build a reasonably accurate mathematical model, then decision makers could design systems with maximum social benefit.

The sample is from 1993 to 2000, and covers 25 urban transit systems active in that period, totaling 194 measurements. They use network complexity measures adapted from mapping theory as possible 'independent variable' parameters to account for ridership levels. The technical approach is fundamentally multiple regression analysis for both benefits and costs, which is quite suitable for this study. That is, they specify a number of possible independent variables, and then statistically analyze the data to determine which variables account for the "dependent variable," the ridership levels (a measure of benefits) and costs. When performed with care for the data structure (what statisticians call "the properties of the dataset,") this approach can lead to fascinating – even counter intuitive – conclusions. The book, *Freakonomics*, by Steven Levitt and Stephen Dubner, offers several successful analyses.

The paper specifically does not include commuter rail systems, which the authors define as those that bring people from suburbs to a major city. This paper does not include METRA in its calculations for Chicago, only CTA (Chicago Transit Authority). It may not include train travel from White Plains, or Westchester, into New York. Train travel from New Haven, Bridgeport or Stamford, all in CT, would be excluded by the definition of "commuter transit" used in this paper. Newark and New Brunswick, NJ could be included in the "Northern New Jersey" system, or one could be excluded as a "suburb."

## Technical Analysis Critique

The first analytic difficulty is with the structure of the data. This is a "natural" type of experiment, inasmuch as the data are selected from the population of running urban transit systems. Natural experiments are common in social science research. Analysis of natural experiments must be careful to assure that independent variables outside the study, or completely unmeasured variables, do not influence the results in strange ways.

To solve multiple regression equations there is a matrix of test conditions that is multiplied by its transpose, inverted, then multiplied by the transposed matrix to form a "solution" matrix. The data is well formed when the solution matrix is near an identity matrix (see Draper and St. John, 1975). Because this data is inherently not well formed, we cannot draw firm conclusions; the statistical calculations do not accommodate sizable deviations from their mathematical assumptions.

The authors cannot be criticized for poorly formed data - the data is what it is. The question here is how well the weak data was accommodated in the analysis.

### Auto Correlation of Data

The data for each system each year is quite similar to the year before; in other words, the data are auto correlated. We really don't have 194 independent data points, but 25. The mathematical analysis <u>assumes</u> that all the measurements are independent. Winston and Maheshri compensate for this assumption by using a Newey-West adjustment, the effect of which is to make the statistical error terms larger and more realistic, but we still have only 25 independent points.

### Outlier - New York City (NYC)

The authors point out that NYC has about two-thirds of all the riders. This could significantly skew the results - we are often comparing NYC with other cities.

The authors note that cities with larger populations tend to have more riders. A plot of number of riders vs. population would show that NYC was larger than any study unit on the population scale, and much larger on the rider mileage scale. Including NYC in this regression forces a result (number of riders depends on city population) that is only valid if all the other characteristics involving NYC are roughly equal to those of the other cities, or if the solution matrix has no large off-diagonal elements.

The authors observe that some NYC characteristics are different than other systems, and they perform some analyses on defined large systems to isolate the difference in coefficients. This further reduces the amount of data available to determine those coefficients. The reader has no way of evaluating the seriousness of outlier effects, and the authors do not discuss the issue.

## Misleading Reliance On $R^2$ - Goodness of Fit

One thing about regression analysis is that there are *a lot* of pitfalls available to the analyst. For example, with n observations and n-1 explanatory, independent variables, then *all* the variation in the dependent variable is explained by the model and $R^2$ is effectively 1 -- perfect. Unfortunately, this fully explained model does not predict the next time period, certainly not with absolute precision. The authors have 15 variables in the demand model, with at most 25 independent measurements. Calculations on selected "large" systems have eight. A year to year comparison of results in a stable system could estimate the statistical measurement error. This approach would be more realistic than claiming irrationally high $R^2$ values.

## Over Fitting of Data

When $R^2$ gets much closer to 1 than the original data precision would warrant, we have an 'over fit' situation. The model has a very small residual error, but it does not predict demand or costs for any time outside the study period. By finding correlations with many variables, it does not describe what actually is going on in the system of urban transit. The model cannot be used for decision-making. There are ways to avoid over fitting of data to make the model predictive within the precision of the source data. This paper apparently does not apply them.

## Why 99% Confidence Use?

The model presumes that average fares will not depend on the ridership (i. e., they are "exogenous," or independent variables). This was evaluated using a Hausman test, and could not reject the assumption of exogenous "at a high level of confidence," footnoted as 99% confidence. This means that the authors will proceed as if the data was exogenous.

The level of confidence is a measure of the alpha risk - the chance of saying there is an effect, when there is none. As the level of confidence selected for the test increases toward 100% (absolute certainty), the alpha risk drops toward 0. But concurrently the beta risk increases - the chance of concluding that there is no effect when an effect actually exists. The statistical test makes a trade-off between falsely proclaiming an effect, and falsely denying it. For most analyses where the consequences are not severe the confidence level is taken as 95%. In many social science studies a confidence level of 90% is used. If there is an effect in the data, we want to see and evaluate it. The analysts are wiling to risk that an effect is not there, in exchange for more frequently detecting less obvious effects.

As best I can assess from the information given in a footnote, the authors would reject the hypothesis of independence at the 92% confidence level. Considering the data involved, that's a pretty good indication that mathematically, average fares are driven by ridership, and thus, that the model includes internal complications. It could simply be that the random variation in sparse data leads to this relationship, but that alternative would raise other questions involving the validity of so many variables in the model.

## Discussion

The Victoria Transport Policy Institute (www.vtpi.org), has evaluated various papers claiming to demonstrate the economic failings of rail transit services. The table below summarizes analysis of impacts that are considered in conventional economic analysis. Conventional transport project evaluation tends to focus on certain direct impacts, such as traffic congestion and vehicle operating costs, but ignores other significant impacts, such as vehicle ownership and parking costs, and indirect impacts, such as induced travel effects (the additional surface street traffic congestion, accidents and pollution emissions caused by expanded highway capacity), and increased sprawl.

***Table 1*** **Impacts Considered and Overlooked** (Litman, 2005)

| Usually Considered | Often Overlooked |
|---|---|
| • Financial costs to governments<br>• Vehicle operating costs (fuel, tolls, tire wear)<br>• Travel time (reduced congestion)<br>• Per-mile crash risk<br>• Project construction environmental impacts | • Downstream congestion impacts<br>• Parking costs<br>• Vehicle ownership costs (depreciation, insurance, etc.)<br>• Impacts on non-motorized travel<br>• Project construction traffic delays<br>• Impacts of generated traffic<br>• Indirect environmental impacts<br>• Strategic land use impacts<br>• Equity impacts<br>• Per-capita crash risk<br>• Impacts on physical activity and public health |

*This table shows which impacts are usually considered and overlooked in conventional transportation economic evaluation. Because critics generally overlook significant impacts hey tend to be undervalue transit improvements, particularly rail transit which attracts discretionary travelers and has significant land use impacts.*

### Selective train type exclusions

By collecting train transportation - heavy and light rail but not commuter rail lines, the authors hope to come up with measures that apply across all systems. I believe they do not include bus transportation in the mix.  Trains are part of a **multi-modal** mix of transportation methods. An analysis that excludes some significant mode will necessarily be incomplete, and may well miss the real explanations (and variables) for why people choose one mode of travel over another.

This exclusion of a major part of transit is roughly the same as looking at the social value of roadways in Wisconsin, without including the major interstate highways. Whatever the result, it would have no value to those planning roads and highways or those wrestling with paying for them.

## Slanted argumentative statements

Winston and Maheshri's analysis is biased to favor automobile transportation over rail. They ignore roadway, parking facility and vehicle costs. Rail pollution reduction benefits are dismissed, although seemingly every possible source of pollution from train use is mentioned. Traffic accidents receive a cursory consideration, and mobility for non-drivers is ignored. The authors may claim that the accident costs are "internalized" by vehicle insurance, so any safety gains from trains are inconsequential, but many studies show that a significant portion of traffic risk is external (Vickrey, 1968; Edlin and Mandic, 2001). These dismissive entries detract from the rigor of the paper. When the authors include virtually superfluous comments, all disparaging train usage, I infer that they are not interested in a rigorous analysis, but want an excuse to slam rail transit systems.

I am a fan of mathematical models (see for example, Santell, Jung, and Warner, 1992). But every mathematical model must meet certain criteria to be useful. It must incorporate the major variables influencing the output, and it must predict rational outcomes for new conditions, or fit the known facts well. If the authors of this paper had concluded that their new model (involving map and topology theory) was an interesting idea, but didn't predict well, I would have no argument with the report.

The authors' model is over fitted, and cannot predict ridership in following years. The cost model neglects major transportation costs. The author's admit as much when they terminated sampling in 2000 on the grounds that the following years were "tumultuous." The next year included 9/11, which severely distorted NYC's ridership figures for reasons unrelated to model parameters. Following years saw relatively dramatic gasoline price volatility. If the cost of auto transportation was accurately included in the model, I am not sure why 2002 through 2005 were not included.

Ridership in the eight newest U.S. rail transit systems all exceeded their long term ridership projections. Six beat the 2010 or 2020 projections in 2006. Although this does not directly contradict Winston and Maheshri's analysis, because the model predicts passenger miles, not ridership, and some systems include commuter rail that was excluded from their study, it nonetheless, makes the model suspect as a predictor.

On Earth Day, 2007, Mayor Bloomberg of NYC announced a plan to reduce automobile traffic into the city with congestion pricing. A major justification is that there is not room in the city for additional parking, nor street space to move the cars around. This disconnect, between real-world transportation problem solving and Winston and Maheshri's recommendation to shut down urban transit illustrates how omitting critical factors can result in technical analysis that provides meaningless results. Their econometric model does not predict a rational outcome. If Winston and Maheshri's model does not apply outside 1993 - 2000, then its predictive value is nil. If the model does apply (2001 excepted) into the 21st century, its predictions are erroneous.

When a model utterly fails to fit the situation, objective analysts can conclude that critical parameters were inadvertently left out, that the input data failed to include significant aspects, or that data failed one or more of the data structure assumptions required for the mathematical analyses. The authors of this paper do none of these; they persist in believing the model. They do not perform the most fundamental and important test of an analytic model; they do not ask, "is this model reasonable?"

## Conclusions:

- The analysis excludes commuter rail -- transit from suburbs to metropolitan areas. This would be a major oversight for evaluating inherently multi-modal transit networks.
- The cost of destination parking for alternative transport is not considered. The authors may underestimate highway congestion costs for alternative transport. Destination street congestion costs are not considered. Some train benefits are neglected, or willfully dismissed.
- The authors do not include alternative costs of transportation for non-drivers.
- The authors' treatment of pollution reduction benefits is intellectually questionable at best, and deceptive at worst.
- Due to auto correlation, statistically independent data is restricted. New York City data frequently is an outlier, distorting possible interpretations.
- Overfitting of data gives a severely distorted picture of model precision, and raises questions of predictive validity.
- Tests of exogenicity may be misinterpreted, hiding significant relationships of the data.
- Inherent characteristics of the data make the mathematical model suspect, limiting the significance of any conclusions.
- The model predicts that the most socially beneficial action would be to shut down all urban train systems, a result that in New York City has recently been flatly contradicted. Ridership from 2002 to 2006 on the eight newest systems has also indicated the invalidity of this conclusion.
- The analytic model either does not apply beyond 2000, or does not predict accurately. Therefore, its usefulness to planners and payers is nil.

## References

Nathaniel Baum-Snow and Matthew E. Kahn (2005), *The Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970 to 2000*, Brookings Papers on Urban Affairs (www.econ.brown.edu/fac/Nathaniel_Baum-Snow/brook_final.pdf).

Antonio M. Bento, Maureen L. Cropper, Ahmed Mushfiq Mobarak and Katja Vinha (2003), *The Impact of Urban Spatial Structure on Travel Demand in the United States*, World Bank Group Working Paper 2007, World Bank (http://econ.worldbank.org/files/24989_wps3007.pdf).

Norm R. Draper and R.C. St. John (1975), "D-Optimality for regression designs: a review," *Technometrics*, V 17(1), pg. 15-23.

Aaron S. Edlin and Pinar Karaca Mandic (2001), *The Accident Externality from Driving*, University of California, Berkeley (http://works.bepress.com/aaron_edlin/21).

Todd Litman (2004a), *Rail Transit In America: Comprehensive Evaluation of Benefits*, VTPI (www.vtpi.org).

Todd Litman (2004b), *Evaluating Public Transit Benefits in St. Louis: Critique of "Light Rail Boon or Boondoggle*," Citizens for Modern Transit (www.cmt-stl.org); available at www.cmt-stl.org/images/litman.pdf.

Todd Litman (2005) *Evaluating Public Transit Benefits and Costs*, VTPI (www.vtpi.org).

Todd Litman (2006), *Responses To "A Desire Named Streetcar,"* Victoria Transport Policy Institute (www.vtpi.org); at www.cfte.org/litmanotoole.pdf.

Todd Litman (2007), *Evaluating Rail Transit Criticism*, VTPI (www.vtpi.org); at www.vtpi.org/railcrit.pdf.

M.P. Santell, J.R. Jung and J.C. Warner (1992), "Optimization in Project Coordination Scheduling Through Application of Taguchi Methods," *Project Management Journal*, XXIII(3), pg. 5-16, Sept. 1992

William Vickrey (1968), "Automobile Accidents, Tort Law, Externalities, and Insurance: An Economist's Critique," *Law and Contemporary Problems*, Vol. 33, pp. 464-487; Available at www.vtpi.org/vic_acc.pdf.

Clifford Winston and Vikram Maheshri (2006), "On the social desirability of urban rail transit systems," *Journal of Urban Economics* (doi:10.1016/j.jue.2006.07.002)