Infrastructure maintenance, regeneration and service quality economics: A rail example

Marc Gaudry¹, Bernard Lapeyre², Émile Quinet²

1. Agora Jules Dupuit (AJD), Département de sciences économiques
   Université de Montréal
   Montréal, marc.gaudry@umontreal.ca

2. Paris-Jourdan Sciences Économiques (PSE)
   École des Ponts ParisTech (ENPC)
   Paris, bernard.lapeyre@univ-paris-est.fr, quinet@enpc.fr

An earlier draft of 6th October 2010 by Gaudry and Quinet, entitled Optimisation de l’entretien et de la régénération d’une infrastructure: exploration d’hypothèses, benefitted from comments by Bernard Caillaud and Matthieu de Lapparent and was presented without econometric tests at the Kuhmo Nectar Conference on Transportation Economics in Stockholm on 1st July 2011 under the title “Joint optimization of continuous maintenance and periodic renewal”. The authors thank Marc Antoni, Richard Arnott, David Meunier and Yves Puttalaz for discussions or comments, Cong-Liem Tran for computing assistance and are grateful to Société nationale des chemins de fer français (SNCF) for financial support and for allowing inclusion in this version of estimates based on databases constructed by Michel Ikonomov and Pascaline Boyer. Exploratory estimates obtained from fixed form regression specifications were presented at the Kuhmo Nectar Conference on Transportation Economics in Berlin on 21st June 2012 through David Meunier’s good offices.

JEL classification codes: C61, D61, D92, H54, R41, R48

Université de Montréal
Agora Jules Dupuit, Publication AJD-136-E
----------------------
Paris-Jourdan Sciences Économiques
Working Paper N° 2011-03

Version 1, January 2011; Version 20, 16th May 2014
Abstract
This paper proposes a formalized framework for the joint economic optimization of continuous maintenance and periodic regeneration of rail transport infrastructure taking into account output consisting not only in traffic levels but also in track service quality. Core features of the formulation readily apply to other infrastructures, such as roads. Derivations are made principally under certainty assumptions, using Pontryagin procedures, but some uncertainty conditions are explored through Bellman processes. The resulting general expression for the marginal maintenance cost nests Newbery’s 1988 known result, derived with roads in mind.

The model equations are tested on national French rail track segment databases using Box-Cox transformations and directed (including spatial) autocorrelation of residuals and results are compared to rail regeneration and maintenance practices prevailing in France. In particular, the inverted U-shaped time profile of maintenance expenditures predicted by the model is verified for the part of the network subjected to regeneration and, for the central bout of the three-phase approach, the estimated total current maintenance cost and track service quality supply equations for the whole network include statistically significant target service and service trajectory correction terms implied by the joint optimization approach.

Keywords: rail infrastructure, road infrastructure, airport infrastructure, current maintenance, regenerative maintenance, infrastructure services, marginal cost pricing, Pontryagin maximum principle, Hamilton-Jacobi-Bellman simulation, Box-Cox transformation, directed autocorrelation, spatial autocorrelation, French rail network, rail track quality supply.

JEL classification codes: C61, D61, D92, H54, R41, R48.
Table of contents

1. Infrastructure service, maintenance and regeneration ............................................ 4
2. Context of the approach adopted for rail .................................................................. 4
3. Jointly optimal maintenance and regeneration ........................................................ 5
   3.1. An optimal maintenance policy for given renewal horizon T ............................. 6
   3.2. Determination of the optimal renewal horizon T .............................................. 9
   3.3. Consequences for optimal intertemporal infrastructure pricing .................... 9
4. Effects of uncertainty on maintenance policy .......................................................... 11
5. Econometric and calibration tests ....................................................................... 13
   5.1. Data and estimation strategy ................................................................. 14
   5.2. Are there three phases in current rail Maintenance databases? ..................... 19
   5.3. Model teachings: cruising Maintenance and Service levels ....................... 22
   5.4. Preliminary estimation of Phase B Target Service level S_c .......................... 23
   5.5. Estimation of Phase B Maintenance expenses u ............................................ 28
   5.6. Estimation of derived Phase B Service supply S .......................................... 33
   5.7. Other calibrations ................................................................................... 34
6. Conclusion .......................................................................................................... 35
7. Acknowledgements .......................................................................................... 36
8. References ....................................................................................................... 37
9. Appendix 1. The optimal duration of a renewal .................................................. 40
10. Appendix 2. Discretization of the singular stochastic control problem ............ 41
11. Appendix 3. Adding speed to maintenance cost models ...................................... 42

List of figures

Figure 1. Longitudinal track degradation index NL and ballast tamping frequency ....... 6
Figure 2. Typical evolution of service quality and maintenance expense for T=35 ....... 8
Figure 3. Analytically derived Pontryagin and Hamilton-Jacobi-Bellman simulation results 12
Figure 4. Maximum allowed speed (km/h) on track segments classified by UIC group ...... 15
Figure 5. Time profile of cost u with respect to cumulative traffic, UIC group 3 segments .. 20
Figure 6. Changes in Service quality, Phase B segments only ..................................... 24
Figure 7. Total Maintenance cost (euros per km) u, censored 2007 sample (673 obs.) ...... 28
Figure 8. Year-to-year track Service quality changes from 2006 to 2008 ....................... 34
Figure 9. Relation between the yearly traffic and the lifetime of renewals ..................... 35

List of frames

Frame 1. Parameters and functions used for the simulations ......................................... 12
Frame 2. Impact of aging factors on the productivity of a unit of maintenance expenses .... 26

List of tables

Table 1. Revenue differences between standard and optimal intertemporal pricing rules ..... 11
Table 2. Estimation sequence for Maintenance and Service equations .......................... 14
Table 3. Mean values of principal variables available by track segment .................... 15
Table 4. Sign conditions for a maximum or a minimum with two BCT on the same X_k ....... 16
Table 5. Main variables used to establish the existence of three maintenance cost phases.... 19
Table 6. Phase tests with cumulative traffic variables, classic line network, 1999 & 2007 ... 21
Table 7. Embedded Target Service S_c function results (4060 obs., 2001-2007) .............. 27
Table 8. SNCF regions in 2007, their number of neighbors and of segments in the sample.. 30
Table 9. Phase B Maintenance cost u without max. speed v (580 obs., 2007) .................. 32
Table 10. Phase B Maintenance cost u without max. speed by train type v_c (580 obs., 2007)32
Table 11. Explaining Service quality changes ΔS by factors defined for 2007 and 2006 ...... 34
Table 12. Ratio between Renewal and Maintenance expenses for several levels of traffic ... 35
Table 13. Phase B Maintenance cost u with max. speed by train type v_c (580 obs., 2007) .... 42
1. Infrastructure service, maintenance and regeneration

Optimization of transport infrastructure maintenance, notably for roads and railways, is a longstanding research concern. Most work belongs to the operations research literature where the types and sequencing of maintenance operations are addressed but little attention is paid to their economic consequences. Other work, more concerned with economics, and notably with pricing, focuses on the econometrics of current maintenance expenditures but takes little account of interactions among types of maintenance, notably between current and regenerative maintenance categories to their service quality impacts. The pricing implications, developed with a rail infrastructure focus, admit of numerous key features common to other transport infrastructure such as roads or airport runways and taxiways, as already pointed out in the first version of this paper (Gaudry & Quinet, 2011), in French.

After a short literature review, the third section presents the optimization model and its resolution, as well as its consequences for pricing, uncertainty being introduced in the next section. The fifth presents econometric estimates and comparisons of results to infrastructure management practices. The conclusion points to their limitations and potential refinements.

2. Context of the approach adopted for rail

The first literature stream just mentioned derives from engineering knowledge managerial rules pertaining to maintenance operations and their scheduling. It includes both academic studies and administrative documents produced by national and international agencies supplying advice and computer-based decision tools.

The operations research orientation of such work is clear from recent reviews (e.g. Gu et al., 2012; Kobayashi et al., 2012). A notable example of computer-based methods produced by agencies is the Highway Development and Management program (currently HDM-4) distributed by the World Bank to make recommendations on the analysis, planning, management and appraisal of road maintenance, improvements and investment decisions. It provides detailed modeling of pavement deterioration and maintenance effects as functions of local characteristics (soil, weather…) and calculates annual costs of road construction, maintenance, vehicle operation, and travel time. But, despite the use of optimization methods allowing for the normative computation of minimum cost output configurations and of present value indicators, such tools are not strictly speaking economic in the sense that they may not typically allow for explicit cost function and pricing analyses, or portend to represent the supply behavior of infrastructure owners.

The second literature stream, centered on the estimation of cost functions, takes a more usual economic approach and renounces any detailed analysis of technical processes to estimate econometric linkages among total costs, outputs and input prices if available. Typically, a behavioral assumption of short run cost minimization by infrastructure owners is made with an eye for pricing implications, especially in European countries where the marginal cost pricing doctrine of the European Union Commission is applied.

A particularly complete synthesis of this stream is found in the numerous 2008-2009 country reports of the Cost Allocation of Transport INfrastructure cost (CATRIN) Consortium, summarized in Wheat et al. (2009), and in more recent related contributions (e.g. Andersson et al., 2012). CATRIN studies notably provide, among other contributions, comparable
coordinated statistical analyses linking annual maintenance expenditures to traffic and technical track characteristics for five European country rail networks. The resulting cost functions rely on minimal technical knowledge and avoid dealing with the interaction between current and regenerative types of maintenance expenditures, a matter abundantly dealt with in literature of the first category where it is addressed primarily in a technical manner. Andersson et al. and Andersson & Björklund (2012) do bring up or study the rarely mentioned matter of regeneration expenditures but without relating it to that of current maintenance expenditures.

Accounting for the interaction between periodic regeneration and continuous maintenance dimensions of assets, and duly deriving implications for pricing, is rare. For Feldstein & Rothschild (1974), current maintenance has no effect on the replacement of capital goods (machines); a closer example is the probabilistic modeling of the choice between maintenance and replacement of a component of an asset of variable service life duration, namely the engine of an urban bus (Rust, 1987). Concerning transport infrastructure, Newbery (1988a, 1988b) stands out with a formulation of a road infrastructure deteriorating with traffic axle loads and recovering its initial service quality when regeneration is carried out after a certain time interval. His main result is that, under stationary network conditions (i.e. constant road traffic on all links and life durations uniformly distributed in time), revenue generated under marginal cost pricing equals the optimal regeneration expense. A key structural feature of his seminal analysis, our starting point, is that current maintenance has no effect whatsoever on regeneration timing or requirements, an assumption shared by other authors distinguishing between current and periodic maintenance (regeneration) expenditure, such as Small et al. (1989) for roads, or Lévi et al. (2010) for rail tracks.

3. Jointly optimal maintenance and regeneration

Consider an infrastructure defined by a vector $K$ of technical characteristics and supporting at any time $t$ a traffic of density $q(t)$ which creates damages that can be remedied by a combination of current maintenance expenditures $u(t)dt$, during any time interval $dt$, and of regeneration expenditures $D$, incurred at intervals to be defined, that fully restore the infrastructure to its initial un-degraded quality state.

The service quality $S(t)$ supplied by the infrastructure at time $t$ depends on repair and degradation functions $C(\cdot)$ and $f(\cdot)$ varying in manners determined by the above technical characteristics $K$, current traffic density and maintenance expenditure $u(t)$, but also by time elapsed $t$ and cumulative traffic $Q(t)$ since the last regeneration. Taking the occurrence of the latter as the origin of times, cumulative traffic is the integral of $q(t)$ from the moment of regeneration to $t$, namely $Q(t) = \int_0^t q(s)ds$, and one may write for service quality:

$$dS(t)/dt = C(K,Q(t),q(t),t,u(t)) - f(K,Q(t),q(t),t).$$

Combined, or joint optimization pertains to both the current expenditure function $u(t)$ and to the sequence of renewals, i.e. to their timing. The objective function of an infrastructure manager assumed to maximize the collective surplus is then defined as the present value of services supplied by the infrastructure minus maintenance expenditures —both current ones $u(t)dt$ for each period $dt$ and regeneration ones $D$ incurred at intervals $T_i$. Assuming that the current value of services supplied at period $t$ is the product of traffic $q$ by the monetary value of incurred service, expressed by the function $g(S)$, the optimization problem may be written, neglecting any service inconvenience caused by maintenance, as:

---

1. In particular, Box-Cox transformations are applied by all to CES, and by some to other specifications.
2. Austria, France, Great Britain, Sweden, Switzerland.
3. Like Newbery (op. cit.), they ignore any pavement life prolongation caused by non regenerative maintenance.
\[ M^* = \max_{u(t)} \{ \sum_{t=0}^{T} \left[ \int_{0}^{t} [-u(t) + q(S(t))] e^{-\mu t} dt - De^{-\beta t_i} \right] e^{-\beta T_i} \} \] , for \( T_0 = 0 \),

(2) with
\[
\begin{align*}
\frac{dS(t)}{dt} &= C(K, Q(t), q(t), t, u(t)) - f(K, Q(t), q(t), t) \\
Q(t) &= \int_{t_i}^{t} q(s) ds
\end{align*}
\]

, for \( T_i < t < T_{i+1} \).

Under the simplifying assumption that traffic \( q(t) \) is constant over time and denoted by \( q \), optimal reignistrations will be regularly spaced at some interval \( T \) and the problem becomes:

\[ M^* = \max_{u(t)} \left\{ \int_{0}^{T} [-u(t) + qg(S(t))] e^{-\mu t} dt - De^{-\beta T} \right\} \frac{1}{1 - e^{-\beta T}} \]

(3)

\[ = \max_{T} \left\{ J(u(t), T) - De^{-\beta T} \right\} \frac{1}{1 - e^{-\beta T}}. \]

To solve it, first search for the current maintenance function \( u(t) \) which maximizes \( M^* \), and therefore \( J \), for given \( T \); then maximize the resulting \( M^*(T) \) with respect to \( T \).

### 3.1. An optimal maintenance policy for given renewal horizon \( T \)

Maximization of \( J(u(t), T) \) for given \( T \) is a classical dynamic programming problem. We first assume that its solution function \( u(t) \) is limited both downward by 0 (to exclude negative maintenance) and upward by \( m \) (to account for the fact that constraints, such as time windows for track works and the cost of regeneration, limit feasible maintenance per unit of time); and we specify functions \( C \) and \( g \) used above. Taking the former first, we write:

\[ dS(t)/dt = h(K, Q(t), t) [u(t)] - f(K, Q(t), q, t) \]

(4)

where \( f(\cdot) \) expresses degradation in the absence of current maintenance and \( h(\cdot) \) the physical productivity of a unit of maintenance expenditure, which depends on technical segment characteristics \( K \), on elapsed time \( t \) and cumulative\(^4\) traffic \( Q(t) \) since the last regeneration, but not on the volume of maintenance expenditure or the prevailing service level.

This double independence from levels in (4) is deemed consistent with a lumpy saw-like process whereby, to maintain a target service level (say a combination of comfort and safety), growing track degradation is periodically reduced (and service increased) by successive current maintenance operations. In Figure 1, this occurs with ballast tamping (immediately followed by preventive\(^5\) rail grinding) effected at an increasing frequency over the lifetime of a ballast, a duration \( T \) effectively prolonged by such maintenance known to decrease by about half the slope of the degradation curve. This key effect of maintenance expenses on \( S \) and \( T \) cannot be ignored: “maintenance causes continuous changes in quality (deterioration)”, as long ago emphasized in a housing maintenance policy model by Arnott et al. (1983).

**Figure 1. Longitudinal track degradation index NL and ballast tamping frequency**

\[ /: \text{track degradation index NL} \quad \downarrow: \text{effect of tamping on NL} \]

\[ \text{Duration } T \text{ between ballast regenerations} \quad \text{Minimum service level} \]

\(^4\) In empirical work below, current traffic \( q(t) \) also matters through \( \partial Q / \partial t \) in \( h' \) of Equation 8.

\(^5\) Grindings done between tampings will have no effect on such longitudinal or transversal degradation indices.

For an example of the effect of tamping only on track settlement, see in Ongaro & Iwnicki (2009, Figure 2) an index, also in physical units, almost identical over time to our Figure 1 drawn from Antoni et al. (1989).
Expectations about $h(\cdot)$ and $f(\cdot)$, both assumed to be positive, are that the productivity of maintenance decreases with $Q$ and $r$ and that, in the absence of any maintenance, service degradation increases (track quality $S$ falls) with current traffic $q$.

Concerning the operator’s attitude to $S$, constant risk aversion is simplest with $g$ specified as:

$$(5) \quad g(S) = -\alpha S^{-\lambda},$$

and, if sensitivity to quality of service varies by traffic category $i$ but aversion $\lambda$ does not, and if we consider that $Q$ is a homogeneous measure of either cumulative tonnage $W$ or cumulative number of trains $N$, the Hamiltonian may be written:

$$(6) \quad H = -u(t) - \sum_{i=1}^{c} q_i \alpha_i S(t)^{-\frac{\lambda}{\alpha_i}} + y(t) \left[ h(K, Q, t) u(t) - f(K, Q, q_1, \ldots, q_c, t) \right],$$

the solution of which, under Pontryagin’s maximum principle, will satisfy three relations:

$$\begin{align*}
Max_u H &= Max_u \left\{ u(t) \left[ h(K, Q, t) y(t) - e^{-\beta y} \right]\right\} \\
H_S + \dot{y} &= 0 \\
H_y &= dS/\text{dt}
\end{align*}$$

with the two-limit constraint on $u(t)$, located in the closed interval $[0$ and $m]$, and the transversality condition $y(T)=0$ satisfied. The latter implies that the solution $u(t)$ cannot be totally within the interval $[0, m]$ and that one must distinguish three phases depending on whether the value of $u$ which maximizes $H$ is within, or at the limits of, the central domain:

**Phase A:** $h(K, Q, t) * y(t) - e^{-\beta y} < 0$. With $u(t)=0$ over this interval $I$, we have:

- $S(t)$ found by integration of $dS/\text{dt} = h(K, Q, t) * u(t) - f(K, Q, q, t)$; it is deduced, the phase starting at moment $t_f$ with $S(t_f)$ denoting service quality at that moment;
- $y(t)$ determined by $H_S + \dot{y} = 0$, which yields $y(t) = y(t_f) - \int_{t_f}^{t} \alpha q \lambda S(t)^{-\frac{\lambda}{\alpha} - 1} e^{-\beta y} \, dv$.

**Phase C:** $h(K, Q, t) * y(t) - e^{-\beta y} > 0$. With $u(t)=m$ over this interval $III$, we have:

- $S(t)$ by integration of $dS/\text{dt} = h(K, Q, t) * m - f(K, Q, q, t)$, the phase starting at time $t_m$;
- $y(t)$ determined by $H_S + \dot{y} = 0$, which yields $y(t) = y(t_m) + \int_{t_m}^{t} q \lambda \alpha e^{-\beta y} \, dv$.

**Phase B:** $h(K, Q, t) * y(t) = e^{-\beta y}$. In reverse order now, we have over this central interval $II$:

- $y(t) = [1/h(K, Q, t)] e^{-\beta y}$, by simple manipulation of the phase condition;
- $S_c(t)$, determined by $H_S + \dot{y} = 0$, and to be called cruising service quality:

$$\dot{y} + H_S = \frac{-j e^{-\beta y}}{h(K, Q, t)} - \frac{\partial h}{\partial t} + \left[ \frac{\partial h}{\partial Q} \frac{\partial Q}{\partial t} \right] e^{-\beta y} + \alpha q \lambda S_c(t)^{-\frac{\lambda}{\alpha} - 1} = 0,$$

which, by simplification of the derivative of $h$ with respect to $t$, may be written:

$$-\alpha q \lambda S_c(t)^{-\frac{\lambda}{\alpha} - 1} = -\frac{1}{h(K, Q, t)} \left( j + \frac{dh}{h dt} \right),$$

or $S_c(t)^{-\frac{\lambda}{\alpha} - 1} = \frac{1}{\alpha q \lambda} \left( j + \frac{dh}{h dt} \right)$, or even, taking into account the above mentioned breakdown of $q$ among $c$ traffic categories $q_i$, applying a logarithmic transformation\footnote{Instead of a logarithmic form, the Box-Cox transformation (BCT) is often used (e.g. Montmarquette & Blais, 1987) as a convenient Arrow-Pratt measure of constant relative risk aversion, concavity ($\lambda < 1$) indicating risk avoidance and convexity ($\lambda > 1$) a preference for risk. Here the BCT will be reserved for the variables in $h(\cdot)$.} and collecting terms:

$$(8) \quad \log[S_c(t)] = \frac{1}{\lambda + 1} \left\{ \log \left( \sum_{i=1}^{c} q_i q_i \right) + \log \left( \frac{\lambda}{\lambda - 1} \right) \right\} - \log \left( 1 + \frac{h^2}{jh^2} \right).$$
with \( u(t) \) then given by:

\[
(9) \quad u(t) = f(K, Q, q, t) + \frac{1}{h(K, Q, t)} \frac{dS_c(t)}{dt}.
\]

One may now determine dates of phase changes and system behavior between \( 0 \) and \( T \) for realistic initial and final conditions.

For system behavior, and starting for convenience at \( T \), the end of the period, transversality condition \( y(T) = 0 \) then requires to be in Phase A, and current maintenance \( u(t) \) to be nil. Two possibilities arise when one starts backing-up in time:

(i) either one stays in Phase A because \( y(t) \) satisfies the corresponding inequality restriction.

The resulting optimal policy is then to perform no current maintenance and to regenerate periodically, as in the road case studied by Newbery (1988a, 1988b);

(ii) or, at a certain instant \( t_f \), one has \( y(t_f) = \left[1/h(K, Q(t_f), t_f)\right]e^{-\beta_f} \), as in Phase B where \( 0 < u(t) < m \), and service quality then follows trajectory \( S_c(t) \). Further, as one approaches period beginning \( t=0 \), a number of possibilities arise depending on how service quality \( S(0) \) achieved by the previous regeneration compares to \( S_c(0) \).

If, as in standard practice illustrated in Figure 2, \( S(0) > S_c(0) \), one again reaches a Phase A state, with \( u(t) = 0 \), i.e. devoid of maintenance. Typical trajectories of service quality and maintenance expenditure are then as shown in Figure 2, with the three phases present.

**Figure 2. Typical evolution of service quality and maintenance expense for \( T=35 \)**

<table>
<thead>
<tr>
<th>A. Service quality (initial value equals 250)</th>
<th>B. Maintenance expenditure (nil close to ( T=35 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Quality of service" /></td>
<td><img src="image" alt="Maintenance expenditure" /></td>
</tr>
</tbody>
</table>

This evolution matches current maintenance practice rather well. For a short period (some months, or at most a few years) after regeneration, maintenance is almost nil (it includes minimal surveillance). During the cruising phase, quality slowly decreases and annual maintenance increases. Shortly before regeneration, maintenance is again nil or minimal.

Concerning dates of phase changes, and starting from a regeneration establishing initial quality \( S_0 \), the first one is reached at time \( t_i \) when, with due degradation, cruising quality \( S_c \) is attained and it is verified that:

\[
(10) \quad S_0 - \int_0^{t_i} f(K, Q, q, t) dt = S_c(t_i).
\]

The next date, \( t_f \), of transition between cruising Phase B and the final Phase C is such that:

\[
(11) \quad h(K, Q, t) \ast y(t_f) = e^{-\beta_f}
\]

and, because \( y(T) = 0 \) also holds, one may deduce from the expression for \( y(t) \) in Phase A:

\[
(12) \quad e^{-\beta_f} = \int_{t_f}^{T} \alpha q \lambda S(t)^{-\lambda-1} e^{-\beta} dv,
\]

which, as an equation, yields the value of \( t_f \) because all the rest is known in terms of \( t_f \), notably function \( S(v) \).
3.2. Determination of the optimal renewal horizon $T$

The assumption of stationary traffic, made to some extent for convenience in the previous section, is now required for an easy resolution. If $J^*(T)$ is, for given $T$, the highest value of $J$ solving $\max_u \{J(u(t), T)\} = J^*(T)$, the optimal duration $T$ is that which maximizes

$$M(T) = \frac{J^*(T) - De^{-\beta T}}{1 - e^{-\beta T}}.$$  

(13)

It is shown in Appendix 1 that, after setting the derivative of $M(T)$ with respect to $T$ equal to 0 to obtain $J^*(T)(1 - e^{-\beta T}) + jDe^{-\beta T} - jJ^*(T)e^{-\beta T} = 0$, the implied optimal solution value for $T$ is a maximum.

3.3. Consequences for optimal intertemporal infrastructure pricing

A full-bloom optimal infrastructure pricing scheme, seriously based on the above results pertaining to cost function phases and on a demand function shifting to allow for substitution among moments of time, should include continuously changing prices and traffics that would be extremely difficult to characterize and would make planning by clients difficult: some stabilizing restrictions are clearly needed.

First assume that demand is not a function of time itself and that quantities demanded are independent across time slices, a weak assumption to the extent that relatively high travel frequencies should not shift in response to modifications in infrastructure charges arising over time from maintenance activities of relatively low frequency. Assume also that the demand function is of the form:

$$q = l(\pi - \alpha S^\lambda) = l(p), \quad \text{or} \quad p = \pi - \alpha S^\lambda = p(q),$$  

(14)

where $\pi$ is the infrastructure charge and $\alpha S^\lambda$ the value of service quality $S$.

The hypothesis previously adopted that annual traffic $q$ is constant over time implies, by this last equation, a charge $\pi$ varying over time in symmetry with service quality. Such a second best assumption, adopted for resolution convenience, is also grounded in practice to the extent that it matches the manager’s stability policy aimed at freight shipper facility investments and at trip scheduling decisions by travelers using competing means of transport. It implies the $p$ is constant over time but that $\pi$ and $S(t)$ can (and effectively do) vary.

Under these assumptions, the optimization of charges involves the maximization with respect to $q$ of actualized collective surplus (SC) arising from managing the infrastructure, with solutions to be determined for the final variables consisting in traffic level $q$ and the sequence of infrastructure charges $\pi(t)$, but involving as intermediate products the sequence of current maintenance expenses $u(t)$ and of renewal expense dates $T_i$ that can be derived from results found in the previous section. With all terms defined, the expression to be maximized is:

$$SC(q) = \frac{1}{1 - e^{-\beta T}} \left\{ \int_0^{T^*} \left[ \int_0^q \left( p(v) + g(S(t)) \right) dv \right] dt - u(t) \right\} e^{-\beta T} dt - De^{-\beta T^*}$$

$$= \frac{1}{1 - e^{-\beta T^*}} \left\{ \int_0^{T^*} \left[ \int_0^q \left( p(v)dv + qg(S(t)) \right) - u(t) \right] dt - De^{-\beta T^*} \right\}$$

$$= \frac{1}{1 - e^{-\beta T^*}} \left\{ \int_0^q p(v)dv + \frac{1}{1 - e^{-\beta T^*}} \left[ \int_0^{T^*} \left[ qg(S(t)) - u(t) \right] dt - De^{-\beta T^*} \right] \right\}$$

$$= \frac{1}{1 - e^{-\beta T^*}} \left\{ \int_0^q p(v)dv + M^*(q), \right\}$$  

(15)

the maximization of which with respect to $q$ yields $p$, then $q$ with values constant over time:
\[ p(q) = -j \frac{dM^*(q)}{dq}. \]

One then calculates infrastructure charges \( \pi(t) \) varying in time and equal to:
\[ \pi(t) = p(q) + \alpha(t) S(t, q)^\lambda, \]
where the first term, constant through time, is the opposite of the actualized mean value of collective marginal surplus increased by service quality \( S(t) \) at moment \( t \). Overall, the generalized cost to the user is constant through time, with variations in infrastructure charges exactly matching changes in service quality.

The completion of optimal tariff constructs requires determining how \( M^*(q) \) varies with \( q \). Note first that variations in \( q \) imply variations in \( T \) but that, by the envelope theorem, no account need be taken of the latter: when traffic increases by \( \delta q \), one considers only variations in \( J^*(T, q) \), the duration of inter-regeneration \( T \) being held constant:
\[ \frac{dM^*(q)}{dq} = \frac{\partial M^*(T, q)}{\partial T} \frac{dT}{dq} + \frac{\partial M^*(T, q)}{\partial q} = \frac{1}{1 - e^{-\beta T}} \frac{\partial J(T, q)}{\partial q}, \]
and then:
\[ p(q) = -j \frac{1}{1 - e^{-\beta T}} \frac{\partial J(T, q)}{\partial q}. \]

Each of the three phases has to be analyzed separately, taking due notice of the fact that the functions \( S(t) \) and \( u(t) \) are continuous within each phase but that, at points of time \( t_i \) and \( t_f \), \( S(t) \) has different RHS and LHS derivatives and \( u(t) \) is discontinuous.

Within Phases A and C, maintenance is nil and the marginal cost equals service degradation. For the instant \( t_i \), then:
\[ \frac{\partial J}{\partial q} = \int_{t_i}^{t_f} \left[ \alpha S(t, q)^{-\lambda} + \alpha q \frac{\partial S(t, q)^{-\lambda}}{\partial q} \right] e^{-\beta t} dt, \]
with a similar relationship for the instant \( t_f \).

By contrast, in the central cruising Phase B, \( \partial J/\partial q \) equals:
\[ \int_{t_i}^{t_f} \alpha S(t, q)^{-\lambda} e^{-\beta t} dt + \int_{t_i}^{t_f} \alpha q \frac{\partial S(t, q)^{-\lambda}}{\partial q} e^{-\beta t} dt - \int_{t_i}^{t_f} \frac{\partial u(t, q)}{\partial q} e^{-\beta t} dt - u(t_i, q) \frac{\partial t_i}{\partial q} + u(t_f, q) \frac{\partial t_f}{\partial q}, \]
or
\[ \frac{\partial J}{\partial q} = \int_{0}^{T} g(S(t)) e^{-\beta t} dt - \int_{0}^{T} \frac{\partial u(t)}{\partial q} e^{-\beta t} dt - \int_{0}^{T} g(S(t)) e^{-\beta t} dt - \int_{0}^{T} g(S(t)) e^{-\beta t} dt. \]

This expression, which easily leads to the calculation of the toll \( \pi \), does comprise \( \partial u/\partial q \), the core component found in standard literature, but additional intertemporal effects now matter. Current traffic leads not only to an increase in contemporaneous maintenance but also to higher maintenance in later years. Also, there is a supply side effect in the sense that increased traffic at the margin leads the operator to improve service, an effort that benefits all other users: it is a positive externality resembling the Mohring (1972) effect for service frequency.

In the initial and final phases, however, the null maintenance cost does not adequately reflect the marginal social cost because service quality is degraded, a nuisance that must be

\footnote{This last term corrects the fact that the variation in \( J \) comprises the gain in service quality by the marginal user who benefits from it but which cannot be counted as an external cost.}
accounted for. Furthermore, due to the discontinuities in \( t_i \) and \( t_f \), two additional terms appear, namely the last ones in Equation (20). As we will see below, all terms in this equation can be estimated through statistical procedures, except the last two. But, using simulations based on sensible parameters like those presented in Section 4 below, it is possible to conjecture that they are of a lower magnitude than the rest of the relationship.

There are now considerations reaching beyond those of the simple current infrastructure maintenance cost, and which may go either way, depending presumably on each case. Moreover, it is now both unadvisable and impossible to split the price in order to distinguish the parts assigned respectively to current maintenance and to regeneration. And the difference between former and new prices can be significant, as simulation examples of Table 1 indicate. For low traffic levels, former charges are too high: as stated previously, this is a sort of Mohring effect in that higher traffic incites the manager to improve service quality, a motivation that progressively decreases.

### Table 1. Revenue differences between standard and optimal intertemporal pricing rules

<table>
<thead>
<tr>
<th>Traffic</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
</tr>
<tr>
<td>300</td>
</tr>
<tr>
<td>200</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Revenue from standard marginal charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.509</td>
</tr>
<tr>
<td>1.080</td>
</tr>
<tr>
<td>0.583</td>
</tr>
<tr>
<td>0.173</td>
</tr>
<tr>
<td>0.057</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Revenue from new optimal charge</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.630</td>
</tr>
<tr>
<td>1.140</td>
</tr>
<tr>
<td>0.588</td>
</tr>
<tr>
<td>0.155</td>
</tr>
<tr>
<td>0.038</td>
</tr>
</tbody>
</table>

Note that the optimal charge now includes the anticipated effect of future cumulative traffic: supplementary current traffic increases not only current expenses but also those of all forthcoming years because it damages the way and makes it susceptible to higher maintenance charges in the future.

### 4. Effects of uncertainty on maintenance policy

It has been assumed in the previous sections that there is no uncertainty. But the maintenance process is in fact quite uncertain, especially concerning the link between traffic and resulting damages. To take this fact into account, we adopt the classical assumption that the evolution of the system depends on a Wiener-like random variable process in accordance with:

\[
dS(t) = h(K, Q, t)u(t) - f(K, Q, q, t)dt + \sigma dz
\]

where the random variable \( dz \) denotes a classical Brownian motion, as in Haussmann & Suo (1995a). We will not analytically solve this optimisation problem but simply perform numerical simulations using the usual Hamilton-Jacobi-Bellman (HJB) equation with partial derivatives:

\[
-J_t = \max_u [-u(t) - \alpha q e^{-\lambda S(t)}] e^{-\mu t} + J_S [h(K, Q, t)u(t) - f(K, Q, q, t)] + \sigma^2 J_{ss}
\]

Note that such a simulation procedure, with parameters and functions found in Frame 2, and (5) specified as \( g(S) = -\alpha e^{-\lambda S} \) instead of \( g(S) = -\alpha S^{-\lambda} \), could have been used to solve the certain case obtained by setting the standard deviation \( \sigma \) to zero: we have consequently compared the results of simulations to the solution obtained analytically by the Pontryagin procedure. The solution values, shown in Figure 3, are drawn from a large number of simulations providing similar results. Those are based on sensible values of the parameters corresponding to the average track segment of the French network and on simple specifications of the functions intervening in the model, as documented in Appendix 2.
Frame 1. Parameters and functions used for the simulations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of years between 2 renewals: ( T )</td>
<td>( T=20 )</td>
</tr>
<tr>
<td>Initial quality of service: ( S(0) )</td>
<td>( S(0)=3,3 )</td>
</tr>
<tr>
<td>Discount rate: ( j )</td>
<td>( j=0,04 )</td>
</tr>
<tr>
<td>Functions for: ( dS(t) = h(K,Q,t)[u(t) - f(K,Q,q,t)]dt + \sigma dz )</td>
<td>( h(t)=7; )</td>
</tr>
<tr>
<td>Limit value of current maintenance ( u(t) )</td>
<td>( 0 )</td>
</tr>
<tr>
<td>Yearly traffic normalized to: ( q )</td>
<td>( q=100 )</td>
</tr>
</tbody>
</table>

Note that technical parameter values denoted by \( K \) earlier are taken into account in the values selected.

Figure 3. Analytically derived Pontryagin and Hamilton-Jacobi-Bellman simulation results

<table>
<thead>
<tr>
<th>Case</th>
<th>A. Analytical Pontryagin result</th>
<th>B. HJB simulation ( \sigma=0 )</th>
<th>C. HJB simulation ( \sigma=0,1 )</th>
<th>D. HJB simulation ( \sigma=0,2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target service quality ( S )</td>
<td>0,976996</td>
<td>0,98</td>
<td>0,97</td>
<td>0,96</td>
</tr>
<tr>
<td>( J ) integral</td>
<td>-0,256568</td>
<td>-0,259 +/-0,0015</td>
<td>-0,261 +/-0,003</td>
<td>-0,263 +/-0,005</td>
</tr>
<tr>
<td>Marginal cost: ( dJ/dq )</td>
<td>0,000919</td>
<td>0,0006 +/-0,0002</td>
<td>0,0006 +/-0,00044</td>
<td>0,0010 +/-0,00007</td>
</tr>
</tbody>
</table>

Target trajectories of the quality of service \( S(0,\ldots,1) \) over 20 years (1,\ldots,20)

The following conclusions may be drawn from Figure 3 findings:

- the simulation procedure brings out a critical path for the quality of service, a target trajectory: as soon as the random variable provides a gap between actual and goal trajectories, the optimal policy consists in eliminating the gap as quickly as possible;
- in the certainty case \( \sigma=0 \), the target trajectory provided by the HJB simulation is the same, due account taken of computing accuracy, as the cruising trajectory which can be analytically found in the certainty situation by Pontryagin maximum principle procedures;
- when the standard deviation of the random Brownian motion differs from zero, the target trajectory is slightly below the « certainty » target trajectory and the moment when the optimal maintenance is set at zero happens sooner. But for sensible values of the parameters, the difference with the certainty situation is small;
- it turns out that, at each point of time, the optimal policy is to make either \( \theta \) (the minimum) or \( m \) (the maximum) current maintenance, depending on whether the quality of service is above or below the cruising level. But averaged over a sufficiently long period, the implied optimal value of maintenance appears “close” to the maintenance level of the certain situation, this closeness increasing as the standard deviation approaches the certainty level;
- marginal costs calculated with uncertainty are of the same magnitude as the marginal costs obtained in the “certainty” situation, for sensible values of the standard deviation.
5. Econometric and calibration tests

Our model, taking regeneration and service level provision into account, implies the existence of the three Service performance and current Maintenance phases shown in Figure 2. But we have found no evidence of the existence of such phases in the rare literature on track performance provision\(^8\) by rail firms (e.g. Martland, 1992; Robert et al., 1997) or in the literature on infrastructure maintenance cost models estimated from track section data\(^9\).

Since the first track segment-level analyses by Idström (2002), Johansson & Nilsson (2002, 2004) and Gaudry & Quinet (2003), such models, all based on short-run maintenance cost minimization assumptions, are nested into (9) by the removal of three key variables: elapsed time \(t\) as well as cumulative traffic \(W\) since the last regeneration, and Target Service \(S_r\).

These simplifications lead explicitly for each segment of a given sample to the specification:

\[
u_t = f(K(h_0, e_0), q(n,w), S(h)),
\]

where the index \(t\) refers to a few consecutive years, typically 1 to 3. This brevity of large track segment panels makes it possible, if not necessary, for the above early and for later authors to abstract from input prices which are assumed constant in space (and during the very short time periods available) for the given national infrastructure provider studied, for instance Swedish in Andersson (2006), Swiss in Marti & Neuenschwander (2006) and Austrian in Link (2009) —all three of which use CES (Log-Log) specifications for \(f(\cdot)\).

Note that classical formulation (23-A), based on short run cost minimization, flushes out variables only summarily defined above and earlier. First, technical characteristics \(K\) are split between planned reference quality or “standing” \(h_0\) (such as maximum allowed design speed \(v\) and UIC group status \(g_k\) used below), current deterioration performance indicators \(S(h)\) (such as age of rails and of sleepers (ties); proportion of wood sleepers; current speed restrictions; etc.) and state characteristics \(e_0\) (such as length; circuitousness; number of tracks, switches and tunnels; electrification; centralized automated traffic control, etc.) of the segment. Second, traffic \(q\) or \(Q\) is explicitly defined by measures of the number of circulating trains \((n\ or\ N)\) and of their weight \((w\ or\ W)\).

Nor could we find reference to three phases in the unique example of a road maintenance cost model (Ben-Akiva & Gopinath, 1995) where explanatory variables do include cumulative traffic \(W\), time elapsed since the last regeneration \(t\), and a current degradation performance\(^10\) indicator \(S(h)\). That “extended classical” specification nests into:

\[
u_t = f(K(h_0, e_0), q(n,w), Q(N,W), t, S(h)),
\]

where the “new” Target Service and Trajectory Correction terms, key features of the joint optimization approach included in forthcoming equation (23-C) of Section 5.3, are missing.

---

\(^8\) There of course exists a literature on the supply of vehicle capacity or seat-km services, notably in urban transit systems where Mohring’s (1972) “frequency effect” due to fixed-capacity vehicles arises, as well as a literature on the recursive nature of transit seat Demand-Supply equation systems estimated from monthly or yearly data (e.g. Gaudry, 1980), but such study streams never deal with the supply of infrastructure quality or with its degradation performance. And literature dealing with railroad and road infrastructure provision itself tends to explain only aggregate quantity measures (e.g. Ingram & Liu, 1997) and avoids the issue of quality supply.

\(^9\) These European models of rail maintenance cost estimated solely from data by track segment should not be confused with models of total rail cost estimated from aggregate firm-wide time-series data, frequently estimated with CES and Trans-Log specifications during the previous two decades, or even with models of maintenance cost estimated from similarly aggregate firm-wide time-series data (e.g. Bereskin, 2000), where the assumption of constant input prices obviously cannot be made due to the length of the time-series used.

\(^10\) In their paper, \(h_0\) is the structural pavement number and \(S(h)\) is a latent variable linearly estimated [viz. their Equation (29)] from a set of condition indicators of current degradation. There are no phases or target levels.
5.1. Data and estimation strategy

Specifics of a multi-phase model. The potential existence of phases and of a Target Service level have empirical implications. In this respect, finding out how much the addition of a Target Service level variable $S_c$ contributes to the explanation of current Maintenance costs will require only the addition of a couple of new terms pertaining to $S_c$, a minor modification of maintenance cost equation specifications (23-A) or (23-B); but testing the existence of three phases does require unprecedented reformulations of rail firm behavioral equations.

In particular, phases imply that Maintenance expenditures may follow an asymmetric $\cap$-shaped time path, different even from a quadratic path if Figure 2.B is adequate, a functional form issue addressed presently as the first fundamental implication of the theory. This implication matters because the calculation of representative marginal costs should exclude from the sample observations pertaining to the short phases at the extremities of interval $T_i$ and retain only those associated with the long central monotonic cruising phase B.

**Data base use strategy.** Starting then with Maintenance phases, can $\cap$-shaped forms be detected in our two pruned databases for the years 1999 and 2007? Moving on afterwards to the analysis of Service levels, the absence of information on this variable in 1999 restricts our estimation possibilities to 2007 data. We adopt the estimation steps indicated in Table 2.

<table>
<thead>
<tr>
<th>Table 2. Estimation sequence for Maintenance and Service equations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current maintenance cost</strong></td>
</tr>
<tr>
<td>1999 Base</td>
</tr>
<tr>
<td>a. Test existence of 3 phases on full sample (978 obs.).</td>
</tr>
<tr>
<td>Find Phase A &amp; C segments.</td>
</tr>
<tr>
<td>b. Retain cruising Phase B censored subsample</td>
</tr>
<tr>
<td>2007 Base</td>
</tr>
<tr>
<td>a. Test existence of 3 phases on full sample (700 obs.).</td>
</tr>
<tr>
<td>Find Phase A &amp; C segments.</td>
</tr>
<tr>
<td>b. Retain cruising Phase B censored subsample (675 obs.).</td>
</tr>
<tr>
<td>c. Estimate cruising Phase B Maintenance Cost $u$ equation</td>
</tr>
<tr>
<td>making use of variables based on Target Service $S_c$</td>
</tr>
<tr>
<td>estimates (580 observations)</td>
</tr>
<tr>
<td><strong>Service provision</strong></td>
</tr>
<tr>
<td>2007 Base</td>
</tr>
<tr>
<td>a. Test existence of 3 phases on full sample (700 obs.).</td>
</tr>
<tr>
<td>Find Phase A &amp; C segments.</td>
</tr>
<tr>
<td>b. Retain cruising Phase B censored subsample (675 obs.).</td>
</tr>
<tr>
<td>c. Estimate cruising Phase B Service Supply $\Delta S$ estimate</td>
</tr>
<tr>
<td>using $S_c$ and $u$ estimates</td>
</tr>
<tr>
<td>Estimate embedded Target Service $S_c$ (580 obs./year)</td>
</tr>
</tbody>
</table>

In an analysis of pruned databases for 1999 and 2007, we first show that high standing track segments belonging to UIC$^{11}$ groups 2 to 6 in fact still contain three Maintenance cost phases, as demonstrated by the asymmetric $\cap$-shaped nature of expenditure profiles over time: this is established by a two-step (a, b) process whereby removal of observations from Phases A and C leaves only monotonic central Phase B subsamples. Second, with such doubly censored cruising phase subsamples, we estimate Maintenance cost ($u$) and realized Service provision ($S_c$) equations for the year 2007 (in each case Step c in Table 2) after a preliminary analysis of embedded Target Service ($S_c$) to be explicated in Section 5.3.

**Available information.** The available variables listed in Table 3 originate from the full French network of about 30 000 km but sample sizes differ by year due to pruning of the raw

---

$^{11}$ The 13 UIC (*Union Internationale des Chemins de Fer*, 1989) groups used by SNCF are measures of standing. Our retained subsets of observations exclude Group 1 segments found at the entrance of the largest stations, include Groups 2-6 segments for which track is periodically regenerated and remaining Groups 7-9 segments that are in principle never regenerated (except on an *ad hoc* basis). The latter are further split as between those “with passenger trains”, labeled 7-9 here, and others classified as “without passenger trains”, labeled 10-12, a SNCF assignment that does not mean that passenger traffic is always necessarily nil.
This difference in segment length shows up clearly in Figure 4 where maximum allowed speed \( v_{\text{ma}} \), the main initial infrastructure quality indicator \( h_0 \) in (23-B), is plotted by track segment classified by UIC group membership. Actual maximum speeds are always a multiple of 10 km/h on physical “traffic” segments (whence, 1999 dots stand for many segments) but its average value calculated over longer administrative “cost” segments blurs this structure and gives the impression that speed maxima for train drivers are more varied and continuously defined than they in fact are on the shorter physical traffic segments.

**Figure 4. Maximum allowed speed (km/h) on track segments classified by UIC group**

<table>
<thead>
<tr>
<th>A. 1999 (985 observations)</th>
<th>B. 2007 (700 observations)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="1999 speed distribution" /></td>
<td><img src="image2" alt="2007 speed distribution" /></td>
</tr>
</tbody>
</table>

---

1. Maximum allowed speed for train drivers is always a multiple of 10 km/h on physical “traffic” segments (whence, 1999 dots stand for many segments) but its average value calculated over longer administrative “cost” segments blurs this structure and gives the impression that speed maxima for train drivers are more varied and continuously defined than they in fact are on the shorter physical traffic segments.
Econometric strategy. The general econometric specification adopted, readily applicable to our cross-sectional data sets on segments \((I, \ldots, t, \ldots, T)\), is the Box-Cox model with a non-spherical distribution of residuals:

\[
\begin{align*}
y_t^{(\lambda)} &= \beta_0 + \sum_i \beta_i X_i^{(\lambda)} + u_t, \\
u_t &= \exp\left(\sum_m \delta_m Z_{im}^{(\lambda_m)}\right)^{1/2}, \\
v_t &= \sum_{i=2}^{T} \rho_{i,1} \left(\sum_{n=1}^{T} \tilde{R}_{i,m} v_n\right) + w_t.
\end{align*}
\]

where, in the first two equations, the Box-Cox transformations (BCT) applicable to any strictly positive\(^{12}\) variable \(\text{Var}_i\) is commonly defined without Tukey’s shift parameter as:

\[
\text{Var}^{(\lambda)} = \begin{cases} 
\left(1/\lambda_i - 1\right)/\lambda_i, & \lambda_i \neq 0, \\
\ln\left(\text{Var}_i\right), & \lambda_i \rightarrow 0,
\end{cases}
\]

and where, in the third equation, \(\tilde{R}_{i,m}\) denotes the typical element of matrix \(\tilde{R}_i\), a row or column-normalized square \((T \times T)\) Boolean matrix \(R_i\) expressing hypothesis \(\ell\) concerning the presence of correlation among designated residuals, which may then behave as substitutes \((\rho_i > 0)\) or complements \((\rho_i < 0)\) to the residual of segment \(t\). And rewriting (24-C) in matrix notation eases the explicitation of matrix \(\tilde{R}_i\), namely

\[
v = \sum_{i=2}^{T} \rho_{i,1} \tilde{R}_i v + w, \text{ with } \tilde{R}_i = \pi_i \left[I - (1 - \pi_i)\tilde{R}_i\right]^{-1} \tilde{R}_i, \quad (0 < \pi_i \leq 1),
\]

where the new proximity parameter \(\pi_i\) measures the relative influence of “near” and “distant” neighbors of \(R_i\), as soon discussed below.

BCT forms. Note that (24-A) contains the necessary intercept (Schlesselman, 1971) and that the \(X_i\) denote any non-dummy term, including a product of variables; in addition, a special case used in Phase tests below consists in using the cumulative traffic variable \(W\) twice:

\[
Q(W_t) = \beta_0 W^{(\lambda_0)} + \beta_1 W^{(\lambda_1)} + \beta_2 W^{(\lambda_2)},
\]

in which case the BCT are distinct \([\lambda_1 \neq \lambda_2]\), the model is identified in terms of transformed variables and the usual alternating signs conditions on \(\beta_1\) and \(\beta_2\), deciding if and whether a maximum or a minimum occurs with a quadratic specification, generalize to those listed in Table 4, from Gaudry et al. (2000), where a given sequence of signs yields a maximum or a minimum depending on the sign of the difference \([\lambda_1 - \lambda_2]\). Phase tests below will not focus on the most general asymmetric turning case \([\lambda_1 \neq \lambda_2]\) but on a particular one \([\lambda_1 = 1; \lambda_2 = 2]\) and on the proper symmetric quadratic \([\lambda_1 = 1; \lambda_2 = 2]\) special case, both nested in (24-F).

<table>
<thead>
<tr>
<th>\text{CASE}</th>
<th>\beta_1</th>
<th>\beta_2</th>
<th>\lambda_1 - \lambda_2</th>
<th>\beta_1(\lambda_1 - \lambda_2) \text{ or } \beta_2(\lambda_2 - \lambda_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\cap)</td>
<td>Maximum 1</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(\cup)</td>
<td>Minimum 1</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>(\cap)</td>
<td>Maximum 2</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(\cup)</td>
<td>Minimum 2</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^{12}\) The dummy variable and 0-replacement methods available to transform variables that contain some 0 are discussed and used in Gaudry & Quinet (2010). Here 0 traffic values are replaced by 0.00001, unless stated otherwise, as for instance in the detection of Phase C segments where the dummy variable method is used.
**Heteroskedasticity.** To obtain both homoskedastic errors \( v_i \) or \( w_i \), and correct BCT forms, particularly for the dependent variable \( y_i \), it is necessary to specify a second instrument for the former of these targets. It is provided by (24-B) where, following Gaudry & Dagenais (1979), the \( Z_m \) variables may also be \( X_k \) variables and the formulation, which necessarily yields positive variances, is flexible enough to include classical heteroskedasticity as a special case if all \( \delta_m \) but one are set at 0 and that remaining one equals 2 (with the matching \( \lambda_m = 0 \)). In our forthcoming tests on Maintenance cost \( (u_i) \), where the optimal BCT on dependent variable is in the vicinity of 0.25, heteroskedasticity in various garbs, classical and more general, was never found: the presentation of results henceforth ignores the issue.

**Directed autocorrelation.** Matrix \( \hat{R} \) in (24-E) results from three steps. In the first, a square matrix \( R \) is defined with typical element \( r_{i,j} = 1 \), to express an hypothesized correlation between any two residuals \( v_i \) and \( v_j \) (and \( r_{i,i} = 0 \) otherwise). It is often called a contiguity matrix because spatial hypotheses were first used to define it; but clearly any criterion, spatial, temporal or socio-economic, may specify a pattern of interdependence among residuals: the researcher’s directed choice\(^{13}\) is not limited by the natural order of the data.

In the second step, it has long been the practice (Ord, 1975) to row or column normalize this matrix and Bolduc (1987) has shown that the resulting normalized matrix \( \hat{R} \) guarantees a convex likelihood function over the stable unit interval of \( \rho \). This formulation is « idiot-proof » in the sense that, based only on zeroes and ones, it avoids the complications of lack of invariance of estimates to changes in units of measurement that arise when the elements of \( R \) are functions of continuous variables, such as distance or income (Bolduc et al., 1989, p. 369). But something might then be needed to compensate for the discreteness of \( R \), as many distributed phenomena are likely to be smooth and their representation by a few « all-or-nothing » slices insufficient: the solution resides in taking due account of the whole set of near and distant “neighbor” slices, as long done with distributed lags in time-series analysis.

**Distributed “spatial” lags.** Following the Blum et al. (1995/1996) approach, powers of \( \hat{R} \) generate a sequence of contiguity matrices \( \{ \hat{R}^2, \ldots, \hat{R}^c, \ldots \} \) which define degrees of neighborliness or proximity, with \( \hat{R}^2 \) denoting neighbors of neighbors, and \( \hat{R}^c \) higher powers. Under the assumption that the impact of these close and distant neighbors decreases geometrically with «distance» \( c \), as in Koyck (1954) distributed lags of time-series (whence the name Autoregressive Contiguous Distributed (AR-C-D) for this analogue process), one obtains \( \hat{R} = \pi_i \left[ I - (1 - \pi_i) \hat{R} \right]^{-1} \hat{R}, \) where the new proximity parameter \( \pi_i \) allows for endogenization of the relative importance of near and distant effects. If \( \pi_i = 1 \), \( \hat{R} = \hat{R} \), in which case only the adjacent neighbors have an impact on the correlations among the residuals selected by the “residue impact criterion” \( R \), exactly as in the classical (Ord, 1975; Cliff & Ord, 1981) case. By contrast, as \( \pi_i \rightarrow 0 \), the near effect is reduced to a minimum in favor of the distant effect. In this R-Koyck formulation, the parameter \( \pi_i \) therefore generally weights the relative importance of near and distant effects, i.e. the sharpness or slope of the decline. Using the L-2.1 algorithm (Tran & Gaudry, 2008), we will test below, in our analysis of Maintenance cost \( (u_i) \) by segment, three specifications of the \( R \) matrix and, singly and jointly, two distributed processes \( \pi_i \) as well.

\(^{13}\) For an example of directed choice involving correlation of residuals across socio-economic groups, see Gaudry & Blum (1988). For distributed processes with transport and trade flow models, see Gaudry (2004).
Log likelihood function and measure of fit. As discussed at length in Dagenais et al. (1987), the proper statistical model to use for problems like ours, where \( y_t \) can be said to have both a lower limit (e.g. maintenance cost cannot be negative but can be close to 0, implying the possibility of a mass point at \(-1/\lambda\) when \( \lambda > 0 \), or in practice at a very small number \( \varepsilon \)) and an upper limit (maintenance cost cannot exceed the periodic reconstruction cost, called “renewal cost”), is the Rosett & Nelson (1975) two-limit Tobit model where the likelihood of observing the vector \( y \equiv (y_1, \ldots , y_n, \ldots , y_T) \) reduces, in the absence of limit observations in the sample, to the maximand \( \Lambda \) actually proposed by Box and Cox (1964) themselves:

\[
\Lambda = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi \sigma^2_w}} \exp \left( -\frac{w_i^2}{2\sigma^2_w} \right) \mid \frac{\partial w_i}{\partial y_i} \mid,
\]

where, under the assumptions of normality and constancy of the variance \( \sigma^2_w \) of the independent error term \( w_i \) of zero mean, the Jacobian of the transformation from \( w_i \) to the observed \( y_i \) is \( \left| \frac{\partial w_i}{\partial y_i} \right| = y_i^{\varepsilon_{i-1}} \). With multiple BCT in (24-A), finding the global maximum of \( \ln(\Lambda) \) is required because its concavity (Kouider & Chen, 1995) need not hold.

But, even in the absence of limit observations in our final doubly censored samples, any observation \( y_i \) in (24-A) is still assumed to be censored both downwards and upwards, within a range \( \varepsilon \leq y_i \leq \nu \), where \( \varepsilon \) and \( \nu \) are respectively the strictly positive lower and upper censoring points common to all observations. In such and forthcoming generalizations of the Tobit model to a doubly censored dependent variable, the expected value of \( y_i \), given that \( \phi(w) \) is the normal density function of \( w \) with 0 mean and variance \( \sigma^2_w \), is given by:

\[
E(y_i) = \varepsilon \int_{-\infty}^{\nu_{i}(\varepsilon)} \phi(u)du + \int_{\varepsilon_{i}(\varepsilon)}^{\nu_{i}(\nu)} y_i \phi(w)dw + \nu \int_{\nu_{i}(\nu)}^{\infty} \phi(w)dw, \quad (\varepsilon \text{ and } \nu > 0).
\]

Elasticities. In all forthcoming tables of results, we report on elasticities of the dependent variable and for brevity neglect the regression coefficients which have lost their intuitive interpretation in BCT models, but we present their \( t \)-statistics computed conditionally upon the value of BCT estimates (Spitzer, 1984). Although the L-1.4 and L-2.1 algorithms used both compute the elasticity of \( E(y) \) defined in (24-H) with respect to all explanatory variables (including the \( Z_m \) if heteroskedasticity is present), we prefer instead, for comparability with published results, to list here the so-called “sample” measure, defined in Dagenais et al. (1987, pp. 464-467) for standard and Boolean dummy variables \( X_i \) and \( X_d \), as:

\[
\eta_i( y, X_k) = \left. \frac{\partial y}{\partial X_k} \right|_y = \beta_{X_k} \frac{X_k^{\varepsilon_{i-1}}}{y^{\varepsilon_{i-1}}},
\]

\[
\eta_i( y, X_d) = \beta_{X_d} \frac{X_d^{\varepsilon_{i-1}}}{y^{\varepsilon_{i-1}}},
\]

where the vertical line means that the expression (for our homoskedastic cases) is “evaluated at \( t \)” (at sample means, in all of our tables) and \( \bar{X}_d \) denotes the sample mean calculated over non-zero observations. Expression (24-J) provides an acceptable estimate of the percentage change in the dependent variable resulting from the presence of the dummy variable, i.e. an “elasticity” measure of the response resulting from the addition of the dummy variable.

---

14 The original Tobin (1958) paper had limits that varied by observation. Here it is assumed that current Maintenance expenditures \( u_t \) include at least minimum surveillance and cannot be larger than the Regeneration cost, and that technically defined Service measure \( S_t \) is limited between a minimum contractually defined threshold level which triggers immediate corrective action and a maximum provided after regeneration.

15 The full results from which tables 6, 7, 9, 10, 11 and 13 are selected are available from the authors.
5.2. Are there three phases in current rail Maintenance databases?

We first establish the existence of phases in Maintenance expense time profiles\textsuperscript{16} using specification (23-B) —which is devoid of Service Target Level and Correction variables.

Definitions of variables. The principal variables used in the tests are listed in the summary specifications of Table 5 where the original cumulative traffic variable $W$ is split between UIC groups to distinguish between the higher groups 2-6 which are regenerated periodically and the lower ones 7-9 which are never regenerated\textsuperscript{17}. In terms of measurement, note also that $t$, time elapsed since the last regeneration, is not observed directly, in contrast with the key initial standing variable \textit{maximum authorized speed}, but only approached by the \textit{average age of rails} on the segment. Note finally that cumulative traffic $W$ is constructed by simple multiplication of time by current traffic in 1999 but is observed\textsuperscript{18} in the 2007 sample.

Specification. All phase form tests pertaining to cumulative traffic variable $W$ neglect observations on “new line” (TGV-only) segments, somewhat heterogeneous with the classical line segments and few in number\textsuperscript{19}. Regression results found in Table 6 are abstracted\textsuperscript{20} from equation specifications using all variables listed in Table 5 except for total traffic $w$ and for UIC group dummy variables\textsuperscript{21} $g_k$, variables which will intervene later.

Table 5. Main variables used to establish the existence of three maintenance cost phases

<table>
<thead>
<tr>
<th>$u = f$</th>
<th>$K(e_0 ; h_0)$</th>
<th>$Q(W_{2.6}; W_{7.9})$</th>
<th>$q(w \ or \ w_c)$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>cost per km</td>
<td>max speed $v$; UIC $g_k$</td>
<td>${0(w)}_{2.6}$</td>
<td>${0(w)}_{7.9}$</td>
</tr>
<tr>
<td>2007</td>
<td>cost per km</td>
<td>max speed $v$; UIC $g_k$</td>
<td>${\sum_{i=1}^{7} w_i}_{2.6}$</td>
<td>${\sum_{i=1}^{7} w_i}_{7.9}$</td>
</tr>
<tr>
<td><strong>Column</strong></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Pruning and anticipations. The detection of symmetric or asymmetric $\cap$ shapes in the pruned databases depends on the residual presence of just renewed Phase A segments and of Phase C segments slated for renewal. The former is unlikely because the minimum value of the average age of rails is 4 years in both databases; the latter is more likely, due account taken of the pruning process used. For 1999, it excluded low traffic observations (such as traffic in gross tons that are not matched by exact circulations)\textsuperscript{22} but kept all high traffic segments; for 2007, it excluded apparent high traffic outliers\textsuperscript{23} but kept low traffic segments. We will soon see that, in spite of this double censoring, both samples still contain enough Phase C segments to require their removal if desired Phase B estimates, corresponding to step $c$ in Table 2, are to be demonstrably monotonic in the cumulative traffic variable $W$.

Results. Table 6 presents maximum likelihood results for parameters of selected variables included in regression phase tests carried out with the fully documented LEVEL 1.4 algorithm (Liem \textit{et al.}, 1993). Reported statistics include the total number of BCT and of $\beta_i$ regression coefficients of each regression variant. Columns A to C document models explaining \textit{total} current maintenance (the sum of surveillance and maintenance) expenses in 1999, and Columns D to E models pertaining only to \textit{surveillance}\textsuperscript{24} expenses in 2007.

---

\textsuperscript{16} Phase length tests on Service degradation (step $a$ in the third column of Table 2), will not be carried out.

\textsuperscript{17} See Appendix 1 to see how low traffic leads to infrastructure maintained indefinitely and never renewed.

\textsuperscript{18} The traffic is observed from 1995 until 2007 and reconstructed beforehand.

\textsuperscript{19} There are 18 such segments for 1999 and 17 for 2007.

\textsuperscript{20} Results for the 6 equations are not provided in full because results of more complete specifications of maintenance cost equations in accordance with 23-C will be supplied below in Tables 9 and 10.

\textsuperscript{21} Groups are defined by linear combinations of tons by train type and maximum allowed speed.

\textsuperscript{22} Data precision conventions used to register traffic in 1999 yielded many such cases for that year, which explains the absence of UIC group 10-12 segments in the sample (see Figure 4).

\textsuperscript{23} Such as two suburban traffic values 3 to 7 times the size of the third highest traffic.

\textsuperscript{24} Tests on the maintenance sub-total and on the total failed to detect turning forms, symmetric and asymmetric.
Results of reference models A and D assume monotonicity \( \beta_{q1} = 0 \) and those found in the next two columns respectively assume strictly quadratic \( \lambda_1 = 1; \lambda_2 = 2 \) and asymmetric turning \( \lambda_1 = 1; \lambda_2 \neq 2 \) special cases of (24-F). We note:

i) the existence of turning forms. The cumulative traffic variable in tons \( W \), broken up between \( W_{2:6} \) for segments belonging to UIC groups 2-6 and \( W_{7:9} \) for those belonging to groups 7-9, yields, for asymmetric and quadratic specifications, gains in fit which are very high and significant for total expenses (considering Log likelihood values for B or C relative to A)\(^{25} \) but very low and of marginal significance for surveillance expenses (considering Log likelihood values for E or F relative to D): in the latter sample, remaining Phase C observations are apparently relatively few.

For the same cumulative traffic variable \( W_{2:6} \), sign results of \( \beta \) and \( \lambda \) parameters documented in (24-F) correspond to those of Maximum1 in Table 4 for models B, C and E, but to those of Maximum2 for model F. Overall, differences in the statistical significance of 1999 and 2007 form results are consistent with anticipations based on the prior pruning;

ii) asymmetry of the \( \cap \) shapes. There is no significant difference between the turning shapes (models B and C) of total expenses in 1999; but, simply selecting the most likely specification on the basis of the highest Log Likelihood, the asymmetric shape model F is slightly preferable to the symmetric shape model E for surveillance expenses in 2007. Although these turning shapes are defined by and for all segments belonging to UIC groups 2-6, it is of interest to show in Figure 5 the evolution of the dependent variable for the most frequent\(^{26} \) segments, which belong to group 3.

Figure 5. Time profile of cost \( u \) with respect to cumulative traffic, UIC group 3 segments

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph A" /></td>
<td><img src="image2.png" alt="Graph B" /></td>
</tr>
</tbody>
</table>

Values shown in Figure 5 are not the proper expected values of the dependent variable \( E(y) \) obtainable from (24-H) but simply the effect of regression component (24-F) on the dependent variable calculated by \( y(W) = \{1 + \lambda_1 [Q(W) + \beta_0 + \sum \beta_i X_{i}^{(1)}] \}^{1/\lambda_1}, X_{i} \neq W, X_{i} = \bar{X}, i.e. \) by “unrolling” the transformed dependent variable \( y \) in (24-A), all other regressors having been set at their sample mean values\(^{27} \). In these ad hoc illustrations of asymmetric turning effects, the estimated value of the BCT on the dependent variable \( \lambda_{y} \) equals 0.28 in model C and 0.19 in model F. In Figure 5.A, one clearly detects a few high traffic segments for which regeneration is no doubt approaching; in Figure 5.B, the asymmetry of the \( \cap \)-shaped profile seems strong, but the regression results for surveillance expenses demonstrate that the existence of this turn is in fact fragile, as presently confirmed by Phase C censoring tests where the removal of as few as 6 of the highest traffic segments for this UIC group (and comparable numbers for the other groups) destroys it in favor of monotonicity;

iii) censoring Phase C observations. To isolate Phase C segments, we set at 0 the value of \( W \) for a subset of segments assumed close to renewal and we simultaneously add a matching Phase C

---

\(^{25} \) Further tests showed that one could always reject turning forms for UIC groups 7-9, as predicted by SNCF infrastructure managers and descriptions of infrastructure management practices (e.g. Rail Concept, 2006).

\(^{26} \) In the 2007 sample, the frequency by group is 28, 141, 123, 105 and 98, respectively, for the 495 segments belonging to groups 2-6.

\(^{27} \) For Figure B, the large negative constant was set at 0 in order to keep observations in the positive quadrant.
dummy variable to preserve the invariance of the estimates of the BCT applied to the modified W* variable now containing some zeroes. If and when the sufficient subset is found, estimation of (24-D) yields two BCT that converge to an identical value, at which point the transformed twin values of W* become collinear and the turning form reverts to a monotonic shape.

Such censored W* and associated Phase C dummy variables corresponding to the highest 5% of cumulative traffic values were constructed for 1999 (31 observations) and 2007 (25 observations). This first construct was found sufficient to reestablish monotonicity for 1999 but not for 2007 where a modified version of the 5% censoring rule, consisting in removing 4-6 of the highest traffic segments by UIC group (25 again in total), was adequate to identify Phase C segments. Detailed results with convergent BCT values in (24-F) naturally cannot be shown because the expression $X^* X'$, consisting in matrices of observations on $X_i$ variables transformed by BCT, is never invertible in the presence of exact collinearity.

<table>
<thead>
<tr>
<th>Table 6. Phase tests with cumulative traffic variables, classic line network, 1999 &amp; 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Specified form of $W_{2,6}$</strong></td>
</tr>
<tr>
<td>----------------------------------</td>
</tr>
<tr>
<td>$W_{2,6}$ (first term; first BCT)</td>
</tr>
<tr>
<td>$W_{2,6}$ (first term; first BCT)</td>
</tr>
<tr>
<td>$W_{2,6}$ (second term; second BCT)</td>
</tr>
<tr>
<td>$W_{2,6}$ (second term; second BCT)</td>
</tr>
<tr>
<td>$W_{2,6}$ (second term; second BCT)</td>
</tr>
<tr>
<td>Estimated form of $W_{2,6}$</td>
</tr>
<tr>
<td>Estimated form of $W_{2,6}$</td>
</tr>
<tr>
<td>Estimated form of $W_{2,6}$</td>
</tr>
<tr>
<td>Log likelihood</td>
</tr>
</tbody>
</table>

| Number of $\beta$ estimated | 20 | 21 | 21 | 17 | 18 | 18 |
| Number of $\lambda$ estimated | 8 | 7 | 8 | 8 | 7 | 8 |
| Difference in D.F. | 0 | 0 | 0 | 0 | 0 | 1 |
| Variant run number | 77 | 84 | 76 | 6 | 3 | 0 |

$\eta$: Elasticity of the dependent y variable with respect to $X$, calculated by (24-I).
$f^*$: Student’s $t$ statistic of $\beta$ coefficient with respect to 0, calculated conditionally upon the estimated value of $\lambda$.
$f^{**}$: Student’s $t$ statistic of $\lambda$ with respect to 0 and to 1, calculated unconditionally.
***: Numerical tests have determined that, due to the large size of the cumulative traffic variable $W_{2,6}$ and to a BCT value in Column C quite close to 2, the digits of the Log Likelihood are not distinct from those of Column B. But, if one replaces the 5 traffic variables by a single total traffic variable, the difference between the maximized Log Likelihood values of cases B and C is only 0.3 (variant runs 91 to 93 corresponding to A, B and C are not shown).

**Do three-phase cycles differ across UIC groups?** The residual existence of Phase C segments, certain in 1999 but of marginal significance in 2007, has then been established in both pruned databases with equation specifications containing the maximum number of distinct traffic by database, respectively 5 and 6, but no dummy variables for UIC groups.

---

28 Simply applying the BCT to the zero values, as sometimes carelessly done in total rail cost studies (e.g. Caves et al., 1980, 1985), will produce uninteresting estimates of the BCT that depend on units of measurement of the transformed variable.

29 Trials with 10% of observations yielded similar results, i.e. a return to monotonicity.

30 The modification consisted in choosing the 25 segments from the 4-6 highest cumulative traffic outliers within each UIC group. The original and modified choices yielded different numbers of censored observations by group, respectively (7;5) for UIC2, (14;6) for UIC3, (3;4) for UIC4, (0;5) for UIC5 and (1;5) for UIC6.

31 One in fact collapses back to specifications A and D with the censored $W_{i,n}$ variable replacing $W_{i,n}$ and with the addition of a matching Phase C dummy.
An interesting behavioral question is then whether the three-phase régime for groups 2-6 and the single phase régime for groups 7-9 are in fact specific to each group, i.e. whether UIC groups, rarely used by other European railroads, consist merely in administrative labels for management purposes or reflect distinct structural optimization and differences in standing. Adding 7 UIC \( g_t \) dummies to model C (where UIC group membership is reconstructed with some error) only increases the Log Likelihood by 10 points and adding 11 such dummies to model E (where UIC group membership is observed) by a mere 8 points \(^{32}\); this suggests that the optimization process is continuous over all segments and that any initial differences in treatment based on UIC standing are progressively erased by corrective maintenance.

For our 2007 database, Step b of Table 2 therefore consists in removing 25 Phase C observations in order to retain only central Phase B segments, leaving 675 observations \(^{33}\) for the analysis of cruising Service and Maintenance determination, to which we now turn.

### 5.3. Model teachings: cruising Maintenance and Service levels

Current Maintenance and Service equation specifications, the twin core endogenous determinations of our approach, should embody the uncertainty dimension proposed in Section 4 and their formulations should be given discrete garbs amenable to estimation from yearly data. Towards that end, all forthcoming specifications shall pertain to the cruising phase associated to the censored sample (excluding the extremities illustrated in Figure 2.A for service and 2.B for maintenance) and the running index \( t \) will refer to full years (with values measured for the year) but, for simplicity, the notation neglects subscripts for cruising phase B. Let us develop the key testable relationships of the model more formally than was just done to perform the phase tests of Table 6 with specification (23-B).

**The behavioural current Maintenance relationship.** We have seen that, during Phase B, maintenance is at each point of time either 0 or \( m \), depending on whether the quality of service is above or below the cruising level, but that on average maintenance \( u_i(t) \) is close to the level of the certain situation. We translate this last point in discrete time by extending (9):

\[
(23-C) \quad u_i = f(K,Q,q,t_i) + \frac{1}{h(K,Q,t_i)} [S_{c_i} - S_{c_{i-1}}] + \alpha(K,Q,t_i) \left[ S_{c_{i-1}} - S_{c_{i-2}} \right] + \epsilon_{u_i},
\]

where the “classical” first term stands for the damages caused by traffic and the two “new” terms bring in the extra maintenance \( [S_{c_i} - S_{c_{i-1}}] \), necessary to maintain the cruise trajectory of the quality of service in the absence of random effects, and the 0 or \( m \) correction \( [S_{c_{i-1}} - S_{c_{i-2}}] \) made by the infrastructure manager to revert to the target cruise trajectory in the presence of deviations from it. The new weight function \( \alpha(\cdot) \) might of course not differ much from \( h(\cdot) \) which, from (4), expresses the efficiency of a unit of maintenance expenditure on a segment: for the moment, both are assumed not to depend on current traffic \( q \).

**The behavioural Supply of Service relationship.** Expression (21) explaining how the quality of service varies in continuous time can be translated in discrete time as:

\[
(25) \quad S_i - S_{i-1} = h(K,Q,t_i) [u_i - f(K,Q,q,t_i)] + \epsilon_{2i},
\]

which, after replacement of \( u_i - f(K,Q,q,t_i) \) by its value from (23-C), may be written:

\[
(26) \quad S_i - S_{i-1} = S_{c_i} - S_{c_{i-1}} + \frac{a(K,Q,t_i)}{h(K,Q,t_i)} \left[ (S_{c_{i-1}} - S_{c_{i-2}}) + [\epsilon_{u_i}] \right] + \epsilon_{2i},
\]

where the random maintenance cost Surprise term \( [\epsilon_{u_i} = u_i - E(u_i)] \) can be estimated from (23-C) by using (24-H) to calculate \( E(u_i) \).

---

\(^{32}\) These variant runs, respectively numbered 72 and 14, are not shown but are available from the authors.

\(^{33}\) We will in fact use 580, for reasons given later.
The embedded Target Service relationship, a necessary preliminary. Note that both equations of interest (23-C) and (26) contain target service level \( S_c \) written from (8) as:

\[
\ln[S_c(t)] = \frac{1}{\lambda + 1} \left\{ \ln\left(\sum_{h=1}^{t} \alpha q_h\right) + \ln(\alpha \lambda) - \ln(h) + \ln \left[ \frac{j - h'}{h} \right] \right\}.
\]

Owing to its complexity, its substitution into (23-C) and (26) would impose intractable estimation constraints on the functions \( f(\cdot) \), \( h(\cdot) \) and \( a(\cdot) \) about which precious little is known except for the anticipated direction of some variations, e.g. \( f(\cdot) \) should be increasing in \( Q \), \( q \) and \( t \) but \( h(\cdot) \) might—or might not—be decreasing in \( Q \) and \( t \), as discussed in Frame 2. Also, effects this substitution per force for 2007 only (because maintenance expenses are only observed for that year) would neglect available information: data on \( S \) and on the \( q_i \) are available over many years (2000-2007) — until 2010 for \( S \).

We prefer to derive the Target Service \( S_c \) from a self-standing estimation of (27-A) as a third econometric relation: with \( S_c \) embedded in \( S_r \), fitted values \( E(S_c) \) and \( E(S_{c-1}) \) may then serve to construct required “new” variables \([S_c - S_{c-1}] \) and \([S_{c-1} - S_{c-2}] \) for (23-C) and (26).

5.4. Preliminary estimation of Phase B Target Service level \( S_c \)
An embedded structure. We wish to recover or extract the infrastructure manager’s target from data on actually provided service. It might first be asked how much track Service quality, measured here by subtracting from 10 the mean standard error of longitudinal deviations of rails from their desired axes\(^{34} \), i.e. \( S = (10 - NL) \), actually changes.

Figure 6.A shows that mean service quality \( S \), in a context of increasing rarity of public funds, has been quite variable over the 11 recent years from 2000 to 2010 and appears to have a minimum around 2004. This turning point coincides with a new awareness of the insufficiency of maintenance expenditures underlined in an audit report (Rivier & Puttalaz, 2005). Figures 6.B and 6.C also show that our measure of quality \( S \) degrades with time and UIC standing. Also, as such groups are largely based on daily (and therefore cumulative) tonnage \( W \), one expects \( S \) to be related to \( W \), but Figure 6.D shows no clear slope.

Specification: variables and form. To specify (27-A) for estimation with 7 years of data on 580 segments\(^{35} \), we momentarily neglect constants, acknowledge that regression coefficients are implicitly divided by \((\lambda + 1)\) and consider the components in turn:

- concerning \( g(\cdot) \), assume that social surplus depends on \( n_i \), the number of train circulations by category. Then, although the first R.H.S. term of (27-A) should also be of logarithmic form in accordance with the specification\(^{36} \) of (6), a myopic view of surplus can be approximated by

\(^{34} \) Subtracting the NL degradation measure from 10 yields Service \( S \) with a correlation of -1.00 to NL, the mean of standard errors of observations on longitudinal geometric deviations measured by segment zone of 200m. Such detailed NL readings, taken by special vehicles (called Mauzin, Matisa and Iris 320) travelling at the reference speed of the segment, are recognized as the best indicators of track quality and form the basis of various derived aggregate track performance indicators, such as the NL d’état which is subjected to contractual agreement between the (SNCF) infrastructure maintenance unit and the (RFF) infrastructure manager: excessive deteriorations automatically give rise to diagnostics and to corrective maintenance measures.

\(^{35} \) As noted above, the doubly censored database contains 675 observations for 2007 (including 2 for 2006), but only 583 for which Service is also measured every year from 2000 to 2007, as noted in Table 3. Reconstructing the average age of rails per segment for years previous to 2007 requires that the time series start in 2001 and that 3 segments with low average ages in 2001 be dropped, leaving 580 observations per year for 2001-2007.

\(^{36} \) Social surplus, defined as a sum of surpluses by train service type, differs from potential aggregate service indices, such as \( q_i = (q_1, ..., q_n)^{1/n} \) or \( q_i^* = \exp[\sum \alpha q_i / N] \), that would obviate the use of approximation by shares because the maximization would yield terms like \( (\alpha_1 \ln q_1 + ... + \alpha_n \ln q_n) / N \) or \( \sum \alpha_i (\ln q_i) / N \).
\[ \beta \ln(n_i) + \sum_{i=2}^{\infty} \beta_i s_{ij} \] if current circulation shares \( s_i \), including the reference one \( i=1 \), are stable across train categories. Replacement of such current measures of circulation by mean or trending values of circulations would instead imply a longer term view of social surplus influence;

* if \( h(\cdot) \), the productivity of a unit of maintenance expenditure, depends on cumulative train tonnage \( W \) and is defined with BCT as \( \exp \left[ \sum \beta_k X_k^{(i_k)} \right] \), \( \ln[h(K,Q,t)] \) is straightforward and \( h'(-)/h(-) \), then equal to \( \beta_k q/Q^{1-E} \), collapses to \( \beta_k q \) (with \( q \) specified as current tonnage \( w \)) under assumed linearity. The last term \( \ln(1-q/h) \) should then be close enough to \( h'/h \) if the latter is small relative to actualization rate \( j \).

The retained basic specification of embedded Target Service \( S_c \) is then, neglecting dummies:

\[ \ln(S_c) = \beta_0 + \beta_{xn} \ln(n_i) + \sum_{i=2}^{\infty} \beta_i s_{ij} + \sum_k \beta_k K_k^{(i_k)} + \beta_q W^{(i_q)} + \beta t^{(i_t)} + \beta_{iK} W_t + \epsilon_{it} , \]

the form estimation of which must maintain the theoretically mandated logarithmic and the assumed linear terms; but is not so restricted for all the terms in \( h(\cdot) \), amenable to BCT.

**Figure 6. Changes in Service quality, Phase B segments only**

**A. Mean track service quality evolution over time, 2000-2010 (594 segments)**

**B. Quality vs time (673 segments, 2007)**

**C. Quality vs UIC groups 2-12 (673 segments, 2007)**

**D. Quality S vs cumulative gross tonnage W (673 segments, 2007)**
More a calibration than an estimation, approach. The parameter estimates and results shown in Table 7 arise from a maximization of the Likelihood of \( S_i \) [with the due Jacobean of the transformation from \( \ln(S_i) \) to \( S_i \)] as if the errors were normally distributed, but without justifying this distributional assumption, using the same LEVEL 1.4 algorithm (Liem et al., 1993) employed for the Table 6 variants above.

Log Likelihood and \( t \)-statistics reported in Table 7 are therefore interpreted as numerical sensitivity measures of a BCT-enriched least squares calibration and should not be given too statistical an interpretation even if we carefully verified\(^{37}\) that the sample did not contain limit observations implying that the assumption of as-if normality for \( \varepsilon_i \) would be untenable.

**Sequence to establish optimal form.** In Table 7, the linearity assumption applied to all variables (excepting the total number of trains and the dependent variable, both in mandated logarithmic form) in Column 1 is progressively relaxed. Firstly in Column 2, by using for three technical variables (segment and track lengths, maximum speed) optimally determined logarithmic forms on the basis of detailed background BCT tests, which greatly improves the Log Likelihood. Secondly in Column 3, by the further application of BCT to \( W \) and \( t \), again yielding large additional gains in fit with optimal BCT power values of about 1.44 and 2.23, respectively. Column 4 simply adds to this last specification 6 yearly dummy variables which, quite significant as a group, may reflect the changing yearly availability of public funds available for maintenance, the reference year being 2001. The pattern of results for these dummy variables, found in Note 1 of Table 7, suggests that a minimum level of service was indeed reached around 2004-2005, as apparent in Figure 6.A.

**Results.** In addition to noticing these considerable gains in fit due to the application of 5 BCT to \( K \), \( Q \) and \( t \) terms of the component function \( h(\cdot) \), a number of remarks are in order:

i) **sensitivities:** as a built-in embedded structure, \( S_i \) should naturally vary little. The remarkably small values of the elasticities of \( S \) (all evaluated at sample means in conformity with (24-I)) with respect to even the most significant of the variables are fully consistent with this expectation;

ii) **signs:** it is appropriate to discuss signs for each component of (27-B) separately.

**In** \( g(\cdot) \): after noticing a regression coefficient sign change of the least significant variable, the *share of locomotive circulation*, all signs conform to expectations: the valuation of total service (total train circulations \( N \)) is positive and highly significant; in terms of mix, all circulation shares \( s_j \) are less valued than TGV circulations, taken here as reference.

**In** \( h(\cdot) \): differences in the productivity of a unit of maintenance expenditures with respect to technical state and standing variables \( K \) and \( v \) are reasonable. For instance, it is quite possible that, *maximum allowed speed* being accounted for (with a positive and very significant effect), expenditure on *suburban* and *high speed* lines be somewhat less productive in terms of service recovery than on classic non-suburban lines; concerning the *electrified lines*, there are also some small positive differences in productivity that do not contradict common sense. Without surprise, the *density of switches* lowers the productivity of a unit of current repair expenditure.

But for *traffic* (\( W \) and \( w \) in tons) and *time elapsed* variables \( Q \) and \( t \) in \( h(\cdot) \), the anticipated impact on the productivity of a unit of repair expenditure is uncertain. It is clear that current physical maintenance, such as tamping, postpones regeneration and becomes less effective over time, but this does not imply that a Euro of tamping becomes less productive between regenerations: it is not possible to tell from engineering knowledge concerning track longitudinal deterioration (*viz.* Fig.1), or from very similar graphs concerning track settlement over time (Ongaro & Iwnicki, 2009), whether increases in aging factors \( t \), \( Q \) and \( q \), the latter of which standing for \( h^2(\cdot)/h(\cdot) \),

\(^{37}\) This includes calculating the probability that the dependent variable be close to the lower limit \( e \) in (24-H).
increase or decrease (our stated expectation concerning Equation 4) the effectiveness of unit maintenance expenditure for given train circulations (N and its composition $s_i$).

Frame 2 compares the reference result of Table 7, Col. 3, to three other results obtained with alternate formulations and concludes that $t$, $Q$ and $q$ decrease it under all four specifications (despite some apparent differences in signs across the cases). This finding of a decreasing efficiency of repair expenses with respect to all aging factors implies that postponement of repairs beyond their due date raises maintenance cost.

**Frame 2. Impact of aging factors on the productivity of a unit of maintenance expenses**

Col. 3 below presents, extracted from the complete reference model found in Col. 3 of Table 7, results for the variables that cause wear and tear of tracks, namely time $t$ (the age of rails), cumulative traffic in tons $W$, and current traffic in tons $w$. The other columns are variants.

In Col. 3A, the BCT on $t$ and $W$ are set at the optimal values, respectively 1 and 0, previously found for those variables in Col. 3B and 3C variants where the traffic variables $W$ and $w$ are specified as levels, instead of densities as in 3 and 3A. Finally, in Col. 3C, train circulation variables (results not shown), also present in the specification (viz. Table 7) of target service, are not the current ones, as in the first three columns, but their mean (2001-2007) values.

To show that, for given train circulations in all four variants, increases in aging factors $t$, $W$ and $w$ always decrease the productivity of a Euro of maintenance expenditures despite some apparent differences in signs, remember first that $W = t \cdot w_{\text{reference year}}$ : indeed, the correlation between $W$ and the product of time $t$ by reference traffic $w_{2007}$ is 0.96 in Col. 3 and 0.93 in Col. 3A-3C. If aging $a$ is formally defined as the joint impact of those three variables, one can then write results for any column approximately as: $a = \beta_1 t^{(\lambda_1)} + \beta_2 (t,w)^{\lambda_2} + \beta_3 w^{(\lambda_3)}$, with derivatives:

$$\partial a / \partial t = \beta_1 \lambda_1 t^{\lambda_1 - 1} + \beta_2 \lambda_2 (t,w)^{\lambda_2 - 1} w$$

and

$$\partial a / \partial w = \beta_2 \lambda_2 (t,w)^{\lambda_2 - 1} t + \beta_3 \lambda_3 w^{\lambda_3 - 1}.$$

With the signed regression coefficient values of Col. 3, the actual derivatives:

$$\partial a / \partial t = \beta_1^{1.23} + \beta_2 (t,w)^{0.44} w$$

and

$$\partial a / \partial w = \beta_2 (t,w)^{0.44} t$$

are clearly always decreasing; and with those of Col. 3A to 3C, the actual derivatives:

$$\partial a / \partial t = \beta_2 / t$$

and

$$\partial a / \partial w = \beta_2 / w$$

are also always decreasing.

<table>
<thead>
<tr>
<th>Column</th>
<th>3</th>
<th>3A</th>
<th>3B</th>
<th>3C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_k$ coefficient, $t$-statistic, $\lambda$</td>
<td>$\beta_k$</td>
<td>$\lambda$</td>
<td>$\beta_k$</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>$t$</td>
<td>-.17E-04</td>
<td>2.23</td>
<td>.17E-02</td>
<td>1</td>
</tr>
<tr>
<td>(t=0)</td>
<td>(-17.88)</td>
<td>(7.89)</td>
<td>(-18.85)</td>
<td></td>
</tr>
<tr>
<td>(t=1)</td>
<td></td>
<td></td>
<td>[4.34]</td>
<td></td>
</tr>
<tr>
<td>$W$</td>
<td>-.11E-09</td>
<td>1.44</td>
<td>.17E-01</td>
<td>0</td>
</tr>
<tr>
<td>(t=0)</td>
<td>(-2.80)</td>
<td>(5.42)</td>
<td>(11.11)</td>
<td></td>
</tr>
<tr>
<td>(t=1)</td>
<td></td>
<td></td>
<td>[1.65]</td>
<td></td>
</tr>
<tr>
<td>$w$</td>
<td>-.29E-06</td>
<td>1</td>
<td>-.35E-06</td>
<td>1</td>
</tr>
<tr>
<td>(t=0)</td>
<td>(-1.08)</td>
<td></td>
<td>(-4.92)</td>
<td></td>
</tr>
<tr>
<td>(t=1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-1399.82</td>
<td>-1358.21</td>
<td>-1343.79</td>
<td>-1327.69</td>
</tr>
<tr>
<td>Belsley-Kuh-Welsch collinearity index</td>
<td>73</td>
<td>109</td>
<td>127</td>
<td>167</td>
</tr>
<tr>
<td>Variant run number</td>
<td>83</td>
<td>95</td>
<td>92</td>
<td>93</td>
</tr>
</tbody>
</table>

**Derivation of Target Service and Trajectory Correction variables.** In principle, all results of the last 3 columns of Table 7, where the successive use of 5 BCT has progressively changed the original form of variables appearing in the $h(\cdot)$ of Column 1, provide acceptable parameter estimates to construct from (24-H) components of the Target Service term $[E(S_{t_{i+1}}) - E(S_{t_{i+1}})]$ and of the Trajectory Correction term $[S_{t_{i+1}} - E(S_{t_{i+1}})]$, the “new” explanatory variables in forthcoming Maintenance cost and Service provision equations.
### Table 7. Embedded Target Service S. function results (4060 obs., 2001-2007)

<table>
<thead>
<tr>
<th>Column</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity $\eta(S), \lambda$, $t$-statistic* of $\beta_k$ coefficient</td>
<td>$\eta(S)$</td>
<td>$\lambda$</td>
<td>$\eta(S)$</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>$\beta_k$</td>
<td>Intercept</td>
<td>(t=0)</td>
<td>n.a.</td>
<td>(t=0)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>(334.96)</td>
<td>(174.17)</td>
<td>(180.63)</td>
<td>(176.66)</td>
</tr>
</tbody>
</table>

#### Social surplus evaluation component $g(\cdot)$

<table>
<thead>
<tr>
<th>$n$</th>
<th>Total number of trains</th>
<th>(t=0)</th>
<th>0.021</th>
<th>(31.38)</th>
<th>0</th>
<th>0.020</th>
<th>(29.25)</th>
<th>0</th>
<th>0.020</th>
<th>(29.38)</th>
<th>0</th>
<th>0.020</th>
<th>(30.22)</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Long distance GL train share (ref.: TGV)</td>
<td>(t=0)</td>
<td>-0.006</td>
<td>(-3.68)</td>
<td>-0.007</td>
<td>(-4.46)</td>
<td>-0.007</td>
<td>(-4.47)</td>
<td>-0.007</td>
<td>(-4.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Regional TER train share (ref.: TGV)</td>
<td>(t=0)</td>
<td>-0.003</td>
<td>(-4.02)</td>
<td>-0.004</td>
<td>(-4.79)</td>
<td>-0.004</td>
<td>(-4.73)</td>
<td>-0.006</td>
<td>(-7.80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ile-de-France regional train share (ref.: TGV)</td>
<td>(t=0)</td>
<td>-0.000</td>
<td>(-0.97)</td>
<td>-0.000</td>
<td>(-1.53)</td>
<td>-0.000</td>
<td>(-1.67)</td>
<td>-0.001</td>
<td>(-4.15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Freight train share (ref.: TGV)</td>
<td>(t=0)</td>
<td>-0.007</td>
<td>(-5.86)</td>
<td>-0.007</td>
<td>(-6.66)</td>
<td>-0.007</td>
<td>(-6.27)</td>
<td>-0.008</td>
<td>(-7.43)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Locomotive only HLP train share (ref.: TGV)</td>
<td>(t=0)</td>
<td>0.000</td>
<td>(0.79)</td>
<td>-0.000</td>
<td>(-0.28)</td>
<td>-0.001</td>
<td>(-1.19)</td>
<td>-0.000</td>
<td>(-0.66)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Productivity of maintenance expenses component $h(\cdot)$

<table>
<thead>
<tr>
<th>$e_k$</th>
<th>Segment length</th>
<th>(t=0)</th>
<th>0.003</th>
<th>(2.56)</th>
<th>1</th>
<th>0.004</th>
<th>(4.57)</th>
<th>0</th>
<th>0.004</th>
<th>(4.13)</th>
<th>0</th>
<th>0.004</th>
<th>(4.12)</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Track length</td>
<td>(t=0)</td>
<td>0.011</td>
<td>(3.80)</td>
<td>1</td>
<td>0.019</td>
<td>(7.18)</td>
<td>0</td>
<td>0.018</td>
<td>(7.02)</td>
<td>0</td>
<td>0.018</td>
<td>(6.75)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Electrified 1500 V (ref.: not electrified)</td>
<td>(t=0)</td>
<td>0.009</td>
<td>(4.25)</td>
<td>1</td>
<td>0.008</td>
<td>(3.70)</td>
<td>0</td>
<td>0.008</td>
<td>(3.71)</td>
<td>0</td>
<td>0.007</td>
<td>(3.17)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Electrified 25000 V (ref.: not electrified)</td>
<td>(t=0)</td>
<td>0.005</td>
<td>(2.29)</td>
<td>1</td>
<td>0.004</td>
<td>(1.99)</td>
<td>0</td>
<td>0.004</td>
<td>(1.76)</td>
<td>0</td>
<td>0.002</td>
<td>(1.05)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Number of switches</td>
<td>(t=0)</td>
<td>-0.002</td>
<td>(-2.79)</td>
<td>1</td>
<td>-0.002</td>
<td>(-2.91)</td>
<td>1</td>
<td>-0.002</td>
<td>(-3.35)</td>
<td>1</td>
<td>-0.002</td>
<td>(-3.34)</td>
<td>1</td>
</tr>
<tr>
<td>$b_0$</td>
<td>Maximum allowed speed</td>
<td>(t=0)</td>
<td>0.061</td>
<td>(21.04)</td>
<td>1</td>
<td>0.055</td>
<td>(27.79)</td>
<td>0</td>
<td>0.054</td>
<td>(27.74)</td>
<td>0</td>
<td>0.053</td>
<td>(26.83)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Suburban line (ref.: other line)</td>
<td>(t=0)</td>
<td>-0.012</td>
<td>(-3.47)</td>
<td>1</td>
<td>-0.011</td>
<td>(-3.50)</td>
<td>-0.011</td>
<td>(-3.42)</td>
<td>-0.005</td>
<td>(-1.52)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High speed rail line (ref.: classic line)</td>
<td>(t=0)</td>
<td>-0.048</td>
<td>(-10.53)</td>
<td>1</td>
<td>-0.022</td>
<td>(-4.02)</td>
<td>-0.017</td>
<td>(-3.04)</td>
<td>-0.019</td>
<td>(-3.70)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W$</td>
<td>Cumulative total tons per km of track</td>
<td>(t=0)</td>
<td>-0.002</td>
<td>(-3.06)</td>
<td>1</td>
<td>-0.000</td>
<td>(-2.04)</td>
<td>-0.001</td>
<td>(-2.80)</td>
<td>1.44</td>
<td>(5.42)</td>
<td>[1.65]</td>
<td>-0.001</td>
<td>(-2.83)</td>
</tr>
<tr>
<td></td>
<td>[t=1]</td>
<td>1</td>
<td>-0.001</td>
<td>(-2.80)</td>
<td>1</td>
<td>1.44</td>
<td>(5.42)</td>
<td>[1.65]</td>
<td>-0.001</td>
<td>(-2.83)</td>
<td>1.41</td>
<td>(5.11)</td>
<td>[1.48]</td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>Time since last regeneration (age of rails)</td>
<td>(t=0)</td>
<td>-0.030</td>
<td>(-15.82)</td>
<td>1</td>
<td>-0.031</td>
<td>(-16.69)</td>
<td>1</td>
<td>-0.026</td>
<td>(-17.88)</td>
<td>2.23</td>
<td>(7.89)</td>
<td>[4.34]</td>
<td>-0.028</td>
</tr>
<tr>
<td></td>
<td>[t=1]</td>
<td>1</td>
<td>-0.026</td>
<td>(-17.88)</td>
<td>1</td>
<td>2.23</td>
<td>(7.89)</td>
<td>[4.34]</td>
<td>-0.028</td>
<td>(-17.84)</td>
<td>2.09</td>
<td>(7.83)</td>
<td>[4.09]</td>
<td></td>
</tr>
</tbody>
</table>

#### Derivative: $h'(\cdot) / h(\cdot)$

| $w$ | Current tons per km of track | (t=0) | 0.000 | (0.23) | 1 | 0.000 | (0.53) | 1 | 0.001 | (1.08) | 1 | 0.001 | (1.06) | 1 |
|     | Degree of freedom | (t=0) | -- | -- | -- | 6 yearly dummies (ref.: 2001) | -- | -- | -- | Note 1 |
| $\beta_k$ | 6 yearly dummies (ref.: 2001) | -- | -- | -- | -- | -- | -- | Note 1 |

#### Log likelihood

| Number of $\beta_k$ | 18 | 18 | 8 | 24 |
| Number of $\lambda_k$ estimated | 0 | 3 | 5 | 11 |
| Number of variant run number | 88 | 89 | 83 | 90 |

Note 1: percentage changes in $\gamma$ ("elasticities" as in Eq. 24-J) caused by the presence of dummies, and the $t$-statistics of their underlying coefficients, are for years 2002 to 2007: -0.001 (-0.55), -0.003 (-1.12), -0.003 (-1.21), 0.001 (0.28), 0.009 (3.10), 0.012 (4.23).

All $t$-statistics are conditional on the value of functional form parameters, linear unless stated otherwise.

However, Column 4 is special in that the use of yearly dummies to capture the rarity of public funds may have improved fit at the expense of biases in parameter estimates, as time period dummies often do in time series models: we therefore select Column 3 results (variant 83)\(^{38}\) to construct\(^{39}\) the needed $E(Sc_t)$ for the required years, namely 2007, 2006 and 2005, because we need to test a two-period (current and one-year lagged) adjustment speed.

---

Note 1: percentage changes in $\gamma$ ("elasticities" as in Eq. 24-J) caused by the presence of dummies, and the $t$-statistics of their underlying coefficients, are for years 2002 to 2007: -0.001 (-0.55), -0.003 (-1.12), -0.003 (-1.21), 0.001 (0.28), 0.009 (3.10), 0.012 (4.23).

---

\(^{38}\) Note in Frame 2 that the Belsley-Kuh-Welsch index (Belsley et al., 1980) is lowest for Col. 3. Although its absolute value has no critical significance and depends on units of measurement of the $X_t$ (Erléek-Rousseu, 1995), comparisons with values obtained with specifications 3A to 3C argue for choosing Col. 3.
Another recursive transport system? These calculations of \( E(Sc_t) \) by year are indeed necessary because, for one, the infrastructure manager’s speed of adjustment of Maintenance expenditures \( u_t \) in 2007 to changes in the pair of “new” variables, assumed above in (23-C) to be contemporaneous \((t=2007)\), may in fact take longer, as tests with \( t=2006 \) will demonstrate. By contrast, the issue of the speed of adjustment in Service \( S_t \), as determined by (26), will be handled by shifting the period of observation of the dependent variable forward from \( t=2007 \) to \( t+1=2008 \) rather than by shifting back the period of the explanatory variables, all then anchored and defined for 2007. As noted in Footnote 8 above, the actual reactivity of supply in seemingly simultaneous transport Demand-Supply systems often shows that they are in fact recursive. In Section 5.5 below, Service Target and Trajectory Correction will affect Maintenance expenditures after a one period lag; in Section 5.6, they will affect Supply in mixed contemporaneous and lagged fashion close to recursiveness.

5.5. Estimation of Phase B Maintenance expenses \( u \)

Equal treatment of segments. In order to explain total Maintenance cost per km of segment with the censored data set \(^{40}\) for 2007, we make two significant changes \(^{41}\) to the specification (23-B) found in Tables 5 and 6. The first consists in using the cumulative tonnage variable \( W \) without distinguishing its role across segments of different UIC standing because trials with the database, now devoid of Phase A and Phase C segments, showed that the coefficient of \( W_{2,6} \), tonnage for segments subjected to periodic renewals, was almost identical to that of \( W_{7,9} \), tonnage for segments never renewed but only maintained.

This important behavioural finding implies that maintenance expenses are not related to standing as such but to objective factors and that the distinction between track that is renewed, and track that is not, is a technical distinction based on optimizing behaviour described by the three-phase model proposed here and involves no expenditure discrimination based on a subjectively lower standing of segments belonging to UIC categories 7-12 \(^{42}\). Concerning those objective factors, Figure 7 graphs of maintenance against \( W \) and \( t \) (agereal) naturally lead one to expect more linkage to the former than to the latter.

Figure 7. Total Maintenance cost (euros per km) \( u \), censored 2007 sample (673 obs.)

<table>
<thead>
<tr>
<th>A. Cost ( u ) and cumulative tonnage ( W )</th>
<th>B. Cost ( u ) and time ( t ) proxy (agereal)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image.png" alt="Graph A" /></td>
<td><img src="image.png" alt="Graph B" /></td>
</tr>
</tbody>
</table>

\(^{39}\) Alternatively, given the logarithmic dependent variable in (27-B), \( E(Sc_t) \) could have been calculated for any sample year \( t=2001, \ldots, 2007 \), by collapsing (24-H) to \( E(y_i) = e^{\exp \left[ \sum \beta X_i^{(j)} \right]} \), where \( k \) is the sample mean of the log-normal random variable \( \exp(w_i) \) for the period.

\(^{40}\) In all forthcoming step c models for (23-C) and (26), estimates will be obtained with 580 observations corresponding to the set available for yearly Service \( S_t \) in Table 7. For maintenance cost \( u_t \), 673 observations are in fact available in 2007, but there is no simple way to estimate the needed \( E(Sc_t) \) for observations beyond the 580 used in Table 7 models.

\(^{41}\) Also, dummy variables for the type of electrification and the designation of suburban and high speed segments were not significant and were removed.

\(^{42}\) For memory, we adopted the SNCF breakdown of UIC categories 7-9 as between “with passenger” (7-9) and “without passengers” (10-12) segments (SNCF, 1999) and use this distinction in Figure 7.C below.

28
The problem of train speed. The second change pertains to Vehicle speed, required on theoretical grounds because track settlement, rail component fatigue and abrasive wear, to say nothing of rolling contact fatigue, are all related to it. But actual speeds by train type differ much from the maximum allowed speed variable $v$ graphed in Figure 4 and summarily used above as a measure of track quality or standing in regressions for Table 6. Although higher quality may as such have higher maintenance costs, it also allows for higher operating speeds causing higher damages, as recognized for instance in the Equivalent Gross Tonne Miles (EGTM) formula used by the UK Office of Rail Regulation (ORR) described in Ongaro & Iwnicki (2009) who indicate that, in the multiplicative format then used, speed had a power of 0.64 for tracks and 1.52 for structures. These estimated powers are not especially high if one remembers that train energy consumption is often deemed to increase with the square of speed (e.g. García et al., 2008). Speed is surely desirable on theoretical grounds.

In view of the ambiguity of the maximum allowed speed variable and of the absence of data on actual speeds by train type by segment, we formulate here two sets of model variants: one without any speed term $v$ and one with speed constructs $v_c$ (c=1, ...,6) deriving maximum allowed speed by train type from maximum allowed speed $v$ by segment. In Tables 9 and 10 no speed variable is used, but in Table 13 of Appendix 3 speed constructs are added to the 6 reference models: it will be seen that they do not influence the remaining parameters of the model but that they obtain BCT values between 1.6 and 3.6, which suggests that the procurement of actual data on speeds would unambiguously improve the reference models.

Reference model results. We first present in Table 9 results of 4 variants specified with traffic aggregates\(^{43}\), the last pair of which (Columns 3.A and 4.A) may be compared to those of Columns 3.B and 4.B in Table 10\(^{44}\), where weight variables are simply disaggregated with a common BCT and the remaining specifications are otherwise unchanged. In all 6 reference cases, cumulative and current traffic $W_t$ and $w_t$ are summed to avoid collinearity\(^ {45}\).

As in other tables, the $t$-statistics with respect to 0 and 1 are computed from first partial derivatives of the Log of $\Lambda$ from (24-G) by the Berndt et al. (1974) method; and we run the known risk of finding coefficients with low $t$-statistics (w.r.t. 0) for some variables with significant BCT powers, a problem of conditional logic in BCT use that is just beginning to be studied (e.g. Cho & Ishida, 2012). Starting with the 4 aggregate models of Table 9, note:

i) specification of variables and flexible form. Column 1 presents a formulation with a fixed multiplicative form as close as possible to that of the UK ORR mentioned above. In this specification, the time proxy variable (agerail) obtains a negative sign which progressively disappears with model refinements of the next three columns; but this fragile variable of expected positive sign is however never really significant and never yields a strong elasticity.

Concerning fit, massive Log Likelihood gains of 110 points are obtained in Column 2 by estimating only 3 BCT, all very different from the logarithmic case value assumed in Column 1, namely: 0.25 for the dependent variable $u$, 0.54 for the three technical state variables $e$ and 0.58 for total tonnage ($W+w$).

ii) speed of adjustment. In the third column, the addition of the “new” Service Target and Trajectory Correction variables required by (23-C) yields the expected positive and negative signs for these variables —with acceptable $t$-statistics— and a significant gain of 11 points in Log Likelihood for a difference of 2 degrees of freedom. But note that the linear\(^ {46}\) terms used are the lagged values $\left[ E(S_{c_{12}}) - E(S_{c_{22}}) \right]$ and $\left[ S_{c_{12}} - E(S_{c_{22}}) \right]$, selected primarily because they are both more

---

43 With cumulative weight variables defined for total traffic, models presented in Table 6 are also “aggregate”.
44 For brevity, Table 10 only includes regressors that differ from those of the corresponding models 3.A and 4.A in Table 9: complete results are available from the authors.
45 For instance correlation as high as 0.99 for Île-de-France trains.
46 BCT cannot be used because these variables are not strictly positive.
statistically significant$^+$ than contemporaneous values $[E(S_i) - E(S_{i-1})]$ and $[S_{i-1} - E(S_{i+1})]$. But there is as second reason. The lag may help to dodge an identification problem that arises with $[E(S_i) - E(S_{i-1})]$, a variable which, in results not shown, is not only less significant than $[E(S_{i-1}) - E(S_{i-2})]$ but also of the wrong (negative) sign. Had $h(\cdot)$ been linear in all terms, we would have had by Equation (8) that $[S_i - S_{i-1}] \approx [\delta_0 + \gamma q]$ and the two new parameters $\delta_0$ and $\gamma$ would have been absorbed by the constant and by the coefficient of $w$ in $f(\cdot)$. Here, however, due to the summing of $w$ and $W$ and to the nonlinearity of that sum$^+$, this identification problem is dodged and, if still present at all, bypassed by the one-period lag. The reasonableness of this interpretation will be confirmed below in the analysis of derived Supplied Service $S$.

Other elasticities are reasonable and stable except for that of segment length, much too significant until the introduction of spatial correlation in Column 4.A. The total ton traffic variable $W+w$ is highly significant with BCT value$^+$ between 0.38 and 0.58.

iii) directed correlation of residuals. The addition of first order directed autocorrelation in Column 4 yields huge further gains (31 points of Log Likelihood for the single additional parameter $\rho_l =0.59$) under the assumption that error terms of segments are correlated if they belong to the same SNCF administrative region. As indicated in Table 8, the number of track segments present in the sample varies across regions, with implications for the structure$^+$ of $R$.

| Table 8. SNCF regions in 2007, their number of neighbors and of segments in the sample |
|---|---|---|---|---|---|---|
| N° | Name* | Neighbors | Segments | N° | Name* | Neighbors | Segments |
| 10 | PARIS EST** | 5 | 5 | 43 | BORDEAUX | 4 | 51 |
| 14 | REIMS | 7 | 17 | 44 | LIMOGES | 4 | 17 |
| 17 | METZ-NANCY | 3 | 36 | 46 | TOURS | 6 | 19 |
| 18 | STRASBOURG | 2 | 31 | 47 | TOULOUSE | 4 | 33 |
| 20 | PARIS NORD** | 7 | 25 | 50 | PARIS SUD EST** | 8 | 24 |
| 23 | LILLE | 2 | 38 | 53 | DIJON | 7 | 30 |
| 24 | AMIENS | 4 | 14 | 54 | LYON | 5 | 35 |
| 30 | PARIS ST-LAZARE** | 5 | 11 | 56 | CLERMONT-FERRAND | 7 | 27 |
| 33 | ROUEN | 5 | 21 | 57 | CHAMBERY | 3 | 31 |
| 34 | PARIS RIVE GAUCHE** | 5 | 25 | 58 | MARSEILLE | 3 | 29 |
| 36 | RENNES | 2 | 14 | 59 | MONTPELLIER | 4 | 23 |
| 37 | NANTES | 5 | 24 | **TOTAL NUMBER OF SEGMENTS 580** |

* In 2012, a merger of regions 10 and 14 and of regions 20 and 24 reduced the total number to 21.

** Administrative regions located in Île-de-France and treated as first neighbors in matrix $R$.

Such positive correlation (denoting substitutes) may reflect the arbitrariness of the administrative assignment of some expenditures across neighboring segments, as implicit in the fact that the size of the coefficient of the segment length variable and its statistical significance are greatly reduced in Column 4 to a low level one would have expected to find from the start$^+$.

In previous work (Gaudry & Quinet, 2003) on the 1999 database with a Generalized Box-Cox specification applied to specification (23-A), and with interactions among explanatory variables accounted for, a different point was made with respect to administrative regions by simply adding to the regression a dummy variable for each of 22 regions, taking region 10 as an arbitrary reference: this improved the Log Likelihood by 24 points and showed that maintenance costs were

$^+$ The correlation between current and lagged Target Service variables is -0.21 and that between Trajectory Correction variables is 0.81; one could envisage using both current and lagged Trajectory correction variables if Maintenance cost data were available for more than the single year 2007.

$^+$ The negative sign of $[E(S_i) - E(S_{i-1})]$ may just correct the adjustment provided by the sum $(W+w)$.

$^+$ A value resembling the 0.49 power for the axle weight variable in the UK ORR formula for track damages.

$^+$ Consider Region 10, with the smallest number of segments, and the Region 43 with the highest: in matrix $R$, the 5 rows corresponding to the former have 4 values of 1.00 and the 51 rows corresponding to the latter have 50 such values. In the row-normalized $\tilde{R}$, the former values become 1/4 and the latter 1/50.

$^+$ Studies produced by the CATRIN consortium (Wheat et al., 2009) tend to show that segment lengths are statistically significant, a result which might indicate an endogeneity of segment lengths.

30
significantly higher in only two regions. The same exercise carried out here yields hardly any gain, which suggests that our current specification of (23-C) with censored Phase B observations is more realistic. Concerning these gains with regional dummies, the weather might be responsible for some marginally higher costs detected for regions 56 and 57 where snow is frequent.

iv) impact of disaggregation. Relaxing in Table 10 the assumption that tonnage coefficients are the same across train types yields in both cases of Columns 3.B and 4.B less than 1 point of Log Likelihood despite the 5 additional coefficients estimated. This is surprising because the additional regression coefficients should reflect the specifics of each train category: actual weights per axle, un-sprung mass and other technical features absorbing the effect of gross axle weights. The almost null regression coefficients estimated for intercity (GL) and Île-de-France regional (IdF) trains may be due to the difficulties faced in the construction of their cumulative values before 1995. The estimated BCT power of weight is increased from 0.39 to 0.73 by the disaggregation.

How to demonstrate the role of speed if speed is not observed? In order to match the very reasonable current practice of the UK ORR mentioned above, we tested in Table 13 of Appendix 3 the inclusion of an operating speed variable in the damage function \( f(v) \). Despite the lack of observed speeds in the database (viz. Table 3), we constructed \( v_r \), an operating speed for each of the 6 train types and \( V \), their sum weighted by their respective shares of total accumulated tons \( W \) (past \( W_p \)), measured at the beginning of each year, and current \( W_c \) by train type. We first assumed that trains were driven at the lowest of 90% of maximum allowed speed \( v \) on the segment and of the maximum speed allowed by train type [assumed to be 200 km/h for main intercity (GL) and regional (TER) trains, 100 km/h for freight (FRT) trains and 50 km/h for locomotives (HLP)].

Table 13.A of Appendix 3 presents the same models as those found in Table 9, but with the addition of the aggregate speed construct \( V \) and under the assumption that speed coefficients, like those of weight, are common to all train types. This assumption is relaxed in Table 13.B, where the speed variable is disaggregated with a single BCT (as for weights).

Comparing Tables 13.A and 13.B to Tables 9 and 10, one notes that the addiction of one aggregate or of many disaggregate speed terms increases fit by only 2 points in Column 2 (at the cost of an additional regression coefficient and of a BCT) and yields relatively smaller statistical gains in the other 5 columns. Concerning Table 13 results, note:

v) reasonable powers of speed. Despite these weak statistical gains, the estimated BCT on speed is reasonable, varying between 1.6 and 3.6 with a mean value around 2.7. Further attempts to improve the specification by adding terms of interaction between speed and weight (either total or by train type) with own BCT never yielded significant gains and were abandoned.

vi) impact of new terms. The introduction of the two “new” lagged terms in Column 3 of Table 13.A reduces the elasticity of the dependent variable with respect to speed, a variable which also becomes less significant —perhaps does component \( S_{ij} \) embody too much past history?

vii) incorrect speed constructs. Also, some three train categories have negative coefficients in Column 4.B of Table 13.B, which suggests that assumed speed maxima are incorrect and that data on actual speeds are required to yield proper results.

---

52 In tests made on Table 13.A, a variant had 18 additional regression coefficients than variant 193 but a gain in Log Likelihood of only 7 points. Regions 56 and 57 appeared to have marginally higher costs and Region 59 to have significantly lower costs those of the 5 reference Île-de-France regions.

53 In Sweden, as much as 80% of track maintenance costs are attributable to snow clearing (Andersson, 2006).

54 Axle weights are typically 22.5 tons for standard locomotives and loaded freight wagons, 17 tons for high speed (TGV) and Île-de-France (IdF) regional train locomotives and 10-12 tons for pulled passenger wagons.

55 For instance, axels of Bo-Bo and Co-Co freight locomotives generate different levels of damages in curves.

56 The construction of distinct speed aggregates for past and current tonnages yielded almost collinear variables.

57 Adding maximum allowed speed (vmax) to runs of Tables 9-10 increases the Log Likelihood in the 6 models by: 0.39 (run 200); 1.71 (run 201); 0.67 (run 202); 0.67 (run 203); 0.62 (run 204); 1.52 (run 205). In the first 4 models of Table 13, maximum allowed speed by train type fails to improve this fit but lowers the BCT on the speed term from about 4.00 in the former cases to about 3.0 in the latter.

58 Most of the gain of 10 points is attributable to the Trajectory correction term.
The sign of freight train speed became positive when the speed maximum was lowered from 100 to 80 km/h.

Table 9. Phase B Maintenance cost $u$ without max. speed $v$ (580 obs., 2007)

<table>
<thead>
<tr>
<th>Column</th>
<th>1</th>
<th>2</th>
<th>3.A</th>
<th>4.A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity $\eta(u), \lambda_0, {t\text{-stat.}=0}^*, {t\text{-stat.}=1}$</td>
<td>$\eta(u)$</td>
<td>$\lambda$</td>
<td>$\eta(u)$</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>$u$ Total maintenance cost per km (dependent variable)</td>
<td>n.a.</td>
<td>0.25</td>
<td>(19.78)</td>
<td>0.24</td>
</tr>
<tr>
<td>$\beta_0$ Intercept</td>
<td>n.a.</td>
<td>(12.05)</td>
<td>n.a.</td>
<td>(10.00)</td>
</tr>
<tr>
<td>$S$ Target $S$ service: 2006-2005 [E(Sx,1) - E(Sx,2)]</td>
<td>0.08**</td>
<td>(4.63)</td>
<td>1</td>
<td>0.10**</td>
</tr>
<tr>
<td>Trajectory correction: (obs.-target)$_{2005}$ [S$_x$-E(Sx,2)]</td>
<td>-0.002</td>
<td>(-3.26)</td>
<td>-0.002</td>
<td>(-3.63)</td>
</tr>
<tr>
<td>$c_0$ Segment length</td>
<td>-0.121</td>
<td>(-2.95)</td>
<td>-0.073</td>
<td>(-1.97)</td>
</tr>
<tr>
<td>Track length</td>
<td>0.118</td>
<td>(1.07)</td>
<td>0.183</td>
<td>(2.34)</td>
</tr>
<tr>
<td>Number of switches</td>
<td>0.189</td>
<td>(9.94)</td>
<td>0.411</td>
<td>(19.91)</td>
</tr>
<tr>
<td>W+w Cumulative+ current total tons (W+w)</td>
<td>0.288</td>
<td>(9.29)</td>
<td>0.262</td>
<td>(9.71)</td>
</tr>
<tr>
<td>$t$ Time since last regeneration (agerail)</td>
<td>-0.101</td>
<td>(-0.97)</td>
<td>-0.021</td>
<td>(-0.27)</td>
</tr>
<tr>
<td>$R_t$ Same region: $\rho_1$</td>
<td>0.589</td>
<td>(6.87)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Log likelihood: -6977.59, -6868.13, -6857.01, -6825.77

* All $t$-statistics of $\beta_0$ are conditional on the value of functional form parameters and those of the $\lambda_0$ are unconditional.

** The sign of $\beta_0$ is shown (the mean value of the variable is negative).

Table 10. Phase B Maintenance cost $u$ without max. speed by train type $v_e$ (580 obs., 2007)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity $\eta(u), \lambda_0, {t\text{-stat.}=0}^*, {t\text{-stat.}=1}$</td>
<td>$\eta(u)$</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>$W_{e+w}$ Disaggregated variables replacing $W_{e+w}$ in Columns 3.A and 4.A of Table 9 are shown.</td>
<td>Total tons ($W_{e+w}$)</td>
<td>Total tons ($W_{e+w}$)</td>
</tr>
<tr>
<td>TGV: high speed trains</td>
<td>0.111</td>
<td>(3.87)</td>
</tr>
<tr>
<td>GL: classic intercity trains (Corail)</td>
<td>0.006</td>
<td>(0.17)</td>
</tr>
<tr>
<td>TER: regional trains</td>
<td>0.048</td>
<td>(1.79)</td>
</tr>
<tr>
<td>IdF: Île-de-France trains</td>
<td>-0.006</td>
<td>(-0.20)</td>
</tr>
<tr>
<td>FRT: freight trains</td>
<td>0.081</td>
<td>(3.27)</td>
</tr>
<tr>
<td>HLP: locomotives</td>
<td>0.052</td>
<td>(1.97)</td>
</tr>
<tr>
<td>$t$ Time since last regeneration (agerail)</td>
<td>0.004</td>
<td>(0.05)</td>
</tr>
<tr>
<td>$R_t$ Same region: $\rho_1$</td>
<td>0.586</td>
<td>(6.59)</td>
</tr>
</tbody>
</table>

Log likelihood: -6856.99, -6826.30

* All $t$-statistics of $\beta_{v_e}$ are conditional on the value of functional form parameters and those of the $\lambda_0$ are unconditional.

59 The sign of freight train speed became positive when the speed maximum was lowered from 100 to 80 km/h.
Concerning the results presented in Tables 9 and 10 and in Appendix 3, note overall:

viii) estimated powers of variables. The only BCT with an optimal value at 0, assumed for all variables in the UK ORR model, is the fragile time proxy variable (agerail); for all other variables, best fit values all differ from the logarithm. For one, the dependent variable has a sturdy average BCT value of about 0.23 in line with all of our previous results with the 1999 database.

xi) other forms of directed autocorrelation. A systematic attempt (with a slightly larger data set of 673 observations) to define a second type of directed autocorrelation $R_1$ based on regions sharing a border with that of the given segment, a number which varies from 2 to 8 as indicated in Table 8, and to define yet another type $R_2$ under the assumption that the error term of a segment could be correlated with those of other segments sharing one or more train lines, yielded no significant correlation in models specified as those found in the 4 Columns 3.A to 4.B of Table 13. The attempt to further estimate proximity parameters $\pi_2$ and $\pi_3$ in addition to $\rho_2$ and $\rho_3$ also failed to provide statistically significant improvements in fit beyond those of models 4.A and 4.B, whether the additional AR-C-D processes were considered separately or jointly.

In toto, the role of within-region SNCF administrative practices is considerable in explaining reported total maintenance cost by segment; unaccounted for influences of neighboring regions have a negligible effect. Also, the fact that segment residuals are uncorrelated with the presence of common train lines suggests that the underlying representation of damage function $f(\cdot)$ is adequate whether speed measures are used or not.

5.6. Estimation of derived Phase B Service supply $S$

As stated in (26), the third relationship of interest in our model says that the existence of Service Target $\left[ E(S_{ct}) - E(S_{ct-1}) \right]$ and Trajectory Correction $\left[ S_{t-1} - E(S_{t-1}) \right]$ factors, combined with reactions to random Surprises in actual maintenance expenditures, $[\varepsilon_{tt} = u_t - E(u_t)]$ mechanically imply a certain Service Supply level path $\Delta S_t$.

But the need to pass from continuous to discrete time and the unknown speed of reaction of the infrastructure firm both raise the possibility of a lag between these three stated path triggers and realized Supply: on these lines, we have just noted in Tables 9 and 10 the presence of a one-period lag between the pair of “new” variables and Maintenance expenses.

We therefore test here the possible existence of such a lag on the determination of Service Supply by stating the trigger values in the period 2007-2006 and looking at realized $\Delta S_t$ both contemporaneously (2007-2006) and one year later (2008-2007). One sees in Figure 8.A and 8.B that these specifications of the dependent variable $S_t - S_{t-1}$ are both reasonably well centred about 0 and independent from time.

Concerning the first two explanatory variables $\left[ E(S_{ct}) - E(S_{ct-1}) \right]$ and $\left[ S_{t-1} - E(S_{t-1}) \right]$ listed in Table 11, their expected value components are derived from variant run 83 (Column 3 in Table 7). The expected value component found in $[\varepsilon_{tt} = u_t - E(u_t)]$, the third variable standing for Surprises, is derived from variant run 105 (Column 4 of Table 9) in Columns 1 and 2 and, in Columns 3 and 4, from run 194 of Column 4 in Table 13.A of Appendix 3.

60 The same holds for all models of Table 6 where that unreported BCT on cost varies between 0.18 and 0.28.
61 A parameter $\pi_i$ would not be meaningful: the matrix $R_i$ partitioned by region is symmetric and block diagonal. It is not clear what its powers would mean.
62 We did not test the differences across the 23 regions, as these require using at least as many subsets of $R_i$.
63 We recall that in Table 9 there is no speed variable but that there is one in Table 13.
Firstly, we aim at reproducing the relation between the traffic level and the lifetime of renewals: we first deduce from the accounts the average cost of the renewal; then from the relations developed in section 3.2 on the optimal lifetime of renewal, it is possible to derive the optimal lifetime for several traffic levels. The relation is plotted in Figure 9 and compared to the relation given by

As the dependent and the three explanatory variables in Table 11 all include negative values, results shown are per force obtained under linearity assumptions. First note that the Surprise term \( c_u = u_t - E(u_t) \) is never significantly different from 0, a relative weakness which may be due to the fact that Maintenance expenses are known to be observed with significant error. The results for the two other variables globally imply that the speed of adjustment is not fully contemporaneous: Target Service has the expected positive sign only in the regressions for 2007-2006 where the trajectory correction term is also most significant: in the regressions for 2008-2007, the former is not of expected sign and the latter becomes less significant but remains strong enough to imply that adjustment is not fully instantaneous.

<table>
<thead>
<tr>
<th>Table 11. Explaining Service quality changes ( \Delta S ) by factors defined for 2007 and 2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column</td>
</tr>
<tr>
<td>Elasticity ( \eta(\Delta S) ) and ( t )-statistic of ( \beta_k )</td>
</tr>
<tr>
<td>( \beta_0 )</td>
</tr>
<tr>
<td>( E(S_{0t}) - E(S_{0t+1}) ) ( (t=2007) )</td>
</tr>
<tr>
<td>( S_{0t} - E(S_{0t}) ) ( (t=2007) )</td>
</tr>
<tr>
<td>( u_t - E(u_t) ) ( (t=2007) )</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
</tr>
<tr>
<td>Number of ( \beta_k ) estimated</td>
</tr>
<tr>
<td>Variant run number</td>
</tr>
</tbody>
</table>

The fact that Supplied Service does not fully respond immediately to the important Trajectory correction term partially confirms the remark made in the previous section concerning the lagged impact of the first two trigger terms on maintenance expenditures. The infrastructure manager’s Maintenance Expenditure-Service Supply system appears to be partially recursive, at least in the context of this very limited two-period analysis.

5.7. Other calibrations

The following calibrations are made for a track link assumed to have the average characteristics of the French network. Firstly, we aim at reproducing the relation between the traffic level and the lifetime of renewals: we first deduce from the accounts of the infrastructure manager the average cost of the renewal; then from the relations developed in section 3.2 on the optimal lifetime of renewal, it is possible to derive the optimal lifetime for several traffic levels. The relation is plotted in Figure 9 and compared to the relation given by...
the doctrine of the infrastructure manager as described in SNCF documents (Rail Concept, 2006). It is clear that the correspondence is very good.

Secondly, we calculate in Table 12 the ratio between the yearly maintenance budget and the yearly renewal budget, showing for various traffic levels both the observed ratio and the ratio deduced from simulations of the model. Here again the correspondence is fairly good.

Such calibrations tend to show, in addition to the econometric tests, that the model is a good representation of the behaviour of the infrastructure manager. The implicit and experimental optimization made by this manager can be reproduced through a formalised optimisation model, as it is often assumed for the cost functions of usual firms.

![Figure 9. Relation between the yearly traffic and the lifetime of renewals](image)

**Table 12. Ratio between Renewal and Maintenance expenses for several levels of traffic**

<table>
<thead>
<tr>
<th>Traffic per day in tons</th>
<th>Observed Ratio renewal/maintenance</th>
<th>Calculated Ratio renewal/maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>82 000</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>41 000</td>
<td>0.32</td>
<td>0.19</td>
</tr>
<tr>
<td>10 250</td>
<td>0.08</td>
<td>0.09</td>
</tr>
</tbody>
</table>

**6. Conclusion**

We show in this article that it is possible to jointly optimize current maintenance, known to postpone regeneration, and regenerative maintenance itself through the introduction of track quality of service, shown by technical expertise to decrease with cumulative traffic, and its link to both traffic and maintenance expense levels.

When such assumed linkages are introduced in a simple model where quality changes are proportionate to maintenance expenses, it becomes possible to characterize the optimal maintenance policy, namely that which maximises the discounted difference between the surplus provided by the quality of service and the amount of current and regenerative maintenance. The following features of the analysis are noteworthy:

- At any time and *ceteris paribus*, the amount of current maintenance and the quality of service increase with the yearly traffic and decrease with the discount rate;

- The trajectory of an optimal maintenance policy consists in three phases. In the first, just after a renewal, current maintenance is *nil* (or, more realistically, limited to surveillance) and the quality of service decreases at a speed which depends on the level of traffic. During the cruising phase, current maintenance is positive and increases at the same time as quality of service remains stable or slowly decreases. In the final third phase just before the next renewal, current maintenance again
decreases to zero (or, more realistically, is limited to surveillance and minimal maintenance) and the quality of service decreases sharply;

- The presence of uncertainty has only weak consequences: the cruising level of service is slightly lowered and current maintenance is stopped somewhat earlier before renewal;

- It is possible to compute the marginal social cost and to deduce from it the optimal pricing strategy whereby the optimal charge differs from the usual marginal infrastructure cost in two respects. First, in the initial and final phases when the maintenance expenses are zero, there is a negative external cost caused by the marginal user to others due to the deterioration of the quality of service. Second, in the cruising phase, there is a positive external cost due to the fact that the optimal infrastructure manager policy response is to increase the quality of service in the presence of the marginal user in a manner similar to the Mohring effect for service frequency. The overall effect naturally depends on the relative importance of this pair of opposite occurrences of externalities not defined, captured or accounted for in current marginal infrastructure cost pricing doctrines;

- In a calibration of this theory made with French data, the behaviour of the model with respect to the renewal horizon and its dependence on traffic level, as well as the overall proportion of maintenance expenses (as between current and renewal), were simulated at country-wide level: the results reproduce rather well the manager’s behaviour. In econometric tests of the model based on data for spatially differentiated track segments resulting from a breakdown of the French rail network in about 800 links for which several variables are available (traffic, quality of service, technical characteristics), one finds that the model accounts well for the data. Indeed there exists multiple phases of current maintenance expenses and of service provision, both of which can be explained satisfactorily: the model specifications with Box-Cox transformations and directed autocorrelation of residuals dramatically improve the quality of adjustments over specifications of fixed functional form (typically multiplicative) or neglecting the spatial correlation of residuals;

- This work should be extended in several directions: from a theoretical point of view first, it would be necessary to give up the assumption of traffic stability and to allow for random variations or drifts; it might also be useful to split the model into parts, one for pure maintenance works and one for surveillance expenses which obey rules different from those of maintenance works proper, and to develop a specific model for such surveillance operations. On the econometric side, the use of observed speeds by train type would much improve the results over those obtained with approximately constructed maximum speed data and allow for calculations of marginal cost by train type better reflecting actual costs than those derived merely from different gross tonnages by train category without train specific track damage parameters.

- Finally a last question arises: it is clear from the econometric validation that the infrastructure manager performs an optimisation in which maintenance expenses and quality of service are weighted, and that he gives a monetary value to this quality of service. But does this quality of service value correspond to the value attributed by the users, which are the passengers and the freight forwarders? In other words, do we have a system equilibrium or an infrastructure manager equilibrium? Research in this direction should be both difficult and important.

7. Acknowledgements

An earlier draft of 6th October 2010 by Gaudry and Quinet, entitled Optimisation de l’entretien et de la régénération d’une infrastructure: exploration d’hypothèses, benefitted from comments by Bernard Caillaud and Matthieu de Lapparent and was presented without econometric tests at the Kuhmo Nectar Conference on Transportation Economics in Stockholm on 1st July 2011 under the title “Joint optimization of continuous maintenance and periodic renewal”. The authors thank Marc Antoni, Richard Arnott, David Meunier and Yves Puttalaz for discussions or comments, Cong-Liem Tran for computing assistance and are grateful to Société nationale des chemins de fer français (SNCF) for financial support and for allowing inclusion in this version of estimates based on databases constructed by Michel Ikonomov and Pascaline Boyer. Exploratory estimates obtained from fixed form regression specifications were presented at the Kuhmo Nectar Conference on Transportation Economics in Berlin on 21st June 2012 through David Meunier’s good offices.
8. References


### 9. Appendix 1. The optimal duration of a renewal

The expression to be maximised is:

\[
M(T) = \frac{J(T) - De^{-JT}}{1 - e^{-JT}}
\]

and its derivative with respect to \( T \) is:

\[
J^*(T)(1-e^{-JT}) + jDe^{-JT} - jJ^*(T)e^{-JT} = 0
\]

where, by the classical properties of the Hamiltonian, \( J^*(T) = H(T) \).

When this equation has a solution (which it might not have, for instance under low traffic conditions if infrastructure is then maintained indefinitely and never renewed —as appears to be the case for track segments belonging to UIC groups 7-9), this solution is a maximum if the second derivative is negative.

The sign of the second derivative of \( M(T) \), which is the same as the sign of \( dS(T)/dT \) because the remaining R.H.S. terms are positive, is negative due to the fact that, in the final phase, \( dS/dT = -f(K,Q,q,t) \). In view of this negativity, the solution of the initial equation corresponds to a maximum.
10. Appendix 2. Discretization of the singular stochastic control problem

We consider here a singular stochastic control problem with dynamic given by:

\[ dS_t = [k(q,t)u(t) - f(q,t)] + \sigma dZ_t, \]

with \( S_0 = x \), where process \( u(t) \) depends on time and randomness but belongs to \([0,m]\), and we want to maximize, among all adapted processes \((u(t), t \geq 0)\), the expected value functional:

\[ \mathbb{E}\left( \int_0^T \left[ -u(t) - \alpha e^{-\beta S_t} \right] e^{-\gamma t} \, dt \right). \]

To solve this problem, and now using \( \mathbb{E}_{t,x} \) to refer to the fact that \( S_t = x \), define

\[ J(t,x) = \max_{u(t) \in [a,m]} \left\{ \mathbb{E}_{t,x}\left( \int_t^T \left[ -u(s) - \alpha e^{-\beta S_s} \right] e^{-\gamma (s-t)} \, ds \right) \right\}, \]

which, following Haussmann & Suo (1995b), will solve (as a viscosity solution) the following Hamilton-Jacobi-Bellman (HJB) equation

\[ J_t(t,x) + \max_{a \in [a,m]} \left[ \left[ (u-a)e^{ax} \right] + (k(q,t)u(t) - f(q,t))J_x(t,x) \right] + \frac{\sigma^2}{2} J_{xx}(t,x) - J(t,x) = 0, \]

now with \( u \in \mathbb{R}, \quad x \in \mathbb{R}, t \leq T \) and \( J(T,x) = 0 \). When \( \sigma \neq 0 \) and no analytically explicit solution can be computed using the maximum principle, we use a numerical procedure, in this case the fully explicit finite difference method. This method is easy to implement but plagued by stability problems: when the space-step \( \Delta x \) is given, the time-step \( \Delta t \) needs to be small enough. These problems can be solved by verifying that we are approximating the initial diffusion problem by a discrete Markov chain constructed on the finite difference grid, a procedure discussed in Kushner & Dupuis (2001) and often called the “Kushner method”.

In order to obtain a stable numerical scheme here, we choose an “upward” finite difference scheme for the convection term \( J_x(t,x) \), given \( (u-a)e^{ax} \), we effectively approximate \( J_x(t,x) \) by \( \frac{J(t,x+\Delta x) - J(t,x)}{\Delta x} \) if \( A \geq 0 \) and by \( \frac{J(t,x) - J(t,x-\Delta x)}{\Delta x} \) if \( A < 0 \), switching regimes in accordance with the sign of \( A \). For \( J_{xx}(t,x) \), we use \( \frac{J(t,x+2\Delta x) - 2J(t,x) + J(t,x-2\Delta x)}{(2\Delta x)^2} \), the standard finite difference scheme. This leads to an approximation of the differential part involved in the HJB equation:

\[ \Delta t \left[ (k(q,t)u(t) - f(q,t))J_x(t,x) + \frac{\sigma^2}{2} J_{xx}(t,x) \right], \]

which can be written

\[ p_{-1}J(t,x-\Delta x) + p_0J(t,x) + p_{+1}J(t,x+\Delta x), \]

where \( p_0 = -p_{-1} - p_{+1} \), and

if \( A > 0 \), then \( p_{-1} = \Delta t(A+B) \) and \( p_{+1} = \Delta tB \), with \( B = \sigma^2 / 2\Delta x^2 \); if \( A \leq 0 \), then \( p_{-1} = \Delta tB \) and \( p_{+1} = \Delta t(-A+B) \).

Having verified that \( p_{-1}, p_{+1} \) and \( 1 - p_{-1} - p_{+1} \) are positive, the solution of the discretized scheme can be related to the value function of a controlled Markov chain. Those conditions fulfilled, convergence of the (viscosity) solution of the initial HJB can be proved when \( \Delta x \) and \( \Delta t \) approach 0.
### 11. Appendix 3. Adding speed to maintenance cost models

#### Table 13.A

<table>
<thead>
<tr>
<th>Column</th>
<th>Number of observations</th>
<th>1</th>
<th>2</th>
<th>3.A</th>
<th>4.A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticeity $\eta(u), \lambda$, ($t$-stat.=$0$)*, ($t$-stat.=$1$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$u$ Total maintenance cost per km (dependent variable)</td>
<td></td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.25</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>($t=0$)</td>
<td>($t=0$)</td>
<td>($t=1$)</td>
<td>($t=1$)</td>
<td>($t=1$)</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=1$)</td>
<td>($r=1$)</td>
<td>($r=1$)</td>
</tr>
<tr>
<td>$S$ Target $\Delta$ service : 2006-2005</td>
<td></td>
<td>0.08**</td>
<td>0.10**</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>[E(Sc,1) - E(Sc,2)]</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=1$)</td>
<td>($r=1$)</td>
</tr>
<tr>
<td></td>
<td>Trajectory correction: (obs.-target)2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Sc,1 - E(Sc,2)]</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=1$)</td>
<td>($r=1$)</td>
</tr>
<tr>
<td>$e_0$ Segment length</td>
<td></td>
<td>-0.125</td>
<td>-0.096</td>
<td>0.52</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
</tr>
<tr>
<td></td>
<td>Track length</td>
<td></td>
<td>0.116</td>
<td>0.179</td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
</tr>
<tr>
<td></td>
<td>Number of switches</td>
<td></td>
<td>0.190</td>
<td>0.415</td>
<td>0.394</td>
</tr>
<tr>
<td></td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
</tr>
<tr>
<td>$(W+w)$ Cumulative+ current total tons</td>
<td></td>
<td>0.280</td>
<td>0.254</td>
<td>0.64</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>($r=0$)</td>
<td>($r=1$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
</tr>
<tr>
<td></td>
<td>$(W+w)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V$ Speed weighted by $(W,w,v_c)$ shares</td>
<td></td>
<td>0.032</td>
<td>0.067</td>
<td>2.65</td>
<td>3.62</td>
</tr>
<tr>
<td></td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
</tr>
<tr>
<td></td>
<td>Time since last regeneration (ageral)</td>
<td></td>
<td>-0.868</td>
<td>0.021</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
</tr>
</tbody>
</table>

#### Table 13.B

<table>
<thead>
<tr>
<th>Column</th>
<th>Number of observations</th>
<th>3.B</th>
<th>4.B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity $\eta(u), \lambda$, ($t$-stat.=$0$)*, ($t$-stat.=$1$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disaggregated variables replacing $W+w$ and $V$ in Columns 3.A and 4.A of Table 13.A are shown.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(W+w)$ Total tons (W+$w$,v)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
</tr>
<tr>
<td></td>
<td>TGV: high speed trains</td>
<td>0.097</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>GL: classic intercity trains (Corail)</td>
<td>0.004</td>
<td>0.588</td>
</tr>
<tr>
<td></td>
<td>TER: regional trains</td>
<td>0.052</td>
<td>0.588</td>
</tr>
<tr>
<td></td>
<td>IdF: Ile-de-France trains</td>
<td>0.004</td>
<td>0.588</td>
</tr>
<tr>
<td></td>
<td>FRT: freight trains</td>
<td>0.089</td>
<td>0.188</td>
</tr>
<tr>
<td></td>
<td>HLP: locomotives</td>
<td>0.051</td>
<td>0.188</td>
</tr>
<tr>
<td></td>
<td>Time since last regeneration (ageral)</td>
<td>0.017</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>($r=0$)</td>
<td>($r=0$)</td>
<td>($r=0$)</td>
</tr>
</tbody>
</table>

#### Log likelihood

| Number of $\beta_k$ estimated | 19 | 19 |
| Number of $\lambda_k$ estimated | 3 | 3 |
| Difference in degrees of freedom | 0 | 1 |
| Variant run number | 92 | 96 |