Predicting road travel time

PAPER FOR THE INFRADAY CONFERENCE, BERLIN 2011

Dan Elliott
Frontier Economics, 71 High Holborn, London WC1V 6DA
Tel. +44 (0) 20 7031 7000
Fax. +44 (0) 20 7031 7001
dan.elliott@frontier-economics.com

Luis Campos
Frontier Economics, 71 High Holborn, London WC1V 6DA
Tel. +44 (0) 20 7031 7000
Fax. +44 (0) 20 7031 7001
luis.campos@frontier-economics.com

Ron Smith
Economics, Birkbeck College, Malet St. London WC1E 7HX
Tel +44 (0) 20 7631 6413
Tel +44 (0) 20 7631 6416
r.smith@bbk.ac.uk

Abstract

This paper summarises the development of a model created for the England’s Highways Agency, which uses time series and cross sectional econometric techniques to predict average journey times on any given section of the England’s major road network. The purpose of the model is to simulate the impact of changes in road conditions on road delays across complex road network systems and over extended periods.

The empirical analysis is based on separate time-series equations for each road section, where average travel time is regressed against several features describing road conditions on that section, including lagged journey time on that section and on adjacent sections. The equations for the separate sections are combined to form an interacting system, where shocks are transmitted between sections. To estimate the relevant parameters we used longitudinal data across more than 2,500 road sections in England over a two year period, with data recorded at 15 minute intervals. Hence the modelling process made use of 190m records.
Although primarily descriptive, two specific features make this analysis particularly suitable for simulation. First, it captures important non-linear features of traffic flow dynamics as, for example, the persistent impact of an accident on road delays after the accident has been cleared and the propagation of delays to different sections of the road network.

Second, it captures the impact of (unobserved) heterogeneity in road conditions across the network on delays by allowing for road-section-specific estimates for each parameter. It uses time-series data to obtain a first set of estimates for each road section. It then improves these estimates using cross-sectional information on the characteristics of the road sections.

The prediction of average travel times on road networks under different conditions can have multiple applications. The Highways Agency has used this model to understand better (1) the main drivers of road delays across its network; (2) the impact of past investments on current road delays; and (3) the likely impact of future investments on road delays.

No one likes traffic jams. They bring uncertainty into our world, make us arrive late to important meetings and are a major source of stress, especially with children in the back seat. Moreover, they increase fuel consumption and harm the environment. In the long term, a badly functioning road network can also be an obstacle to regional development.

These are some of the arguments used to justify road network expansion. While increasing road capacity is probably the most effective way to reduce road delays, it might not always be the most efficient use of public money. Many roads are congested only at peak times and are quite empty much of the rest of the time. Moreover, in many cases there are viable alternatives to building more roads which may provide better ways to reduce delays. For example, by avoiding road works during congested periods, reducing the number of accidents or responding more quickly to incidents on the carriageway.

The best way to tackle delays will very much depend on the nature of the disruption, on the traffic conditions at the time of the disruption, on the features of the road where the disruption occurs and how that road is integrated into the wider network. It follows that to tackle road delays effectively and efficiently, one needs to understand how alternative policies might affect delays on different roads subject to changing traffic conditions.

The Highways Agency (HA) wanted to measure better the relative impact of different disruptions on average delays across its network in England under these multiple scenarios. These measurements would allow it to better:

- assess the impact of past road network investments and road management strategies on average delays across its network; and
- help prioritise future investments based on their likely impacts on average delays across its network.
It concluded that the existing road traffic models could not be used for this purpose and a novel more empirical approach was needed.

**Why an empirical approach is needed**

Most road traffic models measure travel time on a “representative” road under a set of “specific” traffic flow conditions. They tell us what delays to expect under a set of assumptions, but do not explain the delays that actually occur. As **Figure 1** shows, the average travel time on a particular section can vary significantly, even over a relatively short period.¹

**Figure 1.** The average travel time on a road section can change significantly over a week

The HA network is extremely varied and furthermore, road conditions are constantly changing across the network.

These traffic models failed to capture the diverse set of factors affecting traffic conditions on observed road delays on real roads, which is precisely what the HA

---

¹ Examples of the types of models considered by the Highways Agency to measure the effects of changes in road conditions on travel time and economic cost can be seen in the Cost Benefit Analysis (COBA) and Queues and Delays at Roadworks (QUADRO) manuals from the UK Department from Transport. More references are provided at the end of the paper.
was interested in. To predict travel time accurately the HA could not rely on the impact of changes in traffic conditions on a “representative” road section. Instead, it had to take into account the actual traffic behaviour on each particular road section, and the specific road configuration and conditions.

This suggested the need to follow a more empirical approach. In the past, empirical analysis of England road travel times relied on simple comparison between the road delays during a specific period, when an incident or event has occurred (e.g. road works), and delays before (or after) the incident or event.

Such an approach does not permit the measurement of the incremental effect of an incident or event with confidence, because it implicitly assumes that nothing else is affecting the travel time in the two periods compared. It also makes it difficult to compare the impact of reducing different type of incidents, for instance reducing the impact of road works against reducing the number of accidents. So, the HA concluded that, to inform policy they need a new empirical model which would take into account the various factors affecting traffic conditions at any point in time and on any particular road section.

This should be a comprehensive tailor-made model that reflected the particular features of the HA road network and the actual traffic flow behaviour on each part of that network. To do this, we would need to make full use of the information on actual traffic performance over time collected by the HA across its network over time.

**Challenges of the time series approach**

The HA wanted a model that could predict travel times across more than 2,500 interconnected road sections in England. The main challenge was to make this model both comprehensive and tractable given:

- **the large number of factors affecting travel times**, which can be split between factors affecting the conditions on a specific road section such as road demand (or for example, rain), the specific features of the road section (for example, inclination, curvature) and the way each road section is connected to the wider network (for example, having one or two upstream road sections);

- **how these factors vary across the network and over time**;

- **how these factors interact** in complex ways, making it difficult to measure the individual contribution of each of these on road delays;

- **how road delays are likely to persist over time**;

- **how road delays propagate over the network**; and
the difficulties in measuring the relevant impacts given that the information available on traffic conditions, incidents and road characteristics was extensive but incomplete and dispersed across different datasets.

These issues raised both a conceptual and a practical challenge.

- The conceptual challenge was to find out whether we could capture the complex dynamics of traffic flow and the non-linear impact of changes in road conditions on delays with a relatively simple predictive functional formula that worked across most, if not all, road sections, or if a more complex predictive system was needed.

- The practical challenge was to know whether we would have sufficient amounts of data across the network to meaningfully capture the impact of the different possible changes to road conditions on travel time.

The conceptual challenge

The success of the task very much depended on the ability to find a simple predictive formula that could be calibrated for each of the more than 2,500 road sections in the HA network.

The inherent complexity of road traffic dynamics across road sections would suggest the use of a structural approach, whereby each component affecting road delays is explicitly modelled, forming a complex predictive system.

While such a model might predict travel times on a few road sections better than a simpler model, it would become far too complicated to be of any practical use when applied to the wider network of more than 2,500 road sections. Moreover, the impact of (non-observed) heterogeneity in road conditions on delays across the road network might have required us to use different predictive models for different road sections or type of road sections.\(^2\)

We therefore preferred using a simpler descriptive model. However, using such formula to capture the complexity of road dynamics also created its challenges, as exemplified in Figure 2 and Figure 3 below.

Figure 2 shows that an increase in traffic flow is correlated both with an increase and with a decrease in average road speed, depending on the levels of road congestion.

\(^2\) In addition, there is some evidence suggesting that more parsimonious models, with fewer estimated parameters, forecast better. See for example, Clements and Hendry (1998).
Figure 2. The relation between flow and speed (delays) is complex as shown by the observed Flow-Speed curve on a road section between January 2007 and January 2009.

Source: Frontier Economics

When a road is uncongested, increases in the number of vehicles using the road (the flow) have very little impact on average vehicle speeds. Average speeds fall slightly as the flow increases. But as the road reaches maximum capacity, average speeds fall dramatically. If still more traffic tries to use the road, traffic grinds to a halt, which means that average speeds fall and flow reduces, because the traffic is now queuing.

Figure 3 shows that the impact of an accident depends on the road conditions at the time of the accident, for example if the road is already congested before the accident occurred or if the traffic is flowing freely, as illustrated below.
These type of complex non-linear relationships complicate the task of modelling traffic flow dynamics using a simple functional form, especially since road conditions and road features differ significantly across the network and also vary significantly over time.

We therefore had to devise different modelling strategies to handle some of these more complex features of road traffic dynamics. In our first example, instead of traffic flow, the number of cars passing through a link every 15 minutes, we used the alternative variable density, the average number of cars per km on a given road link during a particular 15-minute period.\(^3\)

In our second example, we took advantage of having road traffic information over time to account for the different magnitudes of the impact of incidents or events under different initial road conditions. We have made the magnitude of these impacts dependent on the road conditions at the time, or just before, the accident occurred, as explained in the box below in more detail.

---

\(^3\) The relation between density and speed (or travel time) is much simpler – the higher the density the lower the road speed (delays increase).
How do the dynamics of our model work?

The formula considered by the NDM to predict travel time $J_{it}$ has the following structure

$$J_{it} = a_{i} + b_{1i} P_{it-n} + b_{2i} N_{it-1} + b_{3i} F_{it} + (b_{4i} W_{it} + b_{5i} I_{it} + b_{6i} R_{it}) * J_{it-1},$$

where:

- $P_{it-n}$ includes the travel time in the previous quarter of an hour and in the previous half an hour and daily, weekly and seasonal travel time patterns;

- $N_{it-1}$ includes lagged travel time on the adjacent road links (upstream and downstream);

- $F_{it}$ (density) and $W_{it}$ (weather conditions and road visibility), $I_{it}$ (incidents, such as accidents) and $R_{it}$ (interventions, such as road works and lane closures) are multiplied by $J_{it-1}$, which reflects road conditions 15 minutes before the incident or event.

The dynamic features of the model imply that the impact of a disturbance on a road section on travel time depends on the initial conditions on the road (by including $J_{it-1}$ in RHS of the formula), persists over time ($P_{it-n}$) and propagates over the network ($N_{it-1}$).

The extent of these effects will depend on the specific parameters associated with the simulated road links.
The practical challenge

To make our model accurately predict observed travel time we needed to calibrate the parameters in our descriptive formula to reflect the differences in road conditions across the road network and over time.

We knew that the quality of our analysis would, in the end, be only as good as the data we used. Fortunately, the HA collects an impressive amount of data on traffic conditions across its network, which we used to calibrate our model. All we needed to do was to compile this data, most of which was fragmented across different datasets in different HA departments.

After collating and cleaning all the relevant data we obtained an impressive dataset of more than 190 million records. These records provided relevant road traffic information across more than the 2,500 road sections under the responsibility of the HA over a two year period. Each record had information on traffic flow, incidents and road characteristics for a given road section in a particular 15 minutes period.

Despite the richness of this dataset, we found gaps in the data for a significant number of road sections. This raised an additional challenge as for the “data poor” road sections we were unable to directly estimate the average impact of changes to road conditions on travel time, or at least not in a robust way.

We had to be creative in finding indirect ways of estimating the parameters for these sections, or making these parameters more robust. The way we solved this challenge was by using the information we had about the “data rich” road sections to infer what the likely model parameters would be for the “data poor” road sections.

This was possible because, although we had inadequate traffic data on the “data poor” road sections, we did know a great deal about their layout, location and average conditions. We used this information to adjust the parameter estimates of the “data rich” road sections to reflect the conditions prevailing in “data poor” road sections. These adjustments reflected differences between these two subsets in road characteristics, such as curvature, and in typical road conditions, such average precipitation.

One advantage of using this approach was that it also enabled us to improve the estimates on “data rich” road sections. This is because we were able to compare the estimates we had obtained on these road sections with the estimates we obtained in other road sections with similar characteristics. From these comparisons we were able to narrow down the level of uncertainty surrounding the road section specific estimates and also potentially correct biases affecting those estimates on some of the road sections.
How we calibrated the model in three (simple) steps.

Our statistical model estimates the parameters in three steps, using in the first two steps simple ordinary least squares estimation procedures.

First stage estimation

In this first stage we estimate the parameters using only the time series dimension of our data. In other words, the parameter estimates for a given road section use only the data available for that road section over the two year period.

We obtained robust parameter estimates for some road sections while for other sections we obtained statistically insignificant parameter estimates or, in some cases, no estimates at all due to the lack of available data.

Second stage estimation

In the second stage we estimate the parameters using the cross-section dimension of our data. In other words, our parameters are estimated using the information on average road characteristics across the network.

We start by estimating the importance of road characteristics. For example, we estimate the average impact that road curvature has on accident led delays. We then multiply these average impacts by the road characteristics of each road section to obtain the estimated parameters for those road sections. For example, we multiply the average impact of road curvature on accident led delays across the network and multiply it by the level of road curvature of a particular road section to obtain the parameter estimate of accidents in our model.

“Shrinkage” of first stage and second stage estimates

We are left with two concurrent set of estimates for the same model parameters. In the third stage we combine the two sets of estimates in order to calculate a third set, which is a weighted average of the first two. The weights will reflect the relative robustness of each of the first two sets of estimates used in the average.

---

4 This will happen for most road sections. However, for those road sections for which it was not possible to obtain a first stage estimate we would have only one set of parameter estimates.
In other words, the more robust the first stage estimates the more weight they will have in this average. This third set of estimates is more robust than the previous two, because it combines road section level and network level information.

We provide additional details in our technical annexe at the end to this paper.

Challenges of predicting travel times

Ensuring the stability of the model by avoiding “error contagion”

On the back of the calibrated statistical model, we constructed a simulation model. This simulation model is currently used by the HA to predict the impact of changes to the road conditions (for example, reduction in the number of accidents) and road layout (for example, the addition of a extra lane to a road) on delays across its network.

Using a statistical model to make prediction of travel time is not straightforward, especially for models with strong dynamic features. A statistical model is, by its very nature inaccurate and inaccurate parameters, when combined with strong dynamic features, can result in very unstable predictions. Minor errors in travel time predictions on one road section at a particular time can spread over the network and persist over time, leading to large prediction errors in subsequent periods.

The parameters of our statistical model are estimated within a confidence interval, which will be narrower in “data rich” road sections and larger in “data poor” road sections. This implies that, for example, our estimate of the impact of changes in traffic conditions upstream on travel time downstream will be more accurate for roads with better data.

Therefore, although the accuracy of our estimates improved with our three-stage estimation procedure, there was still a potential risk that our model could not be used to predict travel time on some road sections and, by error contagion, across a wider network and over extended periods.

In order to tackle this problem we included a set of restrictions in our estimation procedure. These ensured that our parameter estimates did not violate the rules of stable road network dynamics in any of the road sections. We found that, although these restrictions were only binding on a small proportion of road sections, they were essential to ensure accurate travel time predictions across the network as a whole.
An example of unstable travel time prediction.

Figure 4 provides an example of a road section for which our unconstrained model calibration generated unstable predictions (left hand side graph). After re-estimating the parameters for this road section under more restrictive conditions the predictions were no longer unstable (right hand side).

Figure 4. Model predictions before and after re-calibration for one road section

We used different modelling techniques to ensure the model generated stable predictions. For example, we changed the scale of our model specification to logarithms to ensure that no negative travel time predictions were generated by our simulation model in any road section. We have also restricted the parameter for past road conditions (captured by the past travel time) on current travel time to be, in most road sections, below one. Without these restrictions, a minor increase in travel time on a particular road section at a particular time could potentially lead to ever larger increases in travel time on that road section and across the network in the future. Even if, in reality even after major road traffic disturbances travel time will return to normal sooner or later.5

Although these model constraints restricted the set of possible outcomes from our estimation, they still allowed for a wide range of different predictions.

5 The conditions that need to be met to ensure the stability of the model are slightly more complicated than what this example suggests. They rely on the interaction between different parameters of the model and the level of the explanatory variables at any given point in time.
The National Delay Model

Ultimately, Frontier’s work was to help the HA gain a better understanding about the key drivers of travel time on its network. The model should help the HA answer questions such as:

- How is travel time on a road section (or across the network) expected to change after several incidents and events occur over a given period of time? And for how long is each of these impacts expected to last?
- What is the relative importance of incidents, road works, traffic flow growth and changes in weather conditions for delays on a particular road section or delays across the network?
- How much is the average travel time on a road section (or across the network) expected to decrease if the HA reduces the time it takes to clear incidents or to finish road works on that road section (or across the network)?
- How does congestion, incidents, road works and lane closures on a road section affect the average travel time on nearby road sections and across the network?
- What is the expected impact of adding a new lane on a road to the average travel time on that road and on nearby road sections?

In conjunction with a software company, WINSQL, we constructed a model, called the National Delay Model (NDM), which can easily retrieve the necessary historical information and parameter estimates for any of the more than 2,500 road sections to answer these questions.

The NDM measures the impact on road delays of changes to the road conditions and/or road layout by comparing the average journey time on a selected set of road sections before and after those changes, i.e. comparing a base case with a scenario case. The impact associated with the scenario is then calculated in terms of:

- average travel time, measured as the change in average travel time over the period considered in the simulation, relative to the base case;
- travel time profile, measured by the change in travel time for each 15 minutes over the simulation period, relative to the base case.
- proportion of average travel time, measured as the change in the contribution of the relevant features of road condition (traffic flow, accidents, etc.) to the average travel time, relative to the base case.
- traffic reliability, measured as the change in the number of periods over which journey time is above a given reliability target for a given road section, relative to the base case.

- lateness, measured as the change in the total number of vehicles-hour above a given reliability measure, relative to a base case.

Taken together, these different measures provided a fairly comprehensive view of the impact of changes in road conditions or road layout on delays on a given set of road links.¹

This model provided the HA with new insights into the contribution of different types of road events on delays. For example, the model showed that, overall, major one-off events, such as accidents, have a much lower impact on total road delays than previously thought. But routine events, such as rainfall, are responsible for a much greater share of total delays. Although an accident will be more disruptive of traffic flow in the short run, accidents fortunately only happen sporadically, while England is no stranger to rain. The quantification of this finding has obvious implications on the type of intervention the HA should be investing on to reduce delays and how it quantifies the performance of the road network.

⁶ In addition, the NDM also calculates and presents different robustness measures associated that are specific to the particular road section and scenario simulated such as average prediction errors and related confidence intervals; confidence intervals for parameter estimates and proportion of directly recorded historical information used in the prediction.
How does the NDM look?

The NDM has easy-to-use interface which allows for a large combination of possible scenario simulations. The scope of possible simulation includes:

- increase/decrease the level of traffic flow by a given percentage;
- increase/decrease in the number and/or duration of collisions, breakdowns, debris and other incidents;
- increase the number of road works, with or without lane closure, during specific chosen slot within a chosen start and end date; and
- increase/decrease the number of permanent lanes.

Figure 5. Illustration of the National Delay Model input sheet

Source: Frontier Economics, WinSQL
How does the NDM look like? (cont.)

The results are presented in a numerical and graphical way. These outputs can be easily manipulated and exported to other platforms or saved on the NDM for future use.

Figure 6. Illustration of the National Delay Model outputs

![Illustration of the National Delay Model outputs](image)

Source: Frontier Economics and WinSQL

The model is flexible. The HA can decide to focus on the impact of a particular incident or event on the entire road or on a road section, as illustrated in the figure below.

Figure 7. Example of a simulation of an accident using the NDM

![Example of a simulation of an accident using the NDM](image)

Source: Frontier Economics, WinSQL
Most pleasingly, given the complexity of predicting road delays, the model’s predictions are accurate, on average, and consistent with the HA operational understanding of traffic flow dynamics on its network. This shows that carefully designed empirical models can become powerful decision tools.

Frontier’s work for the HA shows that a tailor-made analysis can help organisations gain a better understanding about the key drivers of their business. And this can in turn lead to better investment decisions.
Annexe: brief technical note on the empirical analysis of road travel time

This note summarises the main stages of the estimation method considered in the analysis of road travel time and briefly discusses the properties of the resulting estimates in the context of a forecast model.

First stage estimation

The first stage is an estimated time series regression, on data every 15 minutes for each road section, of the form:

\[
y_{it} = \hat{\alpha}_i + \hat{\beta}_i x_{it} + \hat{e}_{it}.
\]

Where \( y_{it} \) is log journey time; \( x_{it} \), a \( k \times 1 \) vector, containing lagged log journey times, incidents, interactions, and other relevant features of traffic conditions; \( \hat{\beta}_i \) is a \( k \times 1 \) vector of estimated coefficients; \( \hat{e}_{it} \) a \( T \times 1 \) vector of residuals. Prime denotes transpose.

Second stage estimation

The second stage is a cross-section multivariate regression across roads explaining the first stage parameters by road characteristics, \( W_i \).

\[
\hat{\beta}_i = \hat{a} + \hat{A}W_i + \hat{U}_i
\]

With predicted values:

\[
\hat{\beta}_i = \hat{\alpha} + \hat{A}W_i
\]

Notice, if there were no road characteristics, \( \hat{\beta}_i \) would just be the mean of the \( \hat{\beta}_i \) coefficients.

Third stage estimation: “Shrinkage”

We then choose final estimates to use in the model \( \hat{\beta}_i \) as a function of \( \bar{\beta}_i \) and \( \hat{\beta}_i \).

The usual shrinkage estimator is a linear function:

\[
\hat{\beta}_i = Q \hat{\beta}_i + (I - Q) \bar{\beta}_i
\]

We use a non-linear function based on a decision tree. For instance, if one cannot estimate \( \hat{\beta}_i \), then one uses \( \bar{\beta}_i \). But there are many other cases where one

Annexe: brief technical note on the empirical analysis of road travel time
would prefer the predicted, which uses information on all the estimates, rather than the estimate for a particular road.

Consider a single parameter and for convenience ignore the covariances between the different parameters in an equation. Call the true value of the parameter for road \( i \), \( \beta_i \), and treat it as random around its expected value \( \beta \). We are not making the dependence of \( \beta \) on the road characteristics explicit here to simplify the notation. Then, \( \beta_i = \beta + \eta_i \) where \( E(\eta_i) = 0; E(\eta_i^2) = \delta \). The first stage estimate of \( \beta_i \) is \( \hat{\beta}_i \) with variance \( V(\hat{\beta}_i) = \sigma^2(X_i'X_i)^{-1} \), which is estimated from the first stage regression, and the second stage estimate of \( \beta \) is \( \bar{\beta} \) with variance \( V(\bar{\beta}) = \delta \), which is estimated by the variance of the second stage regression.

Notice that the variance of \( \hat{\beta}_i \) around \( \beta \) is the sampling error in estimating \( \beta_i \) plus the variance of the random part of the coefficient \( \eta_i \):

\[
E(\hat{\beta}_i - \beta)^2 = V(\hat{\beta}) + \delta.
\]

It is more convenient to work with the precisions, the inverse of the variances, with \( h(\hat{\beta}_i) = V(\hat{\beta}_i)^{-1} \) and \( h(\bar{\beta}) = V(\bar{\beta})^{-1} \). The shrinkage estimator is:

\[
\bar{\beta}_i = Q\hat{\beta}_i + (I - Q)\bar{\beta};
\]

where \( Q = h(\hat{\beta}_i)/(h(\hat{\beta}_i) + h(\bar{\beta})) \).

Since \( \hat{\beta}_i \) and \( \bar{\beta} \) are independent:

\[
V(\bar{\beta}_i) = Q^2V(\hat{\beta}_i) + (I - Q)^2V(\bar{\beta})
\]

\[
V(\hat{\beta}_i) = Q^2 / h(\hat{\beta}_i) + (I - Q)^2 / h(\bar{\beta})
\]

\[
V(\bar{\beta}) = (h(\hat{\beta}_i)/(h(\hat{\beta}_i) + h(\bar{\beta}))^2) + (h(\bar{\beta})/(h(\hat{\beta}_i) + h(\bar{\beta}))^2)
\]

The standard error is the square root of this expression.

Maximizing fit vs. obtaining sensible forecasts

The equation used in the simulation model is then:

\[
y_{it} = \tilde{\alpha}_t + \hat{\beta}_i x_{it} + \tilde{\epsilon}_{it}
\]

Where the intercept is estimated to ensure the residuals average zero;

\[
\tilde{\alpha}_t = \bar{y}_i - \hat{\beta}_i \bar{x}_i
\]
(Note that, since $y_{it}$ is log journey time when we do the transformation to journey time, we also have to do a further adjustment to ensure that the residuals have mean zero.)

A consequence of this procedure is that the in-sample fit of (5) will be worse than that of (1), since the $\hat{\beta}_i$ is the estimator that gives maximum fit. However, the main criteria for choosing $\hat{\beta}_i$ is to give sensible forecasts and model properties, which is different from maximising fit. There is a lot of evidence that using a restricted estimator, like $\hat{\beta}_i$, gives better out of sample forecasting properties and more sensible model properties than using the unrestricted estimator $\hat{\beta}$. Baltagi et al. (2000, 2003) review the issues and provide evidence.

There are two main reasons for these better properties: parameter estimation error and model misspecification.

**Estimating extra parameters** adds extra uncertainty to the forecast and there is a large amount of evidence that simpler, more parsimonious models, with fewer estimated parameters, forecast better; even if the extra parameters are statistically significant. Clements and Hendry (1998) discuss this issue.

**Model misspecification** arises because there are always omitted variables. Suppose the true model is:

\[
y_{it} = \alpha_i + \beta_i' x_{it} + \gamma' w_{it} + e_{it}
\]

And we omit $w_{it}$, and in the sample the correlation between the included and omitted variables is given by:

\[
w_{it} = \delta_{it}' x_{it} + v_{it}
\]

Then our estimated coefficients are given by:

\[
\hat{\beta}_i = \beta_i + \gamma \delta_{it}
\]

These chance sample correlations $\delta_{it}$ are not structural, so are likely to average to zero over roads, and will change as our sample changes, worsening our forecasts. But they can be very large in individual cases, causing a large dispersion in the $\hat{\beta}_i$. This is apparent in our second stage regressions (2) where the $\hat{U}_i$ have a large variance and fat-tails and many of the $\hat{\beta}_i$ are implausible. Notice that if $\delta_{it}$ are large these deviations can be very significant, so we are very cautious about using significance of individual coefficients to guide our choice in the decision tree. Boyd and Smith (2002) discuss these issues in more detail, using macroeconomic examples.
This problem is widely recognised in the literature. Equation (1) picks up the effect of the $w_t$ through its correlation with the $x_{it}$ and so fits better, but we do not want our forecasts and model simulations to be contaminated by that effect, so we prefer to use $\tilde{\beta}_t$ when $\beta_t$ is extreme.

**Combining the equations into a system**

The individual equations can then be combined into a system, which has the form of a non-linear vector autoregression:

\[(10) \ Y_t = Af(Y_{t-1}, X_t) + u_t\]

Where $Y_t$ is a $N \times 1$ vector of log journey times and $A$ is a $N \times N$ matrix. The non-zero entries in $A$ will correspond to the coefficients on lagged values for the road section and for adjacent road sections. This structure allows shocks to be transmitted through the network and enables us to predict the effects of particular shocks, e.g. an incident, on the system as a whole. The non-linearity comes from the interaction of lagged journey times with the other variables in $X_t$, incidents, weather etc. The non-linearity complicates the analysis of the stability of the system, since the time profile of a response to a shock will not be constant, but differ depending on the values of $Y_{t-1}, X_{t-1}$.

**Main references**


Duncan N. C. (1986), Suggested Revision of COBA Treatment of Rural Link Capacity, Transport and Roads Research Laboratory

Halcrow Fox (1992) The Hourly Capacity of Inter Urban Road Links

Incident Cost Benefit Assessment user Manual (INCA) (2009), Matt Macdonald

QUADRO 4 User Manual (2011) Department for Transport