On the Estimation of Temporal Mileage Rates

R. Eddie Wilson
Jillian Anable (Aberdeen), Sally Cairns (TRL/UCL), Tim Chatterton (UWE),
Simon Notley (TRL), John Lees-Miller (Brisol)

Faculty of Engineering
University of Bristol

July 17, 2013

UK MOT (Ministry of Transport) test

VT20 MOT Test Certificate

VOSA Vehicle & Operator Services Agency

This certificate has been issued according to the conditions and notes on the back of this certificate.

Note: If you have doubts as to whether this certificate is valid, please use the service described in note 3 overleaf to check.

MOT test number: 761710136293
Make: VAUXHALL
Odometer reading: 105420 Miles
Registration mark: T203UNP
Model: ASTRA
Test class: IV
Vehicle identification or chassis number: WOLOTGF35X8091395
Colour: WHITE
Approximate year of first use: 1999
Expiry date: AUGUST 25th 2007 (ZERO SEVEN)
Issue date/time: AUGUST 18th 2006 (ZERO SIX) 13:30
Fuel type: Petrol
Authentication number: 0849079144895185556410227
Design gross weight (goods vehicles): kg
Advisory Notice Issued: NO
Test station number: 80572

For all vehicles with more than 8 passenger seats
Seat belt installation checked this test: N/A
Number of seat belts fitted at time of installation check: N/A
Previous installation check date: N/A

Issuer's name in CAPITALS: D. S. BRYANT
Signature of Issuer

Warning: A test certificate is not evidence that the vehicle is in a satisfactory condition.

Check carefully that the above details are correct.

Do not accept a certificate which has been altered.

Reg Mark: T203UNP
Make: VAUXHALL
VTS Number: 80572
MOT Expiry: AUGUST 25th 2007 (ZERO SEVEN)

Peel here

Inspection Authority: HANHAM MOTOR COMPANY 126 BRYANTS HILL ST GEORGE BRISTOL BS5 8RJ

- MOT: the UK's annual safety inspection for all road vehicles older than 3 years
- Since 2005: the results have been captured and stored digitially
- Since November 2010 — the DfT has published this data online - spanning back to 2005.
- Key interest: the *odometer reading* recorded at each test.

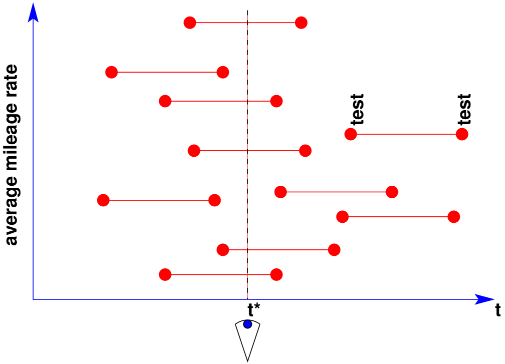
Basic analysis object: *intervals* and their attributes

- Re-arrange blocks of same-vehicle data into consecutive pairs of tests:

Interval	First test			Second test		
	date t_1	miles x_1	place_1	date t_2	miles x_2	place_2
1	2005-08-26	99777	BS	2006-08-18	105420	BS
2	2006-08-18	105420	BS	2007-08-13	113709	BS
3	2007-08-13	113709	BS	2008-08-11	123259	BS
4	2008-08-11	123259	BS	2008-08-11	123259	BS
5	2008-08-11	123259	BS	2009-08-05	132299	BS

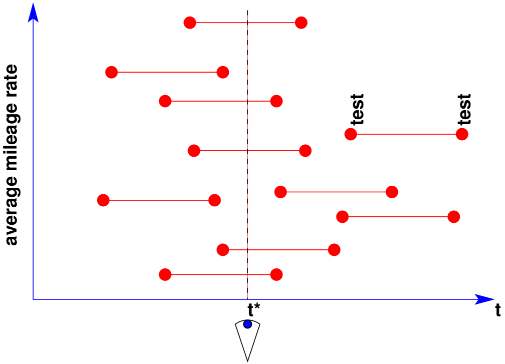
- To which can be linked vehicle-specific attributes:
  VAUXHALL, ASTRA LS 8V, WHITE, P (fuel), 1598 (cc), 1999 (year)
- (Eg) during *interval* 3 — I drove at an average rate of $(123259 - 113709)/364 = 26.24$ miles per day, but we don't know how my mileage was *distributed* during that period.
- These mileage rates are (more or less) complete across the vehicle population — even after cleaning.

Population level statistics: *straddling rate* $\bar{r}(t)$



- Select all N intervals that *straddle* a given *observation date* t^*
- Each interval yields an average (per vehicle) rate r_i.

Population level statistics: *straddling rate* $\overline{r}(t)$

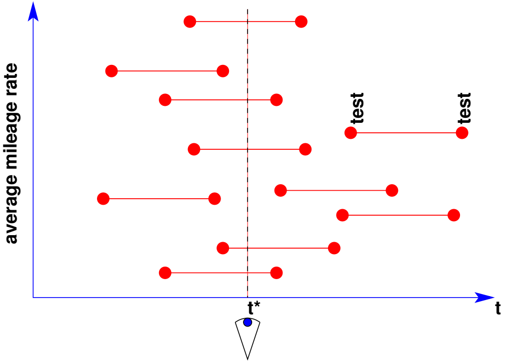


- Select all N intervals that *straddle* a given *observation date* t^*
- Each interval yields an average (per vehicle) rate r_i.

- *Straddling rate* $\overline{r}(t^*)$ is then defined by the **average average**

$$\overline{r}(t^*) = \frac{1}{N}\sum_{i=1}^{N} r_i.$$

Population level statistics: *straddling rate* $\bar{r}(t)$



- Select all N intervals that *straddle* a given *observation date* t^*
- Each interval yields an average (per vehicle) rate r_i.

- *Straddling rate* $\bar{r}(t^*)$ is then defined by the **average average**

$$\bar{r}(t^*) = \frac{1}{N}\sum_{i=1}^{N} r_i.$$

- It is fine for annual statistics: choose $t^* = 1/7/2007$, $1/7/2008$, $1/7/2009$ etc.
- But $\bar{r}(t^*)$ actually incorporates miles driven over the two year span $t^* - 1 \leq t < t^* + 1$.

Central Question for this Paper

Recall that I cannot possibly say anything about an individual's mileage on finer time scales than one year.

But can I derive something about population level mileage over shorter time scales — eg a month?

Central Question for this Paper

Recall that I cannot possibly say anything about an individual's mileage on finer time scales than one year.

But can I derive something about population level mileage over shorter time scales — eg a month?

Possible application: detect the sharp drop in driving in Autumn 2008 following Lehman brothers collapse.

Basic postulate: the population *spot rate* $\phi(t)$

- Suppose there is a population-level *spot rate* $\phi(t)$ that modulates *all* vehicles' mileage (alt. restrict to a population segment).

Basic postulate: the population *spot rate* $\phi(t)$

- Suppose there is a population-level *spot rate* $\phi(t)$ that modulates *all* vehicles' mileage (alt. restrict to a population segment).

- Then each vehicle i has an individual spot rate $\phi_i(t)$ with

$$\phi_i(t) = c_i \phi(t) + \mathrm{noise}.$$

  Here c_i=const.; $\langle c_i \rangle = 1$; and $\langle \mathrm{noise} \rangle = 0$, so that $\phi = \langle \phi_i \rangle$.

Basic postulate: the population *spot rate* $\phi(t)$

- Suppose there is a population-level *spot rate* $\phi(t)$ that modulates *all* vehicles' mileage (alt. restrict to a population segment).

- Then each vehicle i has an individual spot rate $\phi_i(t)$ with

$$\phi_i(t) = c_i \phi(t) + \text{noise}.$$

  Here c_i=const.; $\langle c_i \rangle = 1$; and $\langle \text{noise} \rangle = 0$, so that $\phi = \langle \phi_i \rangle$.

- Let $\psi_i(\tau)$ denote miles driven by i between tests at times $\tau - 1/2$ and $\tau + 1/2$. Then

$$\psi_i(\tau) = \int_{\tau-1/2}^{\tau+1/2} (c_i \phi(s) + \text{noise})\, \mathrm{d}s, \qquad = c_i \int_{\tau-1/2}^{\tau+1/2} \phi(s)\, \mathrm{d}s.$$

From the *spot rate* to the *straddling rate*

- Thus by averaging over tests that straddle t:

$$\bar{r}(t) = \int_{t-1/2}^{t+1/2} \langle \psi_i(\tau) \rangle_i \, \mathrm{d}\tau = \int_{t-1/2}^{t+1/2} \langle c_i \rangle \int_{\tau-1/2}^{\tau+1/2} \phi(s) \, \mathrm{d}s \, \mathrm{d}\tau.$$

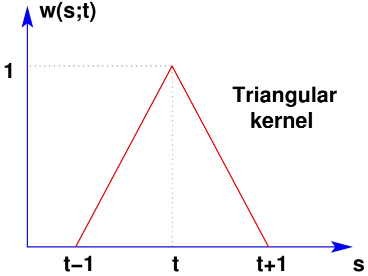
From the *spot rate* to the *straddling rate*

- Thus by averaging over tests that straddle t:

$$\bar{r}(t) = \int_{t-1/2}^{t+1/2} \langle \psi_i(\tau) \rangle_i \, \mathrm{d}\tau = \int_{t-1/2}^{t+1/2} \langle c_i \rangle \int_{\tau-1/2}^{\tau+1/2} \phi(s) \, \mathrm{d}s \, \mathrm{d}\tau.$$

- Simplify integral by $\langle c_i \rangle = 1$ and reverse the order of integration

$$\bar{r}(t) = \int_{t-1}^{t+1} w(s;t)\phi(s) \, \mathrm{d}s,$$



- Thus $\phi(t)$ leads to $\bar{r}(t)$.
  But we want to derive $\phi(t)$ from $\bar{r}(t)$ (which is derivable from data).

From the *straddling rate* to the *spot rate*

- See paper for a whole bunch of Mathematics!!! - upshot:

$$\bar{r}''(t) = \phi(t+1) - 2\phi(t) + \phi(t-1).$$

From the *straddling rate* to the *spot rate*

- See paper for a whole bunch of Mathematics!!! - upshot:

$$\bar{r}''(t) = \phi(t+1) - 2\phi(t) + \phi(t-1).$$

- Isolate $\phi(t+1)$ to derive a time-stepping scheme to evolve $\phi(t)$, with a time-step Δt $(= 1$ month, say)

From the *straddling rate* to the *spot rate*

- See paper for a whole bunch of Mathematics!!! - upshot:

$$\bar{r}''(t) = \phi(t+1) - 2\phi(t) + \phi(t-1).$$

- Isolate $\phi(t+1)$ to derive a time-stepping scheme to evolve $\phi(t)$, with a time-step Δt (= 1 month, say)
- Compute $\bar{r}(t)$ from data at a mesh of points t_i, and estimate $\bar{r}''(t)$ by the divided difference — a natural step size is Δt.
    - in practice: $\bar{r}(t)$ is noisy, so the difference is applied to a smoothing least squares fit spline.

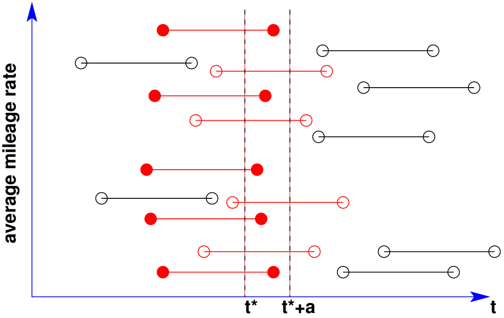
From the *straddling rate* to the *spot rate*

- See paper for a whole bunch of Mathematics!!! - upshot:

$$\bar{r}''(t) = \phi(t+1) - 2\phi(t) + \phi(t-1).$$

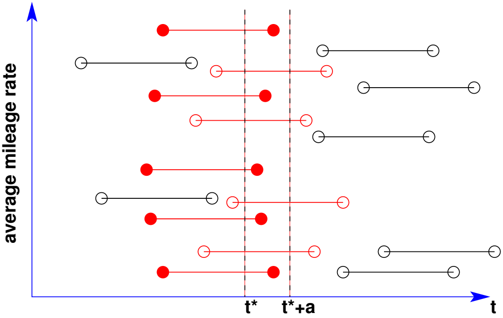
- Isolate $\phi(t+1)$ to derive a time-stepping scheme to evolve $\phi(t)$, with a time-step Δt ($= 1$ month, say)
- Compute $\bar{r}(t)$ from data at a mesh of points t_i, and estimate $\bar{r}''(t)$ by the divided difference — a natural step size is Δt.
  - in practice: $\bar{r}(t)$ is noisy, so the difference is applied to a smoothing least squares fit spline.
- Unfortunately: 2 years of initial data for $\phi(t)$ are required — at the fine scale resolution Δt.

Refinement of the *straddling rate* idea



- Select only the intervals that *straddle* t^* and with right hand ends before $t^* + \alpha$, with $\alpha \leq 1$ year.
- Call resulting average average *straddle rate* $\bar{r}_\alpha(t)$

Refinement of the *straddling rate* idea

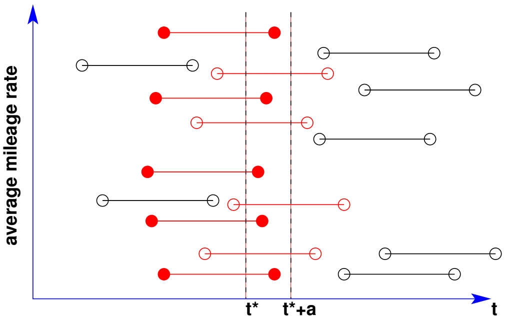


- Crank the handle to give:

$$\bar{r}''_\alpha(t) = \frac{1}{\alpha}\left[\phi(t+\alpha) - \phi(t)\right]$$
$$-\frac{1}{\alpha}\left[\phi(t-1+\alpha) - \phi(t-1)\right]$$

- Select only the intervals that *straddle* t^* and with right hand ends before $t^* + \alpha$, with $\alpha \le 1$ year.
- Call resulting average average *straddle rate* $\bar{r}_\alpha(t)$

Refinement of the *straddling rate* idea

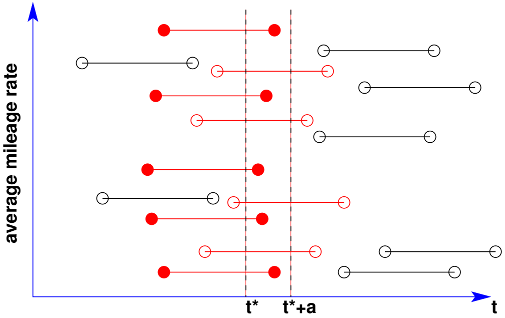


- Select only the intervals that *straddle* t^* and with right hand ends before $t^* + \alpha$, with $\alpha \leq 1$ year.
- Call resulting average average *straddle rate* $\bar{r}_\alpha(t)$
- Crank the handle to give:

$$\bar{r}''_\alpha(t) = \frac{1}{\alpha}\left[\phi(t+\alpha) - \phi(t)\right] - \frac{1}{\alpha}\left[\phi(t-1+\alpha) - \phi(t-1)\right]$$

- Gives time-stepping scheme: but only $1 + \alpha$ years of initial data required.

Refinement of the *straddling rate* idea



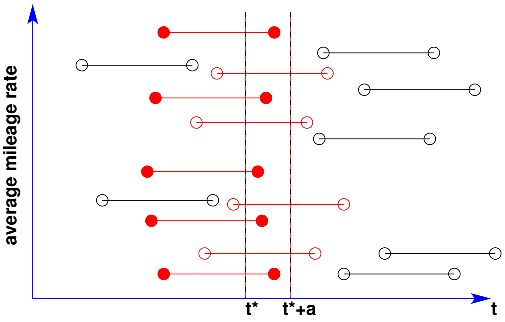
- Select only the intervals that *straddle* t^* and with right hand ends before $t^* + \alpha$, with $\alpha \leq 1$ year.
- Call resulting average average *straddle rate* $\overline{r}_\alpha(t)$

- Crank the handle to give:

$$\overline{r}''_\alpha(t) = \frac{1}{\alpha}\left[\phi(t+\alpha) - \phi(t)\right] - \frac{1}{\alpha}\left[\phi(t-1+\alpha) - \phi(t-1)\right]$$

- Gives time-stepping scheme: but only $1+\alpha$ years of initial data required.
- So interest is in $\alpha \to 0$, which gives $\overline{r}'_\alpha(t) \simeq \phi'(t) - \phi'(t-1)$ (natural meaning)

Refinement of the *straddling rate* idea



- Select only the intervals that *straddle* t^* and with right hand ends before $t^* + \alpha$, with $\alpha \leq 1$ year.
- Call resulting average average *straddle rate* $\bar{r}_\alpha(t)$

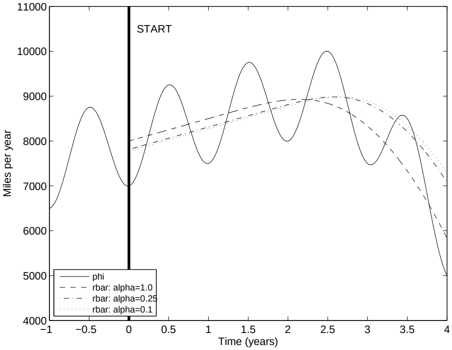
- Crank the handle to give:

$$\bar{r}''_\alpha(t) = \frac{1}{\alpha}\left[\phi(t+\alpha) - \phi(t)\right] - \frac{1}{\alpha}\left[\phi(t-1+\alpha) - \phi(t-1)\right]$$

- Gives time-stepping scheme: but only $1+\alpha$ years of initial data required.
- So interest is in $\alpha \to 0$, which gives $\bar{r}'_\alpha(t) \simeq \phi'(t) - \phi'(t-1)$ (natural meaning)
- $\alpha \to 0$ means fewer and fewer intervals, means noisy $\bar{r}_\alpha(t)$

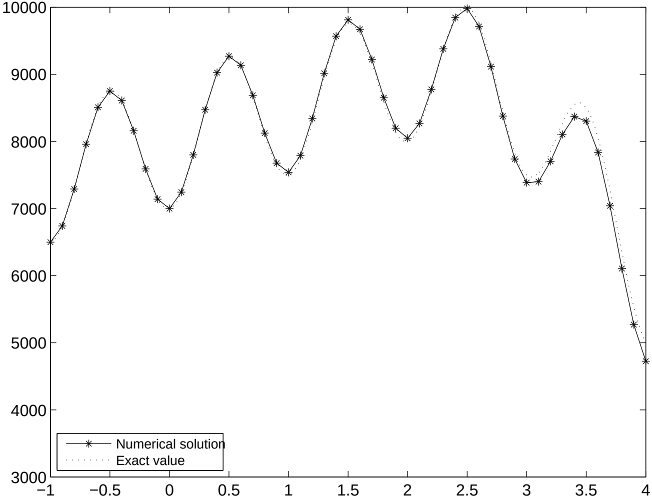
Synthetic data set-up

- Choose *spot rate*
$$\phi(t) = 8000 + 500t - 1000\cos 2\pi t - 1000\left[t-2\right]_+ (t-2)^2,$$
- 10^6 vehicles with tests 1 year apart, test dates uniformly distributed through calendar year
- Vehicle i daily mileage drawn from a distribution modulated by $\phi(t)$ and (random) c_i.
- Odometer readings on test dates are synthesised by adding individual vehicle daily totals



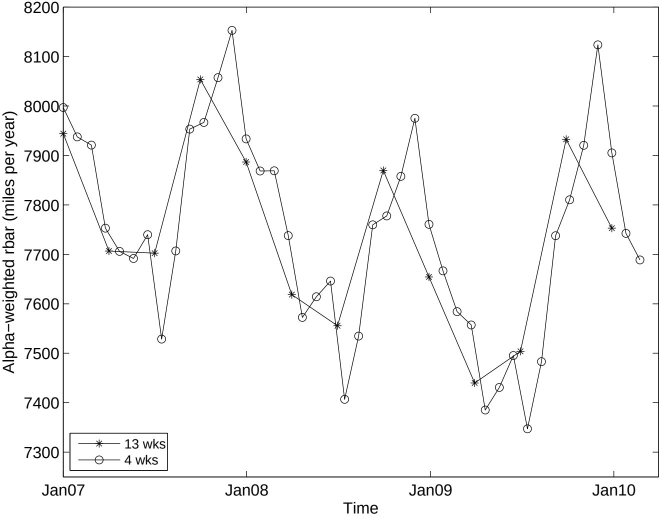
- Periodic component in spot rate $\phi(t)$ is suppressed in straddling rates $\bar{r}_\alpha(t)$

Results with synthetic data: $\alpha = \Delta t = 0.1$ years



- Reconstructed $\phi(t)$ almost indistinguishable from ground truth.

Straddling rates $\bar{r}_\alpha(t)$ for real-world data



- Seasonal component shouldn't be there: underlying assumptions of the theory are broken

Implicit assumptions in the theory. . .

A1 We assume that tests (odometer readings) are exactly one year apart.

Implicit assumptions in the theory...

A1 We assume that tests (odometer readings) are exactly one year apart.

- OKish — theory can be generalised.

Implicit assumptions in the theory...

A1 We assume that tests (odometer readings) are exactly one year apart.

- OKish — theory can be generalised.

A2 We assume that tests occur at same frequency on average throughout year.

Implicit assumptions in the theory. . .

A1 We assume that tests (odometer readings) are exactly one year apart.

- OKish — theory can be generalised.

A2 We assume that tests occur at same frequency on average throughout year.

- Not true — but easy to fix theory.

Implicit assumptions in the theory. . .

A1 We assume that tests (odometer readings) are exactly one year apart.

- OKish — theory can be generalised.

A2 We assume that tests occur at same frequency on average throughout year.

- Not true — but easy to fix theory.

A3 We assume that a vehicle's mileage rate is independent of the time of year of at which it is tested (and its odometer is read).

Implicit assumptions in the theory. . .

A1 We assume that tests (odometer readings) are exactly one year apart.

- OKish — theory can be generalised.

A2 We assume that tests occur at same frequency on average throughout year.

- Not true — but easy to fix theory.

A3 We assume that a vehicle's mileage rate is independent of the time of year of at which it is tested (and its odometer is read).

- Completely wrong. And very hard to fix.

On **A3**: fails because a pattern in new vehicle registrations throughout the year (in the UK).

Conclusions and Further Work

- Methods developed which extract population-level *spot rate* mileage from widely spaced individual vehicle odometer readings. Success with synthetic data.

Conclusions and Further Work

- Methods developed which extract population-level *spot rate* mileage from widely spaced individual vehicle odometer readings. Success with synthetic data.
- UK MOT data set: some fixes/patches to theory are needed.

Conclusions and Further Work

- Methods developed which extract population-level *spot rate* mileage from widely spaced individual vehicle odometer readings. Success with synthetic data.
- UK MOT data set: some fixes/patches to theory are needed.
- Please contact me if you know of other datasets (international) in which odometer readings are systematically collected.
- These methods have the potential to complement / replace existing survey-based / link-flow techniques for estimating population-level mileage.