



HAL
open science

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

Marc Gaudry, Bernard Lapeyre, Emile Quinet

► **To cite this version:**

Marc Gaudry, Bernard Lapeyre, Emile Quinet. Infrastructure maintenance, regeneration and service quality economics: A rail example. 2015. <halshs-00559637v3>

HAL Id: halshs-00559637

<https://shs.hal.science/halshs-00559637v3>

Preprint submitted on 9 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS

WORKING PAPER N° 2011 – 03

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

**Marc Gaudry
Bernard Lapeyre
Émile Quinet**

JEL Codes: C61, D61, D92, H54, R41, R48

Keywords:



PARIS-JOURDAN SCIENCES ÉCONOMIQUES

48, Bd JOURDAN – E.N.S. – 75014 PARIS

TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10

www.pse.ens.fr

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 Ikononov 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^o 2011-03

Version 1, January 2011; Version 21, 27th February 2015

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 part 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.

Résumé

La présente analyse utilise un cadre formalisé d'optimisation de la maintenance d'une infrastructure, les activités en question comportant d'une part des opérations continues d'entretien courant et d'autre part, à intervalles espacés, des opérations de régénération de l'infrastructure. On prend aussi en compte le fait que la fonction de coût de maintenance comporte comme produits non seulement les niveaux de trafic mais aussi la qualité de service fournie aux usagers des voies. Même si la formulation est articulée pour des infrastructures ferroviaires, nombre de ses considérations centrales s'appliquent également à d'autres types d'infrastructures, routières notamment. On supposera à titre principal qu'on est en avenir certain, et alors le problème est traité en utilisant les équations de Pontryagin, mais on explorera aussi la situation d'incertitude par le biais de simulations numériques utilisant les procédures Hamilton-Bellman-Jacobi. On en déduit une expression du coût marginal social de maintenance qui comprend comme cas particulier emboîté le résultat obtenu par Newbery en 1988 dans un contexte routier.

On teste économétriquement les équations du modèle à l'aide de transformations Box-Cox en tenant compte de l'autocorrélation dirigée (y compris spatiale) des erreurs ; et on compare les résultats, estimés à partir de données sur les tronçons du réseau ferroviaire national français, aux règles de décision sur la régénération appliquées en France. En particulier, le profil temporel en U renversé des dépenses d'entretien prédit par la formulation choisie est vérifié pour la partie du réseau qui donne lieu à de la régénération et, pour la longue phase centrale du profil triphasé, les équations explicatives des dépenses totales d'entretien courant et de la qualité de voie offerte qui sont estimées pour l'ensemble du réseau comprennent bien les variables de service cible et de correction de la trajectoire du service introduites par le cadre d'optimisation conjointe adopté, variables qui s'avèrent significatives.

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	7
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. Effects of tamping frequency	6
Figure 2. Typical evolution of service quality and maintenance expense for $T=35$	9
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	11
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 common 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 maintenances, and is only weakly related to extant technical knowledge on infrastructure degradation and its consequences for service quality.

We build, at the juncture of these two research directions, a model embodying some core knowledge of degradation mechanisms but simultaneously linking policies on 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; Sathay & Madanat, 2011). 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. Tsunokawa & Schofer (1994) use optimal control theory to deduce optimal pavement policy, but do not include in their analysis any reference to quality of service and deliberately adopt a goal of optimal technical policy. Friesz & Fernandez (1979) propose a model which takes into account quality of service, but in an environment of wide uncertainty, and with a view to maintenance policy optimization.

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 analyses¹ linking annual maintenance expenditures to traffic and technical track characteristics for five² 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), in contrast with Rust's (1987) 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.

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)³ 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 T_i to be defined, that fully restore the infrastructure to its initial un-degraded quality state.

This assumption implies that the current maintenance can be approximated by a continuous function, while maintenance appears at discrete points of time. This approximation implies that current maintenance operations are numerous within the time unit, and for the average length of the link under consideration; we will see that, owing to the data used for calibration, these conditions are fulfilled. Furthermore, Tsunokawa & Schofer (1994) have shown that a series of discrete maintenance operations and the corresponding sawtooth path curve of the quality of service can be approximated by a continuous function.

We for instance have notably in mind the lumpy saw-like process whereby tampings reduce the rate of track settlement over time, as in Figure 1.A from Ongaro & Iwnicki (2009, Figure 2), and such successive current maintenance operations (immediately followed by preventive⁴ rail grinding) to maintain a target service level (say a combination of comfort and safety), reduce the rate of track service quality degradation (as measured by index NL in Figure 1.B from Antoni *et al.*, 2009) when effected at an increasing frequency over the lifetime of a ballast. Such maintenance is known to decrease by about half the slope of the

¹ In particular, Box-Cox transformations are applied by all to CES, and by some to other specifications.

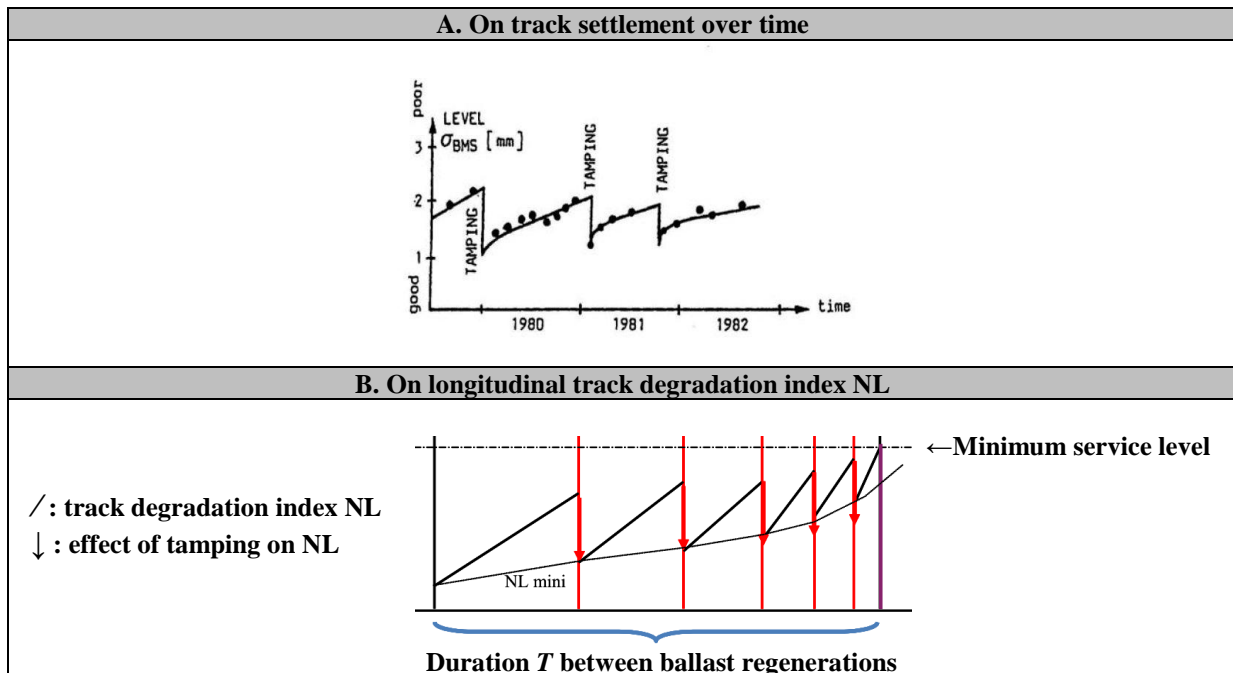
² Austria, France, Great Britain, Sweden, Switzerland.

³ Like Newbery (*op. cit.*), they ignore any pavement life prolongation caused by non regenerative maintenance.

⁴ Grindings done between tampings will have no effect on such longitudinal or transversal degradation indices.

degradation curve and to prolong T_i , a key effect of maintenance observed⁵ on service S and on T_i . Indeed, “maintenance causes continuous changes in quality (deterioration)”, as long ago emphasized in a housing maintenance policy model by Arnott *et al.* (1983).

Figure 1. Effects of tamping frequency



The service quality supplied by the infrastructure at time t is designated $S(t)$. Its derivative at time t depends on the current maintenance $u(t)$, on track segment technical characteristics K , on the cumulated traffic borne by the track segment from the last renewal to present, $Q(t)$, on the level of the service quality $S(t)$, on the current traffic level $q(t)$ and on time elapsed since the last renewal t .

The objective function of an infrastructure manager assumed to maximize the collective surplus is then defined, at interest rate j , as the present value of services supplied by the infrastructure minus maintenance expenditures — both current ones for each period dt and regeneration ones D incurred at intervals T_i . We also make the simplifying assumption that traffic $q(t)$ is perfectly known and constant over time and denoted by q . These assumptions go beyond mere tractability and stem notably from the fact that, in many situations found in developed countries, the trend of traffic is almost *nil* and shows little volatility from year to year (at a sufficient level of aggregation). Under such assumptions, the optimal regenerations will be regularly spaced at some interval T and the stated problem can be expressed as:

$$\begin{aligned}
 (1) \quad & \left. \begin{aligned}
 M^* = \text{Max}_{u(t), T} \left\{ \sum_{n=0}^{n=\infty} \left[\int_{nT}^{(n+1)T} [-u(t) + a * q * S^{-\lambda}] e^{-jt} dt - D e^{-j(n+1)T} \right] e^{-jnT} \right\} \\
 (2) \quad & \left. \begin{aligned}
 \text{or, denoting } \int_0^T [-u(t) + a * q * S^{-\lambda}] e^{-jt} dt \text{ by } J(u(t), T) : \\
 (3) \quad & \left. \begin{aligned}
 M^* = \text{Max}_{u(t), T} \left\{ \frac{J(u(t), T) - D e^{-jT}}{1 - e^{-jT}} \right\} \\
 (4) \quad & \left. \begin{aligned}
 \frac{dS(t)}{dt} = h(K, Q(t), t) * u(t) - f(K, Q(t), q(t), t) \\
 (5) \quad & \left. \begin{aligned}
 Q(t) = \int_{T_i}^t q(s) ds
 \end{aligned} \right\}, \quad \text{for } T_i < t < T_{i+1}.
 \end{aligned}
 \end{aligned}
 \end{aligned}
 \end{aligned}
 \end{aligned}
 \end{aligned}$$

⁵ An analysis of the evolution of service quality, as measured here for 11 years over 581 segments, shows that, for the overwhelming majority of segments where maintenance has produced large year-to-year service improvements, the speed of degradation, measured before and after over 3 years, is considerably reduced.

Clearly, we make several simplifying assumptions: first, we assume that the rate of change of service quality is a linear function of maintenance u ; second, we assume that the utility provided by service quality exhibits risk aversion to account for the fact that it stands for the possibility of rupture or accident. Third, we assume that this rate of change of service quality does not depend on service quality S itself. This assumption is perhaps not intuitive, as another specification in which the rate of change is homogenous of degree -1 in S (building in the fact that dS/S is independent of S) seems more intuitive; but such a specification is in fact equivalent to the present one as may be noted by changing S into $\log(S)$ and $aS^{-\lambda}$ into $a \exp(-\lambda S)$, another quite sensible risk averse utility function.

The following assumptions are made concerning the derivatives of functions $h(\cdot)$ and $f(\cdot)$:

- (i) $h(\cdot)$ is positive and decreasing in $Q(t)$ and t : for a given quality of service, the efficiency (physical productivity) of one Euro spent on maintenance decreases with cumulative traffic, represented below by cumulative tonnage $W(t)$ or cumulative number of trains $N(t)$, and time elapsed since the last renewal;
- (ii) $f(\cdot)$ is positive and increasing in $Q(t)$, $q(t)$ and t (as a group called aging factors in Frame 2 below): for a given quality of service, quality degradation increases with cumulative traffic $Q(t)$, current traffic $q(t)$ consisting in c traffic categories q_i , and t , time elapsed since the last renewal;
- (iii) $u(t)$ 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 work and the cost of regeneration, limit feasible maintenance per unit of time).

To solve this programme, 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. The Hamiltonian may be written, introducing the co-state function $y(t)$, as:

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

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

$$(7) \quad \begin{aligned} \text{Max}_u H &= \text{Max}_u \{ u(t) * [h(K, Q, t) * y(t) - e^{-jt}] \} \\ H_s + \dot{y} &= 0 \\ H_y &= dS/dt \end{aligned}$$

with the two-limit constraint on $u(t)$, located in the closed interval $[0, m]$, 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: $h(K, Q, t) * y(t) - e^{-jt} < 0$. With $u(t)=0$ over this first interval, we have:

- $S(t)$ found by integration of $dS/dt = h(K, Q, t) * u(t) - f(K, Q, q, t)$: 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 $y(t) = y(t_f) - \int_{t_f}^t \alpha q \lambda S(t)^{-\lambda-1} e^{-jv} dv$.

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

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

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

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

$$\dot{y} + H_s = \frac{-j e^{-jt}}{h(K, Q, t)} - \frac{\partial h / \partial t + [\partial h / \partial Q][\partial Q / \partial t]}{h^2(K, Q, t)} e^{-jt} + \alpha q \lambda S_c(t)^{-\lambda-1-jt} = 0,$$

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

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

taking into account the above mentioned breakdown of q among traffic categories q_i , applying a logarithmic transformation⁶ and collecting terms :

$$(8) \quad \text{Log}[S_c(t)] = \frac{1}{\lambda + 1} \left\{ \text{Log} \left(\sum_{i=1}^{i=c} \alpha_i q_i \right) + \text{Log} \left(\frac{\lambda}{j} \right) - \text{Log} \left[\frac{1}{h} + \frac{h'}{j h^2} \right] \right\},$$

where h' notably contains $\partial Q / \partial t$, *i.e.* current traffic $q(t)$. Then $u(t)$ is 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:

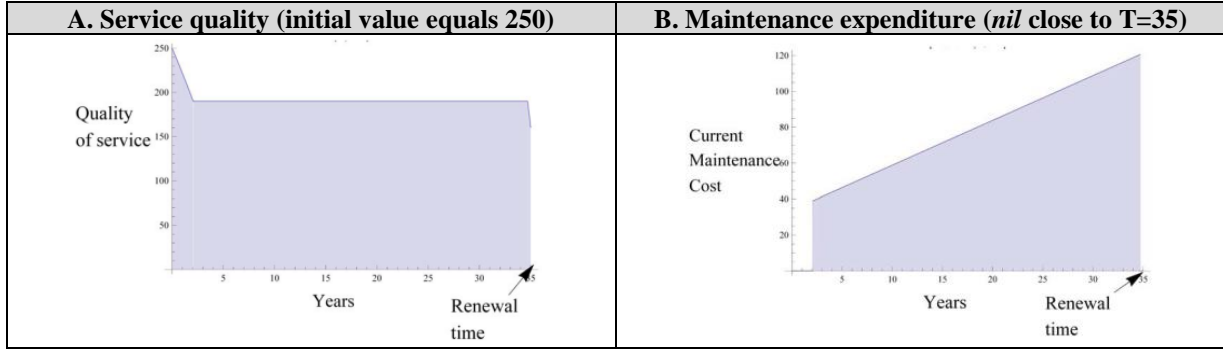
- either one stays in Phase A because $y(t)$ satisfies the corresponding inequality restriction. The resulting optimal policy is then to perform no current maintenance and to regenerate periodically, as in the road case studied by Newbery (1988a, 1988b);
- or, at a certain instant t_f , one has $y(t_f) = [1/h(K, Q(t_f), t_f)] e^{-jt_f}$, 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 $S_c(t)$.

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.

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.

⁶ 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 $h(\cdot)$.

Figure 2. Typical evolution of service quality and maintenance expense for T=35



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:

$$(11) \quad h(K, Q, t) * y(t_f) = e^{-jt_f}$$

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

$$(12) \quad e^{-jt_f} = \int_{t_f}^T \alpha q \lambda S(t)^{-\lambda-1} e^{-jv} dv,$$

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 $\underset{u}{Max}\{J(u(t), T)\} = J^*(T)$, the optimal duration T is that which maximizes

$$(13) \quad M(T) = \frac{J^*(T) - De^{-jT}}{1 - e^{-jT}}.$$

It is shown in Appendix 1 that the second derivative of $M(T)$ with respect to T is negative, so there is an extremum and it is a maximum.

3.3. Consequences for optimal intertemporal infrastructure pricing

In the previous section we have made a first assumption 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. The relation between traffic q , service quality S and infrastructure charges π are assumed to have the following form:

$$(14) \quad q = l(\pi - \alpha S^{-\lambda}) = l(p), \quad \text{or} \quad p = \pi - \alpha S^{-\lambda} = p(q),$$

where π 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 π 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 π and $S(t)$ can (and effectively do) vary.

Under these assumptions, the optimization of charges involves the maximization with respect to q of the discounted surplus (SC) arising from managing the infrastructure:

$$(15) \quad SC(q, u(t), T) = \frac{1}{1 - e^{-jT^*}} \left\{ \int_0^{T^*} [(\int_0^q (p(v) - aS^{-\lambda}) dv) - u(t)] e^{-jt} dt - De^{-jT^*} \right\} = \frac{1}{j} \int_0^q p(v) dv + M(q, u(t), T)$$

with

$$(16) \quad M(q, u(t), T) = \sum_{i=0}^{i=\infty} \left[\int_{T_i}^{T_{i+1}} [-u(t) + a * q * S^{-\lambda}] e^{-jt} dt - De^{-jT_{i+1}} \right] e^{-jT_i} = \left\{ [J(u(t), T) - De^{-jT}] \frac{1}{1 - e^{-jT}} \right\}$$

Using the results of the previous section, and writing $M^*(q) = \underset{u(t), T}{Max} [M(q, u(t), T)]$, the maximization of SC with respect to $u(t)$ and T gives $SC^*(q) = (1/j) \int_0^q p(v) dv + M^*(q)$ and the maximization of $SC^*(q)$ with respect to q yields p and q through:

$$(17) \quad p(q) = -j \frac{dM^*(q)}{dq}.$$

One then calculates infrastructure charges $\pi(t)$ varying in time and equal to:

$$(18) \quad \pi(t) = p(q) + aS(t, q)^{-\lambda} = -j \frac{dM^*(q)}{dq} + aS(t, q)^{-\lambda}.$$

Overall, the generalized cost to the user p 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 δq , one considers only variations in $J^*(T, q)$, the duration of inter-regeneration T being held constant. Then:

$$(19) \quad \frac{dM^*(q)}{dq} = \frac{1}{1 - e^{-jT}} \frac{\partial \left\{ \int_0^T [-u(t) - aS^{-\lambda} q] e^{-jt} dt \right\}}{\partial q},$$

In the derivation of the integral, 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:

$$(20) \quad \frac{\partial J}{\partial q} = \int_0^{t_i} \left[\alpha S(t, q)^{-\lambda} + \alpha q \frac{\partial (S(t, q)^{-\lambda})}{\partial q} \right] e^{-jt} dt,$$

with a similar relationship for the instant t_f .

By contrast, in the central cruising Phase B, the derivation of the integral — call it J — is:

$$\frac{dJ}{dq} = \int_{t_i}^{t_f} \alpha S(t, q)^{-\lambda} e^{-jt} dt + \int_{t_i}^{t_f} \alpha q \frac{\partial S(t, q)^{-\lambda}}{\partial q} e^{-jt} dt - \int_{t_i}^{t_f} \frac{\partial u(t, q)}{\partial q} e^{-jt} dt - u(t_i, q) \frac{\partial t_i}{\partial q} + u(t_f, q) \frac{\partial t_f}{\partial q}.$$

This expression, which easily leads to the calculation of the toll π , does comprise $\partial u/\partial q$, but also 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's (1972) effect for service frequency.

The difference between infrastructure marginal cost and optimal charges can be significant, as simulation cases of Table 1 indicate. For low traffic levels, infrastructure marginal cost charges are higher than optimal charges, indeed a sort of Mohring effect in that higher traffic incites the manager to improve service quality, a motivation that progressively decreases. 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.

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

Traffic	Revenue from standard marginal charge	Revenue from new optimal charge
400	1,509	1,630
300	1.080	1.140
200	0,583	0,588
100	0,173	0,155
50	0,057	0,038

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:

$$(21) \quad dS(t) = h(K, Q, t)[u(t) - f(K, Q, q, t)]dt + \sigma dz$$

where the random variable dz denotes a classical Brownian motion, as in Hausmann & 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:

$$(22) \quad -J_t = \text{Max}_u \{ [-u(t) - \alpha q e^{-\lambda S(t)}] e^{-j t} + J_s [h(K, Q, t)u(t) - f(K, Q, q, t)] + \sigma^2 J_{SS} \}.$$

Note that such a simulation procedure, with parameters and functions found in Frame 1, where the operator's attitude to S is 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 σ to zero: we have consequently compared the results of simulations to the solution obtained analytically by the Pontryagin procedure.

Frame 1. Parameters and functions used for the simulations

Number of years between 2 renewals: $T=20$.	Initial quality of service: $S(0)=3,3$.
Limit values of current maintenance $u(t)$: $0; 0,2$.	Yearly traffic normalized to: $q=100$.
Discount rate: $j=0,04$.	
Functions for: $dS(t) = h(K, Q, t)[u(t) - f(K, Q, q, t)]dt + \sigma dz$:	$h(t)=7$; $f(q, t)=[(0,25+0,16(q/400)^2)*(1+0,04(q/400)t)]/[7e^{-0,02}]$.
Standard deviation of Wiener random variable: $\sigma=0$; then $\sigma=0,1$; then $\sigma=0,2$. [the measured standard deviation of the observed quality of service is approximately 0,08].	
Parameters of the function expressing the value of the quality of service $-\alpha e^{-\lambda S}$: $\alpha=10$; $\lambda=10$.	
<i>Note that technical parameter values denoted by K earlier are taken into account in the values selected.</i>	

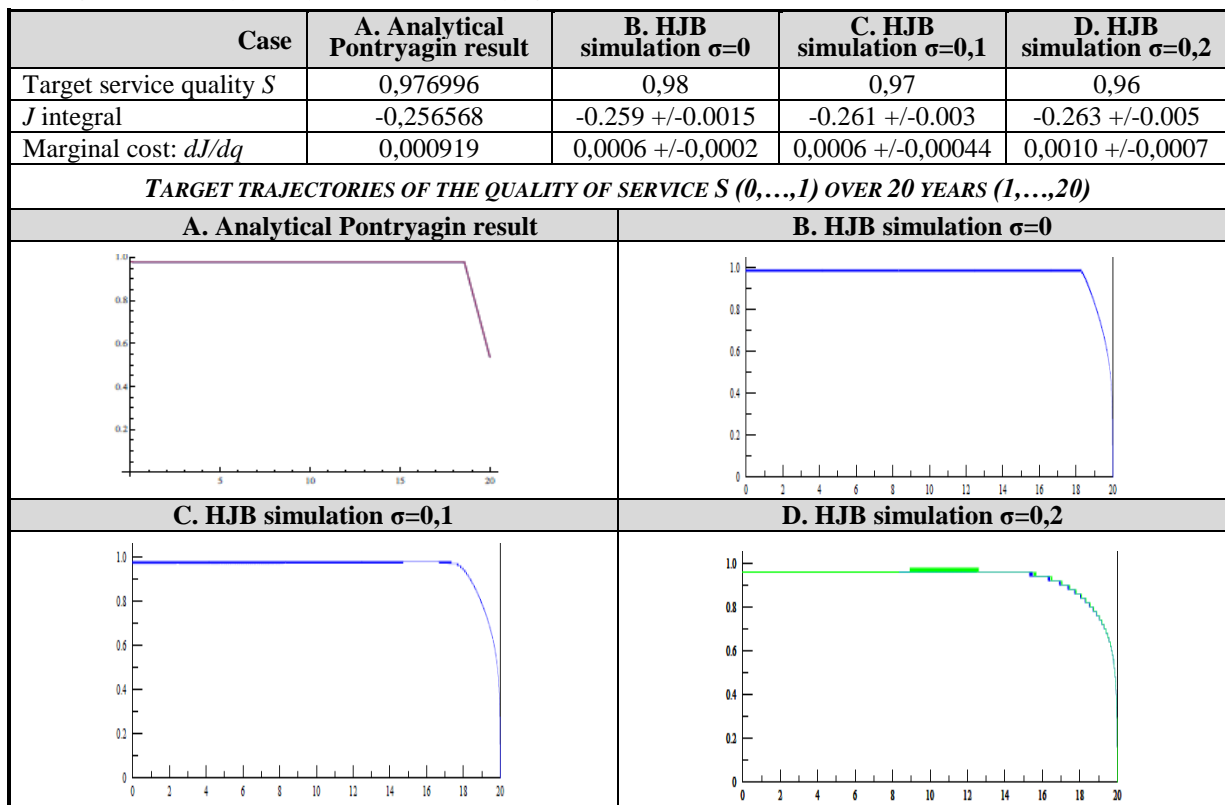
The solution values, shown in Figure 3, are drawn from a large number of simulations providing similar results. Those are based on sensible values of the parameters corresponding

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.

Let us note first that the simulation procedure brings out, not the real, but 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.

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



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 provision⁷ by rail firms (*e.g.* Martland, 1992; Robert *et al.*, 1997) or in the literature on infrastructure maintenance cost models estimated from track section data⁸.

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

$$(23-A) \quad u_t = f(K(h_0, e_0), q(n, w), S(h))_t,$$

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, fleshes 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 performance⁹ indicator $S(h)_t \equiv S_t$. That “extended classical” specification nests into:

$$(23-B) \quad u_t = f(K(h_0, e_0), q(n, w), Q(N, W), t, S(h))_t,$$

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.

⁷ 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.

⁸ 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.

⁹ 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.

In an analysis of pruned databases for 1999 and 2007, we first show that high standing track segments belonging to UIC¹⁰ 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_t) and realized Service provision (S_t) 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.

Table 2. Estimation sequence for Maintenance and Service equations

Current maintenance cost		Service provision
1999 Base	2007 Base	2007 Base
a. Test existence of 3 phases on full sample (978 obs.). Find Phase A & C segments.	a. Test existence of 3 phases on full sample (700 obs.). Find Phase A & C segments.	a. Test existence of 3 phases on full sample (700 obs.). Find Phase A & C segments.
	▼▼	
b. Retain cruising Phase B censored subsample	b. Retain cruising Phase B censored subsample (675 obs.)	b. Retain cruising Phase B censored subsample (675 obs.).
	▶	▼▼
	c. Estimate cruising Phase B Maintenance Cost u equation making use of variables based on Target Service S_c estimates (580 observations)	Estimate embedded Target Service S_c (580 obs./year)
	▶	▼▼
		c. Estimate cruising Phase B Service Supply ΔS equation using S_c and u estimates

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

¹⁰ 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.

data and because the 1999 base consists in operational “traffic” segments and the 2007 base in 49% longer administrative “cost” segments covering more than one traffic segment.

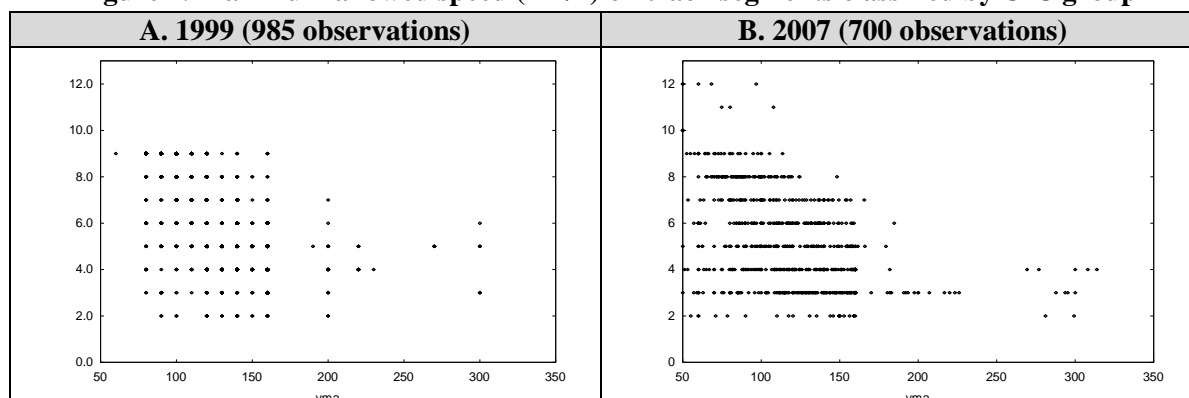
Table 3. Mean values of principal variables available by track segment

Variables		1999 (sample size 985)		2007 (sample size 700)	
u_t	Current maintenance expenses				
	total cost per km (current Euros)	68 899		52 432	
	<i>surveillance</i>	--		14 821	
	<i>maintenance</i>	--		37 611	
e₀	Technical state variables K				
	length of segment (meters)	18 991		28 219	
	length of all tracks (meters)	by number of tracks		46 186	
	electrified or not; tension (1,5 or 25 kV) ¹	yes		yes	
	number of switches per segment	21,41		25,48	
h₀	Initial standing				
	maximum allowed speed (km/h)	127,77		114,69	
	UIC group classification	(reconstructed)		observed	
	high speed rail line	yes		yes	
	suburban line	yes		yes	
h_t	Current condition				
	average age of rails (years)	26,59		30,56	
	average age of sleepers (years)	26,87		28,07	
S_t	Service quality of the track				
	NL index of longitudinal track rectitude (mm)	--		1,451 ²	
q_t	Traffic per day: trains and gross tons	trains	weight (per train)	trains³	weight³ (per train)
	GL : long distance passenger trains (VFE)	6,04	3 151 (522)	17,99	10 255 (570)
	<i>TGV: high speed trains</i>	----	----	9,25	5 612 (601)
	<i>Classic intercity trains (Corail)</i>	----	----	8,74	4 643 (531)
	TER : regional passenger trains	5,09	1 115 (219)	24,44	5 559 (228)
	IdF : Île-de-France passenger trains	5,99	2 378 (397)	16,21	6 215 (384)
	Fret : freight trains	6,10	6 403 (1049)	13,70	15 484 (1131)
	HLP : locomotives	1,40	138 (99)	2,84	275 (97)
	Total for the six categories of trains	24,76	13 185 (533)	75,16	37 788 (503)

¹Out of a total network of about 30 000 km, 11 582 km (including 1 884 km of TGV lines) have alternating tension of 25 kV and 5 863 km have continuous tension of 1,5 kV; 126 km are electrified otherwise (third rail, etc.). ²Available over the period 2000-2010 for a subset of 608 observations. ³Available over the period 1995-2007 for all 700 observations.

This difference in segment length shows up clearly in Figure 4 where *maximum allowed speed* (*v_{ma}*) *v*, the main initial infrastructure quality indicator *h₀* 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



Econometric strategy. The general econometric specification adopted, readily applicable to our cross-sectional data sets on segments (I, \dots, t, \dots, T), is the Box-Cox model with a non-spherical distribution of residuals:

$$(24-A) \quad y_t^{(\lambda_y)} = \beta_0 + \sum_k \beta_k X_{kt}^{(\lambda_k)} + u_t,$$

$$(24-B) \quad u_t = \left[\exp\left(\sum_m \delta_m Z_{mt}^{(\lambda_m)}\right) \right]^{1/2} v_t,$$

$$(24-C) \quad v_t = \sum_{\ell=1}^{\ell=2} \rho_\ell \left(\sum_{n=1}^{n=T} \tilde{r}_{\ell,m} v_n \right) + w_t.$$

where, in the first two equations, the Box-Cox transformations (BCT) applicable to any strictly positive¹¹ variable Var_v is commonly defined without Tukey's shift parameter as:

$$(24-D) \quad Var_v^{(\lambda_v)} \equiv \begin{cases} [(Var_v)^{\lambda_v} - 1]/\lambda_v, & \lambda_v \neq 0, \\ \ln(Var_v) & , \quad \lambda_v \rightarrow 0, \end{cases}$$

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

$$(24-E) \quad v = \sum_{\ell=1}^{\ell=2} \rho_\ell \tilde{R}_\ell v + w, \text{ with } \tilde{R}_\ell = \pi_\ell \left[I - (1 - \pi_\ell) \bar{R}_\ell \right]^{-1} \bar{R}_\ell, \quad (0 < \pi_\ell \leq 1),$$

where the new proximity parameter π_ℓ measures the relative influence of “near” and “distant” neighbors of R_ℓ , as soon discussed below.

BCT forms. Note that (24-A) contains the necessary intercept (Schlesselman, 1971) and that the X_k 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:

$$(24-F) \quad Q(W_t) = \beta_{Q1} W_t^{(\lambda_{Q1})} + \beta_{Q2} W_t^{(\lambda_{Q2})},$$

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 β_1 and β_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 [$\lambda_1 = 1; \lambda_2 \neq 2$] and on the proper symmetric quadratic [$\lambda_1 = 1; \lambda_2 = 2$] special case, both nested in (24-F).

Table 4. Sign conditions for a maximum or a minimum with two BCT on the same X_k

CASE	β_1	β_2	$\lambda_1 - \lambda_2$	$\beta_1(\lambda_1 - \lambda_2)$ or $\beta_2(\lambda_2 - \lambda_1)$
\cap Maximum 1	+	-	-	-
\cup Minimum 1	+	-	+	+
\cap Maximum 2	-	+	+	-
\cup Minimum 2	-	+	-	+

¹¹ 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_t or w_t and correct BCT forms, particularly for the dependent variable y_t , 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 δ_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_t), where the optimal BCT on the 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 \tilde{R}_ℓ in (24-E) results from three steps. In the first, a square matrix R_ℓ is defined with typical element $r_{\ell,m}=1$, to express an hypothesized correlation between any two residuals v_t and v_n (and $r_{\ell,m}=0$ 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* choice¹² 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 \bar{R}_ℓ guarantees a convex likelihood function over the stable unit interval of ρ_ℓ . 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_ℓ 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_ℓ , 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 \bar{R}_ℓ generate a sequence of contiguity matrices ($\bar{R}_\ell^2, \dots, \bar{R}_\ell^c, \dots$) which define degrees of neighborliness or proximity, with \bar{R}_ℓ^2 denoting neighbors of neighbors, and \bar{R}_ℓ^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 $\tilde{R}_\ell = \pi_\ell [I - (1 - \pi_\ell)\bar{R}_\ell]^{-1} \bar{R}_\ell$, where the new proximity parameter π_ℓ allows for endogenization of the relative importance of near and distant effects. If $\pi_\ell=1$, $\tilde{R}_\ell = \bar{R}_\ell$, in which case only the adjacent neighbors have an impact on the correlations among the residuals selected by the “residue impact criterion” R_ℓ , exactly as in the classical (Ord, 1975; Cliff & Ord, 1981) case. By contrast, as $\pi_\ell \rightarrow 0$, the near effect is reduced to a minimum in favor of the distant effect. In this R-Koyck formulation, the parameter π_ℓ therefore generally weighs 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_t) by segment, three specifications of the R_ℓ matrix and, singly and jointly, two distributed processes π_ℓ as well.

¹² 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 0, implying the possibility of a mass point at $-1/\lambda$ when $\lambda > 0$, or in practice at a very small number ε) 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 $y \equiv (y_1, \dots, y_t, \dots, y_T)$ reduces, in the absence of limit observations in the sample, to the maximand Λ actually proposed by Box and Cox (1964) themselves:

$$(24-G) \quad \Lambda = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{w_t^2}{2\sigma_w^2}\right) \left| \frac{\partial w_t}{\partial y_t} \right|,$$

where, under the assumptions of normality and constancy of the variance σ_w^2 of the independent error term w_t of zero mean, the Jacobian of the transformation from w_t to the observed y_t is $|\partial w_t / \partial y_t| = y_t^{\lambda y_t - 1}$. With multiple BCT in (24-A), finding the global maximum of $\ln(\Lambda)$ is required because its concavity need not hold (Kouider & Chen, 1995).

But, even in the absence of limit observations in our final doubly censored samples, any observation y_t in (24-A) is still assumed to be censored both downwards and upwards, within a range $\varepsilon \leq y_t \leq v$, where ε and v are respectively the strictly positive lower and upper censoring points common to all observations. In such and forthcoming generalizations of the Tobit model¹³ to a doubly censored dependent variable, the expected value of y_t , given that $\phi(w)$ is the normal density function of w with 0 mean and variance σ_w^2 , is given by:

$$(24-H) \quad E(y_t) = \varepsilon \int_{-\infty}^{w_t(\varepsilon)} \phi(u) du + \int_{w_t(\varepsilon)}^{w_t(v)} y_t \phi(w) dw + v \int_{w_t(v)}^{\infty} \phi(w) dw, \quad (\varepsilon \text{ and } v > 0).$$

Elasticities and t-statistics of the β_k . In all forthcoming tables of results, we report on elasticities of the dependent variable and for brevity neglect the regression coefficients¹⁴ 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_k and X_d , as:

$$(24-I) \quad \eta_t(y, X_k) = \frac{\partial y}{\partial X_k} \cdot \frac{X_k}{y} = \beta_{X_k} \frac{X_k^{\lambda_k}}{y^{\lambda_y}} \Bigg|_{X_{kt}, y_t},$$

$$(24-J) \quad \eta_t(y, X_d) = \beta_{X_d} \frac{\bar{X}_d^+}{y^{\lambda_y}} \Bigg|_{y_t},$$

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.

¹³ 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.

¹⁴ 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 profiles¹⁵ 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 regenerated¹⁶. 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 *maximum authorized speed*, but only approached by the *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 observed¹⁷ 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 number¹⁸. Regression results found in Table 6 are abstracted¹⁹ from equation specifications using all variables listed in Table 5 except for total traffic w and for UIC group dummy variables²⁰ g_k , variables which will intervene later.

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

	$u = f$	[$K(e_0 ;$	$h_0),$	$Q(W_{2-6};$	$W_{7-9}),$	$q(w$ or $w_c),$	t]
1999	cost per km	track & segment lengths; switches	max speed v ; UIC g_k	$[(t)(w)]_{2-6}$	$[(t)(w)]_{7-9}$	$\sum_{c=1}^{c=5} w_c$	gross tons by train type c mean age
2007	cost per km	track & segment lengths; switches	max speed v ; UIC g_k	$(\sum_t w_t)_{2-6}$	$(\sum_t w_t)_{7-9}$	$\sum_{c=1}^{c=6} w_c$	gross tons by train type c mean age

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)²¹ but kept all high traffic segments; for 2007, it excluded apparent high traffic outliers²² 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 *et al.*, 1993). Reported statistics include the total number of BCT estimated by variant²³ and of β_k regression coefficients of each regression variant. Columns A to C document models explaining *total* current maintenance (the sum of surveillance and maintenance) expenses in 1999, and Columns D to E models pertaining only to *surveillance*²⁴ expenses in 2007. Results of reference models A and D assume monotonicity

¹⁵ Phase length tests on Service degradation (step a in the third column of Table 2), will not be carried out.

¹⁶ See Appendix 1 to see how low traffic leads to infrastructure maintained indefinitely and never renewed.

¹⁷ The traffic is observed from 1995 until 2007 and reconstructed beforehand.

¹⁸ There are 18 such segments for 1999 and 17 for 2007.

¹⁹ 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.

²⁰ Groups are defined by linear combinations of tons by train type and maximum allowed speed.

²¹ 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).

²² Such as two suburban traffic values 3 to 7 times the size of the third highest traffic.

²³ The variants found in this table retain up to 7-8 distinct BCT after estimating 13-14.

²⁴ Tests on the maintenance sub-total and on the total failed to detect turning forms, symmetric and asymmetric.

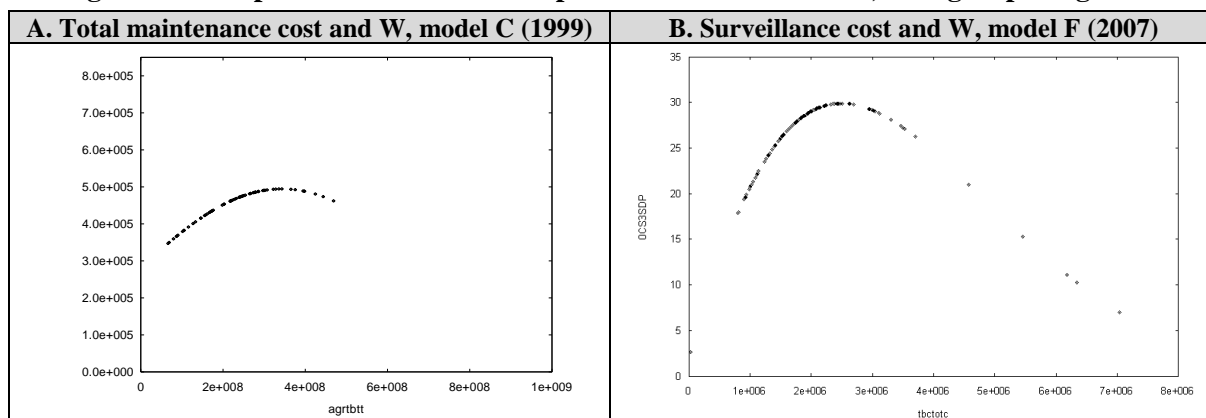
[$\beta_{Q_1} = 0$] and those found in the next two columns respectively assume strictly quadratic [$\lambda_1 = 1; \lambda_2 = 2$] and asymmetric turning [$\lambda_1 = 1; \lambda_2 \neq 2$] special cases of (24-F). We note:

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 β and λ 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, which maintained the spatial representativity of the samples;

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 frequent²⁶ segments, which belong to group 3.

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



Values shown in Figure 5 are not the proper expected values of the dependent variable $E(y_i)$ obtainable from (24-H) but simply the effect of regression component (24-F) on the dependent variable calculated by $y(W)_i = \left\{ 1 + \lambda_y \left[Q(W_i) + \beta_0 + \sum_k \beta_k X_{kr}^{(\lambda_k)} \right] \right\}^{1/\lambda_y}$, $X_k \neq W$, $X_{kr} = \bar{X}_k$, *i.e.* by “unrolling” the transformed dependent variable y in (24-A), all other regressors having been set at their sample mean values²⁷. In these *ad hoc* illustrations of asymmetric turning effects, the estimated value of the BCT on the dependent variable λ_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 invariance²⁸ 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 sufficient²⁹ to reestablish monotonicity for 1999 but not for 2007 where a modified³⁰ 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_k variables transformed by BCT, is never invertible³¹ in the presence of exact collinearity.

Table 6. Phase tests with cumulative traffic variables, classic line network, 1999 & 2007

Cost per km of segment (u):	1999 sample (obs. 12-978)			2007 sample (obs. 1-683)		
	Total maintenance cost			Surveillance cost		
Model	A	B	C	D	E	F
<i>Specified form of W_{2-6}</i>	<i>Monotonic</i>	<i>Quadratic</i>	<i>Asymmetric</i>	<i>Monotonic</i>	<i>Quadratic</i>	<i>Asymmetric</i>
W_{2-6} ($\eta_1(u)$)		0,258	0,267		0,302	-0323
(first term; first BCT)		(3,03)	(5,28)		(3,10)	(-2,78)
λ_1 (t^{**})		1.00	1.00		1,00	1,00
		--	--		--	--
W_{2-6} ($\eta_2(u)$)	0,033	-0,062	-0,070	0,171	-0,094	0,498
(second term; second BCT)	(1,89)	(-3,36)	(-5,55)	(4,03)	(-2,66)	(3,75)
λ_2 (t^{**})	0,24	2.00	1,94	0,11	2,00	0,564
	(0,51;-1,53)	--	(3,41;1,65)	(0,41;-3,30)	--	(1,13;-0,88)
Estimated form of W_{2-6}	/	∩	asymmetric ∩	/	∩	asymmetric ∩
W_{7-9} ($\eta_k(u)$)	0,000	0,001	0,001	0,122	0,010	0,058
(one BCT)	(2,59)	(0,04)	(3,63)	(3,35)	(1,10)	(2,52)
λ_k (t^{**})	3,40	3,39	3,40	0,12	1,00	0,45
	(1,65;1,16)	(2,49;1,75)	(2,41;1,70)	(0,34;-2,52)	(0,56;0,00)	(0,71;-0,88)
Estimated form of W_{7-9}	/	/	/	/	/	/
Log likelihood	-12925,054	-12905.205	-12905.***	-7171,25	-7172,21	-7170,59
Number of β_k estimated	20	21	21	17	18	18
Number of λ_k estimated	8	7	8	8	7	8
Difference in D.F.	0	0	1	0	0	1
Variant run number	77	84	76	6	3	0
η : Elasticity of the dependent y variable with respect to X , calculated by (24-I). t^* : Student's t statistic of β coefficient with respect to 0, calculated conditionally upon the estimated value of λ . t^{**} : Student's t statistic of λ 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 traffics 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_{2-6}^* variable replacing W_{2-6} 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_k 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 points³²: 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 observations³³ 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 $u_c(t)$ is close to the level of the certain situation. We translate this last point in discrete time by extending (9):

$$(23-C) \quad u_t = f(K, Q, q, t)_t + \frac{1}{h(K, Q, t)_t} [Sc_t - Sc_{t-1}]_t + a(K, Q, t)_t [S_{t-1} - Sc_{t-1}]_t + \varepsilon_{1t},$$

where the “classical” first term stands for the damages caused by traffic and the two “new” terms bring in the extra maintenance $[Sc_t - Sc_{t-1}]_t$, necessary to maintain the cruise trajectory of the quality of service in the absence of random effects, and the 0 or m correction $[S_{t-1} - Sc_{t-1}]_t$ made by the infrastructure manager to revert to the target cruise trajectory in the presence of deviations from it. The new weight function $a(\cdot)$ 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:

$$(25) \quad S_t - S_{t-1} = h(K, Q, t)_t \cdot [u_t - f(K, Q, q, t)_t] + \varepsilon_{2t},$$

which, after replacement of $u_t - f(K, Q, q, t)_t$ by its value from (23-C), may be written:

$$(26) \quad S_t - S_{t-1} = Sc_t - Sc_{t-1} + \frac{a(K, Q, t)_t}{h(K, Q, t)_t} \{ (S_{t-1} - Sc_{t-1}) + [\varepsilon_{1t}] \} + \varepsilon_{2t},$$

where the random maintenance cost Surprise term $[\varepsilon_{1t} = u_t - E(u_t)]$ can be estimated from (23-C) by using (24-H) to calculate $E(u_t)$.

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:

$$(27-A) \quad \ln[S_c(t)] = \frac{1}{\lambda+1} \left\{ \ln \left(\sum_{i=1}^{i=c} \alpha_i q_i \right) + \ln(\alpha\lambda) - \ln(h) + \ln \left[j - \frac{h'}{h} \right] \right\}.$$

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, effecting 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_t , fitted values $E(S_{c_t})$ and $E(S_{c_{t-1}})$ may then serve to construct required “new” variables $[S_{c_t} - S_{c_{t-1}}]$ and $[S_{t-1} - S_{c_{t-1}}]$ 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 axes³⁴, *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 segments³⁵, 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 RHS term of (27-A) should also be of logarithmic form in accordance with the specification³⁶ 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 $q_n^* = (q_1^{\alpha_1} \dots q_n^{\alpha_n})^{1/N}$ or $q_n^* = \exp \left[\sum_n \alpha_n q_n / N \right]$, that would obviate the use of approximation by shares because the maximization would yield terms like $(\alpha_1 \ln q_1 + \dots + \alpha_n \ln q_n) / N$ or $\sum_n (\alpha_n / N) q_n$.

$\beta_{s_n} \ln(n_t) + \sum_{i=2}^{i=c} \beta_{s_i} s_{n,t}$ 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;

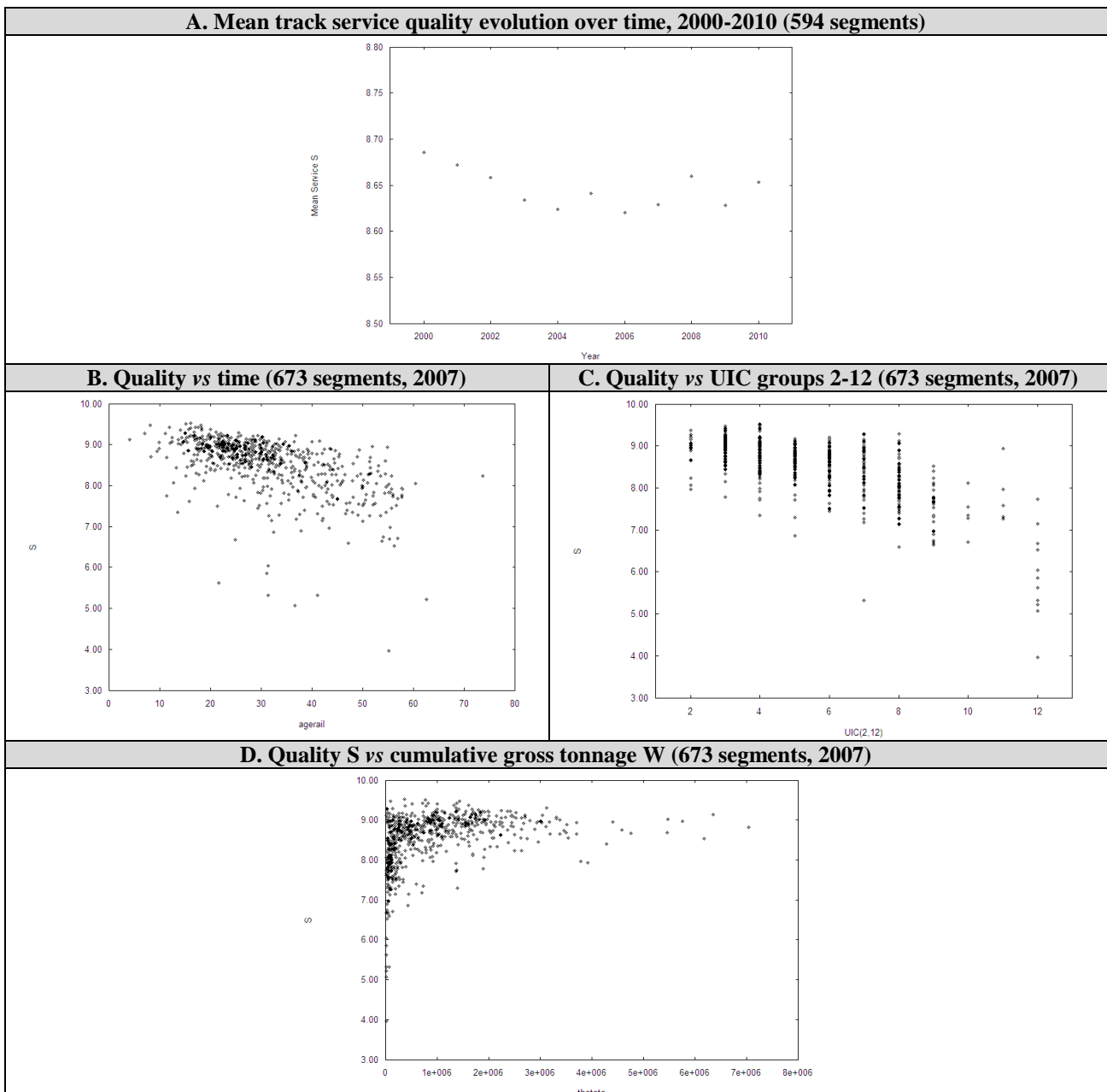
- if $h(\cdot)$, the productivity of a unit of maintenance expenditure, depends on cumulative train tonnage W and is defined with BCT as $\exp\left[\sum_h \beta_h X_{ht}^{(\lambda_h)}\right]$, $\ln[h(K, Q, t)]$ is straightforward and $h'(\cdot)/h(\cdot)$, then equal to $\beta_h q / Q^{1-\lambda_q}$, collapses to $\beta_h q$ (with q specified as current tonnage w) under assumed linearity. The last term $\ln(j - h'/h)$ 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:

$$(27-B) \quad \ln(S_t) = \beta_0 + \beta_{s_n} \ln(n_t) + \sum_{i=2}^{i=c} \beta_{s_i} s_{n_i} + \sum_K \beta_K K_{Kt}^{(\lambda_K)} + \beta_Q W^{(\lambda_w)} + \beta_t t^{(\lambda_t)} + \beta_{h^*} w_t + \varepsilon_{0t},$$

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



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_t [with the due Jacobean of the transformation from $\ln(S_t)$ to S_t] *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 verified³⁷ that the sample did not contain limit observations implying that the assumption of *as-if normality* for ε_{0t} 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 distinct BCT (7 have been estimated) 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_c 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_i 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 settlement over time (*viz.* Fig.1.A), or concerning track longitudinal deterioration (*viz.* Fig.1.B), whether increases in aging factors t , Q and q , the latter of which standing for $h'(\cdot)/h(\cdot)$, increase or decrease (our stated expectation

³⁷ This includes calculating the probability that the dependent variable be close to the lower limit ε in (24-H).

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 $W \approx t \cdot w_{reference\ year}$: 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: $a = \beta_t t^{(\lambda_t)} + \beta_{tw} (t \cdot w)^{(\lambda_{tw})} + \beta_w w^{(\lambda_w)}$, with derivatives:

$$\partial a / \partial t = \beta_t^{\lambda_t - 1} + \beta_{tw} (t \cdot w)^{\lambda_{tw} - 1} w \text{ and } \partial a / \partial w = \beta_{tw} (t \cdot w)^{\lambda_{tw} - 1} t + \beta_w^{\lambda_w - 1}.$$

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

$\partial a / \partial t = \beta_t^{1.23} + \beta_{tw} (t \cdot w)^{0.44} w$ and $\partial a / \partial w = \beta_{tw} (t \cdot w)^{0.44} t$ are clearly always decreasing; and with those of Col. 3A to 3C, the actual derivatives:

$$\partial a / \partial t = \beta_{tw} / t \text{ and } \partial a / \partial w = \beta_{tw} / w \text{ are also always decreasing.}$$

Column		3		3A		3B		3C	
β_k coefficient, t -statistic, λ		$\beta_k(t)$	λ	$\beta_k(t)$	λ	$\beta_k(t)$	λ	$\beta_k(t)$	λ
t	Age of rails ($t=0$) [$t=1$]	-1.17E-04 (-17.88)	2.23 (7.89) [4.34]	-.19E-02 (-18.85)	1	-.19E-02 (-18.80)	1	-.18E-02 (-16.83)	1
W	Total cumulative tons ($t=0$) [$t=1$]	per km of track		per km of track		absolute level		absolute level	
		-.11E-09 (-2.80)	1.44 (5.42) [1.65]	.17E-01 (11.11)	0	.17E-01 (10.96)	0	.16E-01 (8.33)	0
w	Total current tons ($t=0$)	per km of track		per km of track		absolute level		absolute level	
		.29E-06 (1.08)	1	-.35E-06 (-4.92)	1	-.19E-06 (-7.30)	1	-.17E-06 (-6.43)	1
Log likelihood		-1399.82		-1358.21		-1343.79		-1327.69	
Belsley-Kuh-Welsch collinearity index		73		109		127		167	
Variant run number		83		95		92		93	

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 $[E(S_{c_t}) - E(S_{c_{t-1}})]$ and of the Trajectory Correction term $[S_{t-1} - E(S_{c_{t-1}})]$, the “new” explanatory variables in forthcoming Maintenance cost and Service provision equations.

Table 7. Embedded Target Service S_c function results (4060 obs., 2001-2007)

Column		1		2		3		4	
Elasticity $\eta(S)$, λ , t -statistic* of β_k coefficient		$\eta(S)$	λ	$\eta(S)$	λ	$\eta(S)$	λ	$\eta(S)$	λ
β_0	Intercept ($t=0$)	n.a. (334.96)		n.a. (174.17)		n.a. (180.63)		n.a. (176.66)	
Social surplus evaluation component $g(\cdot)$									
n	Total number of trains ($t=0$)	0.021 (31.38)	0	0.020 (29.25)	0	0.020 (29.38)	0	0.020 (30.22)	0
	Long distance GL train share (ref.: TGV) ($t=0$)	-0.006 (-3.68)		-0.007 (-4.46)		-0.007 (-4.47)		-0.007 (-4.85)	
	Regional TER train share (ref.: TGV) ($t=0$)	-0.003 (-4.02)		-0.004 (-4.79)		-0.004 (-4.73)		-0.006 (-7.80)	
	Ile-de-France regional train share (ref.: TGV) ($t=0$)	-0.000 (-0.97)		-0.000 (-1.53)		-0.000 (-1.67)		-0.001 (-4.15)	
	Freight train share (ref.: TGV) ($t=0$)	-0.007 (-5.86)		-0.007 (-6.66)		-0.007 (-6.27)		-0.008 (-7.43)	
	Locomotive only HLP train share (ref.: TGV) ($t=0$)	0.000 (0.79)		-0.000 (-0.28)		-0.001 (-1.19)		-0.000 (-0.66)	
Productivity of maintenance expenses component $h(\cdot)$									
e₀	Segment length ($t=0$)	0.003 (2.56)	1	0.004 (4.57)	0	0.004 (4.13)	0	0.004 (4.12)	0
	Track length ($t=0$)	0.011 (3.80)	1	0.019 (7.18)	0	0.018 (7.02)	0	0.018 (6.75)	0
	Electrified 1500 V (ref.: not electrified) ($t=0$)	0.009 (4.25)		0.008 (3.70)		0.008 (3.71)		0.007 (3.17)	
	Electrified 25000 V (ref.: not electrified) ($t=0$)	0.005 (2.29)		0.004 (1.99)		0.004 (1.76)		0.002 (1.05)	
	Number of switches ($t=0$)	-0.002 (-2.79)	1	-0.002 (-2.91)	1	-0.002 (-3.35)	1	-0.002 (-3.34)	1
h₀	Maximum allowed speed ($t=0$)	0.061 (21.04)	1	0.055 (27.79)	0	0.054 (27.74)	0	0.053 (26.83)	0
	Suburban line (ref.: other line) ($t=0$)	-0.012 (-3.47)		-0.011 (-3.50)		-0.011 (-3.42)		-0.005 (-1.52)	
	High speed rail line (ref.: classic line) ($t=0$)	-0.048 (-10.53)		-0.022 (-4.02)		-0.017 (-3.04)		-0.019 (-3.70)	
W	Cumulative total tons per km of track ($t=0$) [$t=1$]	-0.002 (-3.06)	1	-0.000 (-2.04)	1	-0.001 (-2.80)	1.44 (5.42) [1.65]	-0.001 (-2.83)	1.41 (5.11) [1.48]
t	Time since last regeneration (age of rails) ($t=0$) [$t=1$]	-0.030 (-15.82)	1	-0.031 (-16.69)	1	-0.026 (-17.88)	2.23 (7.89) [4.34]	-0.028 (-17.84)	2.09 (7.83) [4.09]
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
β_d	6 yearly dummies (ref.: 2001)	--		--		--		Note 1	
	Log likelihood	-1476.74		-1415.12		-1399.82		-1384.11	
	Number of β_k	18		18		8		24	
	Number of λ_k estimated	0		0		2		2	
	Degrees of freedom	0		3		5		11	
	Variant run number	88		89		83		90	
Note 1: percentage changes in y ("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 construct³⁹ the needed $E(S_c)$ for the required years, namely 2007, 2006 and 2005, because we need to test a two-period (current and one-year lagged) adjustment speed.

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_k (Erkel-Rousse, 1995), comparisons with values obtained with specifications 3A to 3C argue for choosing Col. 3.

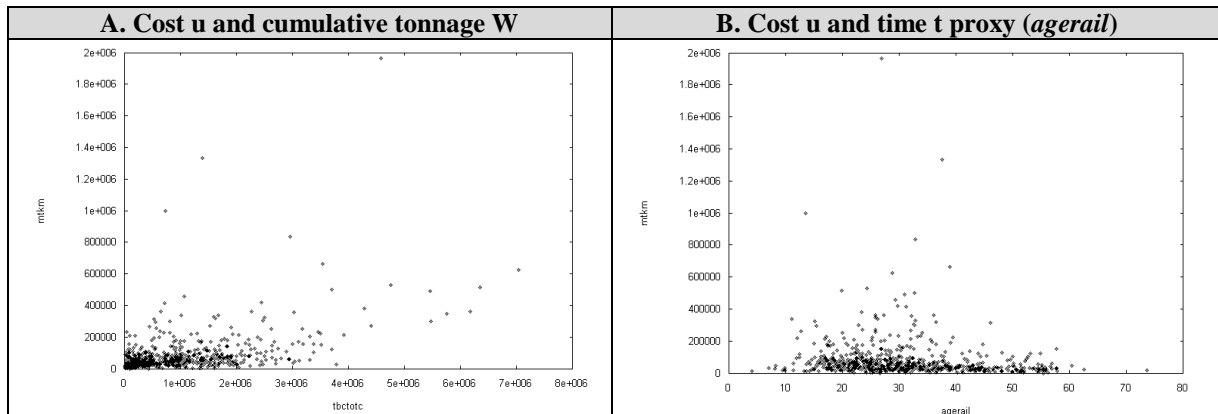
Yet another recursive transport system? These calculations of $E(Sc_t)$ by year are indeed necessary because, for one, the infrastructure manager’s speed of adjustment of Maintenance expenditures u_t in 2007 to changes in the pair of “new” variables, assumed above in (23-C) to be contemporaneous ($t=2007$), 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 set⁴⁰ for 2007, we make two significant changes⁴¹ 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 7-12⁴². Concerning those objective factors, Figure 7 graphs of maintenance against W and t (*agerail*) 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.)



³⁹ Alternatively, given the logarithmic dependent variable in (27-B), $E(Sc_t)$ could have been calculated for any sample year $t = 2001, \dots, 2007$, by collapsing (24-H) to $E(y_t) = k \left\{ \exp \left[\sum_h \beta_h X_{ht}^{(\lambda_h)} \right] \right\}$, where k is the sample mean of the log-normal random variable $\exp(w_t)$ 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.

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 ($c=1, \dots, 6$) 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 aggregates⁴³, 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 10⁴⁴, 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 traffics W_t and w_t are summed to avoid collinearity⁴⁵.

As in other tables, the t -statistics with respect to 0 and 1 are computed from first partial derivatives of the Log of Λ 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 ($W+w$). Up to 7 individual BCT were estimated for Table 9 or Table 10 models before the reduced number of individual values indicated were retained;

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 linear⁴⁶ terms used are the lagged values $[E(Sc_{t-1}) - E(Sc_{t-2})]$ and $[S_{t-2} - E(Sc_{t-2})]$, 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 significant⁴⁷ than contemporaneous values $[E(Sc_t) - E(Sc_{t-1})]$ and $[S_{t-1} - E(Sc_{t-1})]$. But there is a second reason. The lag may help to dodge an identification problem that arises with $[E(Sc_t) - E(Sc_{t-1})]$, a variable which, in results not shown, is not only less significant than $[E(Sc_{t-1}) - E(Sc_{t-2})]$ but also of the wrong (negative) sign. Had $h(\cdot)$ been linear in all terms, we would have had by Equation (8) that $[Sc_t - Sc_{t-1}] \approx [\delta_0 + \gamma q]$ and the two new parameters δ_0 and γ 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 sum⁴⁸, 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_t .

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 value⁴⁹ 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 $\rho_1=0,59$) 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 structure⁵⁰ of R_1 .

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_2 .

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 start⁵¹.

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

⁴⁷ 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.

⁴⁸ The negative sign of $[E(Sc_t) - E(Sc_{t-1})]$ may just correct the adjustment provided by the sum $(W+w)$.

⁴⁹ A value resembling the 0,49 power for the axle weight variable in the UK ORR formula for track damages.

⁵⁰ Consider Region 10, with the smallest number of segments, and the Region 43 with the highest: in matrix R_1 , 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}_1 , the former values become 1/4 and the latter 1/50.

⁵¹ 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.

significantly higher in only two regions. The same exercise carried out here yields hardly any⁵² 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 frequent⁵³.

- 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 axle⁵⁴, un-sprung mass and other technical features absorbing the effect of gross axle weights⁵⁵. 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 $f(\cdot)$. Despite the lack of observed speeds in the database (*viz.* Table 3), we constructed v_c , an operating speed for each of the 6 train types and V , their sum weighted by their respective shares of total accumulated tons [past (W_c), measured at the beginning of each year, and current (w_c)] by train type⁵⁶. 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 addition⁵⁷ 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 addition 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_{t-1} embody too much past history⁵⁸.
- 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 incorrect⁵⁹ and that data on actual speeds are required to yield proper results.

⁵² In tests made on Table 13A, 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 than 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* (v_{ma}) 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.

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

Column		1		2		3.A		4.A	
Elasticity $\eta(u), \lambda, (t\text{-stat.}=0)^*, [t\text{-stat.}=1]$		$\eta(u)$	λ	$\eta(u)$	λ	$\eta(u)$	λ	$\eta(u)$	λ
u	Total maintenance cost per km (dependent variable) ($t=0$) [$t=1$]	n.a.	0	n.a.	0,25 (19.78) [-60.42]	n.a.	0,24 (18.48) [-58.20]	n.a.	0,22 (17.40) [-61.08]
β_0	Intercept ($t=0$)	n.a. (12.05)		n.a. (10.00)		n.a. (8.86)		n.a. (10.17)	
S	Target Δ service : 2006-2005 [$E(S_{ct-1}) - E(S_{ct-2})$] ($t=0$) Trajectory correction: (obs.-target) ₂₀₀₅ [$S_{t-2} - E(S_{ct-2})$] ($t=0$)					0.08** (4.63)	1	0.10** (5.18)	1
						-0.002 (-3.26)		-0.002 (-3.63)	
e_0	Segment length ($t=0$)	-0.121 (-2.95)	1	-0.073 (-1.97)	0.54 (8.37) [-7.25]	-0.070 (-1.92)	0.59 (8.70) [-6.09]	-0.028 (-0.75)	0.50 (8.37) [-8.34]
	Track length ($t=0$)	0.118 (1.07)		0.183 (2.34)		0.145 (1.87)		0.174 (2.26)	
	Number of switches ($t=0$)	0.189 (9.94)		0.411 (19.91)		0.391 (19.22)		0.434 (19.18)	
W+w	Cumulative+ current total tons (W+w) ($t=0$) [$t=1$]	0.288 (9.29)	0	0.262 (9.71)	0.58 (3.10) [-2.23]	0.279 (10.62)	0.38 (2.70) [-4.36]	0.247 (9.06)	0.39 (2.42) [-3.81]
t	Time since last regeneration (agerail) ($t=0$)	-0.101 (-0.97)			-0.021 (-0.27)	0	0.033 (0.40)	0	0.043 (0.56)
R₁	Same region: ρ_1 ($t=0$)							0.589 (6.87)	
	Log likelihood	-6977.59		-6868.13		-6857.01		-6825.77	
	Number of β_k estimated	6		6		8		8	
	Number of λ_k estimated	0		0		2		2	
	Difference in degrees of freedom	0		4		6		6	
	Variant run number	103		97		104		105	

* All t -statistics of β_k are conditional on the value of functional form parameters and those of the λ_k are unconditional.
 ** The sign of β_k is shown (the mean value of the variable is negative).

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

Column		3.B		4.B	
Elasticity $\eta(u), \lambda, (t\text{-stat.}=0)^*, [t\text{-stat.}=1]$		$\eta(u)$	λ	$\eta(u)$	λ
<i>Disaggregated variables replacing W+w in Columns 3.A and 4.A of Table 9 are shown.</i>		Total tons (W_c+w_c)		Total tons (W_c+w_c)	
W_c+w_c	TGV: high speed trains ($t=0$)	0.111 (3.87)	0.73 (4.33) [-1.62]	0.102 (3.32)	0.74 (3.85) [-1.38]
	GL: classic intercity trains (Corail) ($t=0$)	0.006 (0.17)		0.005 (0.12)	
	TER: regional trains ($t=0$)	0.048 (1.79)		0.039 (1.57)	
	IdF: Île-de-France trains ($t=0$)	-0.006 (-0.20)		0.001 (0.03)	
	FRT: freight trains ($t=0$)	0.081 (3.27)		0.072 (2.95)	
	HLP: locomotives ($t=0$)	0.052 (1.97)		0.041 (1.62)	
t	Time since last regeneration (agerail) ($t=0$)	0.004 (0.05)	0	0.036 (0.33)	0
R₁	Same region: ρ_1 ($t=0$)			0.586 (6.59)	
	Log likelihood	-6856.99		-6826.30	
	Number of β_k estimated	13		13	
	Number of λ_k estimated	2		2	
	Difference in degrees of freedom	0		1	
	Variant run number	106		107	

* All t -statistics of β_k are conditional on the value of functional form parameters and those of the λ_k 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 database⁶⁰.
- 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_2 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_3 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 proximity⁶¹ parameters π_2 and π_3 in addition to ρ_2 and ρ_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.
- xii) *within-region autocorrelation diversity*. Results presented above assume that within-region correlation is common to all administrative regions. But large gains in fit (not shown here) are made by allowing for different ρ_1 across the regions, demonstrating that Northern and Eastern groupings of regions matching the networks of the firms merged into SNCF in 1938 exhibit almost no correlation, in contrast with regions from the rest of the country (Gaudry & Quinet, 2014).
In toto, the role of within-region SNCF administrative practices⁶² 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 $[E(S_{c_t}) - E(S_{c_{t-1}})]$ and Trajectory Correction $[S_{t-1} - E(S_{c_{t-1}})]$ factors, combined with reactions to random Surprises in actual maintenance expenditures $[\varepsilon_{1t} = u_t - E(u_t)]$, mechanically imply a certain Service Supply level path ΔS_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 ΔS_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 $[E(S_{c_t}) - E(S_{c_{t-1}})]$ and $[S_{t-1} - E(S_{c_{t-1}})]$ 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 $[\varepsilon_{1t} = u_t - E(u_t)]$, the third variable

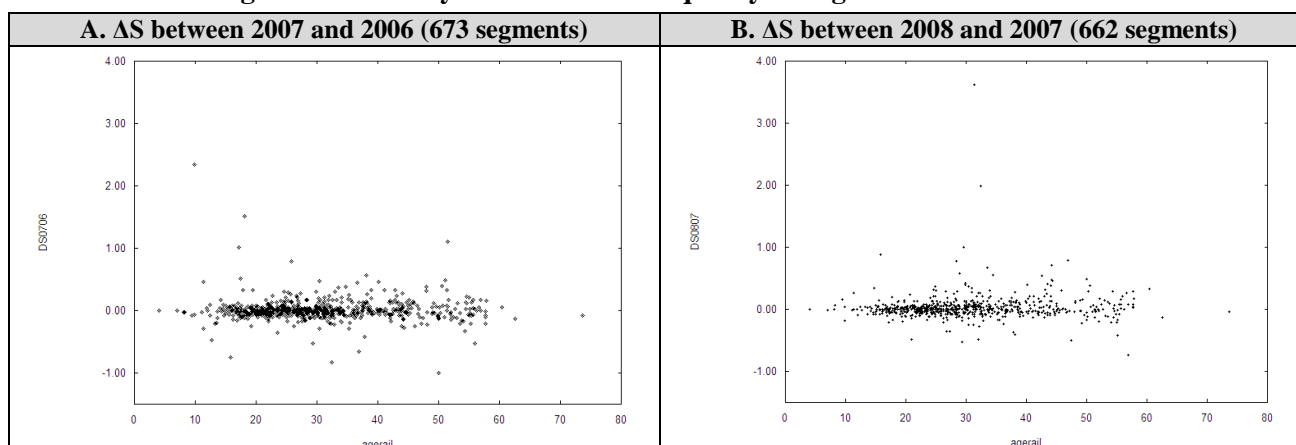
⁶⁰ The same holds for all models of Table 6 where that unreported BCT on cost varies between 0,18 and 0,28.

⁶¹ A parameter π_1 would not be meaningful: the matrix R_1 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_1 .

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 3⁶³.

Figure 8. Year-to-year track Service quality changes from 2006 to 2008



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 $[\varepsilon_{1t} = u_t - E(u_t)]$ 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 11. Explaining Service quality changes ΔS by factors defined for 2007 and 2006

Linear model: dependent variable ΔS_t :		2007-2006	2008-2007	2007-2006	2008-2007
Column		1	2	3	4
Elasticity $\eta(\Delta S)$ and t -statistic of β_k		$\eta(\Delta S)$	$\eta(\Delta S)$	$\eta(\Delta S)$	$\eta(\Delta S)$
β_0	Intercept ($t=0$)	n.a. (3.76)	n.a. (3.59)	n.a. (3.76)	n.a. (3.59)
$E(S_{ct}) - E(S_{ct-1})$ ($t=2007$)	Δ Target Service (2007-2006) ($t=0$)	0.209 (0.95)	-0.111 (-0.64)	0.209 (0.94)	-0.111 (-0.64)
$S_{t-1} - E(S_{ct-1})$ ($t=2007$)	Trajectory Corr. (obs.-target) ₂₀₀₆ ($t=0$)	-0.027 (-4.50)	-0.008 (-1.67)	-0.027 (-4.50)	-0.008 (-1.67)
$u_t - E(u_t)$ ($t=2007$)	Maintenance cost Surprise 2007 based on Column 4.A run of Table 9 ($t=0$)	0.003 (0.52)	0.000 (0.04)		
	Maintenance cost Surprise 2007 based on Column 4.A run of Table 13 ($t=0$)			0.003 (0.53)	0.000 (0.03)
Log likelihood		205.78	55.52	205.79	55.52
Number of β_k estimated		4	4	4	4
Variant run number		17	18	15	16

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

⁶³ We recall that in Table 9 there is no speed variable but that there is one in Table 13.

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

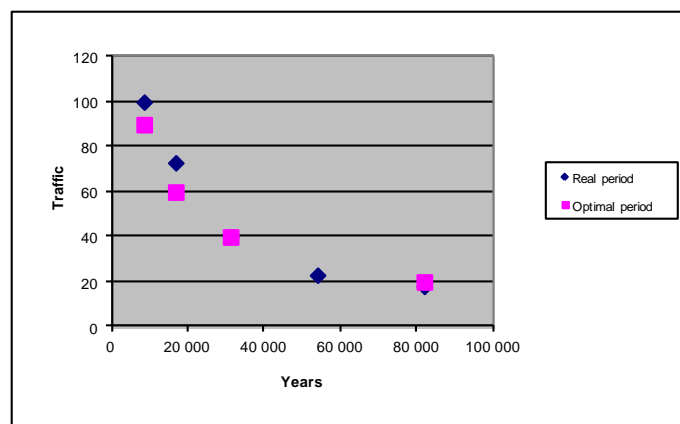


Table 12. Ratio between Renewal and Maintenance expenses for several levels of traffic

Traffic per day in tons	Observed Ratio renewal/maintenance	Calculated Ratio renewal/maintenance
82 000	0,34	0,34
41 000	0,32	0,19
10 250	0,08	0,09

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 Ikononov 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

- Andersson, M. (2006). *Marginal railway infrastructure cost estimates in the presence of unobserved effects*. Annex to Deliverable D3 of GRACE (Generalisation of Research on Accounts and Cost Estimation). ITS, University of Leeds, Leeds, 34 pages.
- Andersson, M., Björklund, G. (2012). Marginal Railway Track Renewal Costs: A Survival Data Approach. CTS Working Paper 2012:29, Centre for Transport Studies, 27 p. Royal Institute of Technology, Stockholm.
- Andersson, M., Smith, A., Wikberg, A., Wheat, P. (2012). Estimating the marginal cost of Railways Track Renewals using corner solution models. *Transportation Research A* 46, 6, 954-964.
- Antoni, M., Meier-Hirmer, C., Lévi, D. (2009). Modélisation des dépenses d'entretien du ballast. *Revue Générale des Chemins de Fer* 188, 21-35, novembre.
- Arnott, R., Davidson, R., Pines, R. (1983). Housing Quality, Maintenance and Rehabilitation. *The Review of Economic Studies* 50, 3, 467-494, July.
- Belsley, D., Kuh, E., Welsch, R.E. (1980). *Regression Diagnostics*. John Wiley, New York.
- Ben-Akiva, M., Gopinath, D. (1995). Modelling infrastructure performance and user costs. *Journal of Infrastructure Systems* 1, 1, 33-43.
- Bereskin, C.G. (2000). Estimating Maintenance-of-Way Costs for U.S. Railroads After Deregulation. *Transportation Research Record* 1707, 13-21.
- Berndt, E.K., Hall, B.H., Hall, R.E., Hausman, J.A. (1974). Estimation and Inference in Nonlinear Structural Models. *Annals of Economic and Social Measurement*, 3, 653-665.
- Blum, U.C.H., Bolduc, D., Gaudry, M.J.I. (1995/1996). From Correlation to Distributed Contiguities: a Family of AR-C-D Autocorrelation Processes. *Zeitschrift für Regionalwissenschaft* 16/17, 99-128.
- Bolduc, D. (1987). *On the estimation of models with generalized SAR(1) processes on the residuals of a regression*. Publication CRT-508, Centre de recherche sur les transports, Université de Montréal.
- Bolduc, D., Dagenais, M.G., Gaudry, M.J.I. (1989). Spatially Autocorrelated Errors in Origin-Destination Models: A New Specification Applied to Aggregate Mode Choice. *Transportation Research B* 23, 5, 361-372.
- Box, G.E.P., Cox, D.R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society, Series B*, 26, 211-243.
- Caves, D.W., Christensen, L.R., Tretheway, M.W. (1980). Flexible Cost Functions for Multiproduct Firms. *Review of Economics and Statistics*, 477-481 (Notes Section).
- Caves, D.W., Christensen, L.R., Tretheway, M.W., Windle, J. (1985). Network effects and the measurement of returns to scale and density for U.S. railroads. Ch. 4 in A. F. Daughety (ed.), *Analytical Studies in Transport Economics*, 97-120, Cambridge University Press, Cambridge.
- Cho, J.S., Ishida, I. (2012). Testing for the effects of omitted power transformations. *Economics Letters* 117, 1, 287-290, October.
- Cliff, A.D., Ord, J.K. (1981). *Spatial Processes: Models and Applications*. Pion Limited, London.
- Dagenais, M.G., Gaudry, M.J.I., Liem, T.C. (1987). Urban Travel Demand: the Impact of Box-Cox Transformations with Nonspherical Residual Errors. *Transportation Research B* 21, 6, 443-477.
- Erkel-Rousse, H. (1995). Détection de la multicolinéarité dans un modèle linéaire ordinaire: quelques éléments pour un usage averti des indicateurs de Belsley, Kuh et Welsch. *Revue de Statistique Appliquée* 43, 4, 19-42.
- Feldstein, M.S., Rothschild, M. (1974). Towards an economic theory of replacement investment. *Econometrica* 42, 3, 393-423.
- Friesz, T.L., Fernandez, J.E. (1979). A model of optimal transport maintenance with demand responsiveness. *Transportation Research B*, 13, 4, 317-339.
- García Álvarez, A., Martín Cañizares, M. del Pilar (2008). *Consumo de energía y emisiones asociadas al transporte por ferrocarril*. Monografía 16, Monografías Enertrans, 82 p.
- Gaudry, M. (1980). A Study of Aggregate Bi-Modal Urban Travel Supply, Demand, and Network Behavior Using Simultaneous Equations with Autoregressive Residuals. *Transportation Research B* 14, 1-2, 29-58.
- Gaudry, M. (2004). The Robust Gravity Form in Transport and Trade Models. *Zeitschrift für Kanada-Studien* 24, 1, Band 44, 144-161.
- Gaudry, M.J.I., Blum, U.C.H. (1988). An Example of Correlation Among Residuals in Directly Ordered Data. *Economics Letters* 26, 4, 335-340.

- Gaudry, M., Dagenais, M.G. (1979). Heteroscedasticity and the Use of Box-Cox Transformations. *Economics Letters* 2, 3, 225-229.
- Gaudry, M., Quinet, É. (2003). *Rail track wear-and-tear costs by traffic class in France*. Publication AJD-66, Agora Jules Dupuit, Université de Montréal, www.e-ajd.net, et W.P. 04-02, Centre d'Enseignement et de Recherche en Analyse Socio-économique, École nationale des Ponts et Chaussées (ENPC-CERAS), www.enpc.fr/ceras/labo/accueil.html, 26 pages, November.
- Gaudry, M., Quinet, É. (2010). *Track Wear and Tear Cost by Traffic Class: Functional Form, Zero-Output Levels and Marginal Cost Pricing Recovery on the French Rail Network*. Publication AJD-130, Agora Jules Dupuit, Université de Montréal, et Working Paper N° 2009-32, Paris-Jourdan Sciences Économiques, 50 pages, 26 avril.
- Gaudry, M., Quinet, É. (2011). *Optimisation de l'entretien et de la régénération d'une infrastructure: exploration d'hypothèses*. Publication AJD-136, Agora Jules Dupuit, Université de Montréal, et Working Paper N° 2011-03, Paris-Jourdan Sciences Économiques 19 pages, Version 2 du 15 janvier.
- Gaudry, M., Quinet, É. (2014). Correlation within SNCF administrative regions among track segment maintenance cost equation residuals of a country-wide model. In Funck, R., Rothengatter, W. (Hrsg.), *Man, Environment, Space and Time - Economic Interactions in Four Dimensions. Karlsruhe Papers in Economic Policy Research*, Volume 34, Nomos Verlag, p. 207-232.
- Gaudry, M., Blum, U., Liem, T. (2000). Turning Box-Cox, including Quadratic Forms in Regression. In Gaudry, M., Lassarre, S., (eds), *Structural Road Accident Models: The International DRAG Family*, Pergamon, Elsevier Science, Oxford, Ch. 14, 335-346.
- Gu, W., Ouyang, Y., Madanat, S. (2012). Joint optimisation of pavement maintenance and resurfacing planning. *Transportation Research B* 46, 511-519.
- Hausmann, U.G., Suo, W. (1995a). Singular optimal stochastic controls I: Existence. *SIAM Journal on Control and Optimization* 33, 3, 916-936, May.
- Hausmann, U.G., Suo, W. (1995b). Singular optimal stochastic controls I: Dynamic programming. *SIAM Journal on Control and Optimization* 33, 3, 937-959, May.
- Idström, T. (2002). *Reforming the Finnish railway infrastructure charge – an econometric analysis of the marginal costs of rail infrastructure use*. In Finnish. Masters Thesis. Department of Economics. University of Jyväskylä.
- Ingram, G.K., Liu, Z. (1997). *Motorization and the provision of roads in countries and cities*. Policy Research Paper 1842, Policy Research Dissemination Center, The World Bank, 38 p., November.
- Johansson, P., Nilsson, J.-E. (2002). *An Economic Analysis of Track Maintenance Costs*. Deliverable 10 Annex A3 of UNITE (UNification of accounts and marginal costs for Transport Efficiency), Funded by EU 5th Framework RTD Programme. ITS, University of Leeds, Leeds, 23 p.. Online: <http://www.its.leeds.ac.uk/projects/unite/>.
- Johansson, P., Nilsson, J.-E. (2004). An economic analysis of track maintenance costs. *Transport Policy* 11, 277-286.
- Kobayashi, K., Kaito, K., Lethanh, L. (2012). A statistical deterioration forecasting method using hidden Markov model for infrastructure management. *Transportation Research B* 46, 544-561.
- Kouider, E., Chen, H. (1995). Concavity of Box-Cox log-likelihood function. *Statistics and Probability Letters* 25, 171-175.
- Koyck, L. (1954). *Distributed Lags and Investment Analysis*. North Holland, Amsterdam.
- Kushner, H.J., Dupuis, P.G. (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer-Verlag.
- Lévi, D., Antoni, M., Meier-Hirmer, C. (2010). Politique de maintenance des voies, deuxième partie : l'influence du taux d'actualisation sur la politique de renouvellement. *Revue Générale des Chemins de Fer* 190, 6-15, janvier.
- Liem, T.C., Dagenais, M., Gaudry, M. (1993). *LEVEL: The L-1.4 program for BC-GAUHESEQ regression —Box-Cox Generalized AUtoregressive HEteroskedastic Single EQUation models*. Publication CRT-510, Centre de recherche sur les transports, Université de Montréal, 41 p. www.e-ajd.net.
- Link, H. (2009). CATRIN (Cost Allocation of TRansport INfrastructure cost), Deliverable 8, Rail Cost Allocation for Europe – Annex 1C - *Marginal costs of rail maintenance in Austria*. VTI, Stockholm, February, 23 pages.

- Marti, M., Neuenschwander, R. (2006). *Case study 1.2E: Track Maintenance Costs in Switzerland*. Annex to Deliverable 3 of GRACE (Generalisation of Research on Accounts and Cost Estimation). University of Leeds, 23 pages.
- Martland, C.D. (1992). Rail freight service productivity from the manager's perspective. *Transportation Research A* 26, 6, 457-469, November.
- Mohring, H. (1972). Optimization and Scale Economies in Urban Bus Transportation. *American Economic Review*, 591-604.
- Montmarquette, C., Blais, A. (1987). A Survey Measure of Risk Aversion. *Economics Letters* 25, 1, 27-30.
- Newbery, D.M.G. (1988a). Charging for roads. *World Bank Research Observer* 3, 2, 119-138.
- Newbery, D.M.G. (1988b). Road damage externalities and road user charges. *Econometrica* 56, 2, 295-316, March.
- Ongaro, D., Iwnicki, S. (2009). CATRIN (Cost Allocation of TRansport INfrastructure cost), D8 – Rail Cost Allocation for Europe - Annex 2 - *Railway Infrastructure Wear and Tear for Freight and Passenger Trains in Sweden*. VTI, Stockholm, 21 pages, February.
- Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 349, 120-126.
- Rail Concept (2006). *Tarifification d'usage de l'infrastructure : Assistance à la détermination de la marginalité des coûts de régénération du réseau*. Rapport d'étude, Version 8, Direction de la Stratégie, Société nationale des chemins de fer français (SNCF), 50 p., 12 décembre.
- Rivier, R., Puttalaz, Y. (2005). *Audit sur l'état du réseau ferré national français*. Laboratoire d'Intermodalité des Transports et de Planification, École Polytechnique fédérale de Lausanne, 30 pages, 7 septembre.
- Robert, W.E., Chapman, J.D., Martland, C.D. (1997). Effects of Axle Loads and Train Capacity on Heavy Haul Network Performance. *Proceedings of the 39th Meeting of the Transportation Research Forum*, Volume 2, 24 pages, October.
- Rosett, R.N., Nelson, F.D. (1975). Estimation of the two-limit Probit regression model. *Econometrica* 43, 141-146.
- Rust, J. (1987). Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. *Econometrica* 55, 5, 999-1033, September.
- Sathay N., Madanat, S. (2011) A bottom up solution for the multi-facility optimal pavement resurfacing problem. *Transportation Research B* 45, 1004–1017.
- Schlesselman, J. (1971). Power families: a note on the Box and Cox transformation. *Journal of the Royal Statistical Society Series B* 33, 307-31.
- Small, K.A., Winston, C., Evans, C.A. (1989). *Road Work. A New Highway and Investment Policy*. The Brookings Institution, 128 p.
- Spitzer, J. J. (1984). Variance estimates in models with the Box-Cox transformations: Implications for estimation and hypothesis testing. *Review of Economics and Statistics* 66, 645-652.
- SNCF (1989). *Armement, ballastage et entretien de la voie, généralités classement des lignes en groupes au point de vue de la maintenance de la voie*. Notice générale EF 2A 1 n°1, SNCF Équipement, 5 p., 12 décembre.
- Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica* 25, 24-36.
- Tran, C.-L., Gaudry, M. (2008). *LEVEL: The L-2.1 estimation procedures for BC-DAUHESEQ (Box-Cox Directed AUtoregressive HEteroskedastic Single EQuation) regression*. Publication AJD-108, Agora Jules Dupuit, Université de Montréal, 26 p., September 2008. www.e-ajd.net.
- Tsunokawa, K., Schofer, J.L. (1994). Trend curve optimal control model for highway pavement maintenance: case study and evaluation. *Transportation Research A*, 28, 2, 151-166.
- UIC, Union Internationale des Chemins de Fer (1989). *Classification des voies des lignes au point de vue de la maintenance de la voie*. Fiche n° 714 R, 7 pages, Paris, 3^{ème} édition, 1^{er} janvier.
- Wheat, P., Smith, A., Nash, C. (2009). *CATRIN (Cost Allocation of TRansport INfrastructure cost), Deliverable 8 – Rail Cost Allocation for Europe*. Funded by Sixth Framework Programme. VTI, Stockholm. 70 pages, March.

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 , to be set equal to 0, is:

$$(A-1) \quad \frac{J'(T)(1 - e^{-jT}) + jDe^{-jT} - jJ(T)e^{-jT}}{(1 - e^{-jT})^2} = 0.$$

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 with very low traffic), this solution is a maximum if the second derivative is negative.

The sign of this second derivative is the same as the sign of its numerator, which is:

$$N = J''(1 - e^{-jT}) + jJ'e^{-jT} - j^2De^{-jT} - jJ'e^{-jT} + j^2Je^{-jT}$$

namely, replacing D by its expression deduced from (A-1):

$$N = J''(1 - e^{-jT}) + jJ'(1 - e^{-jT}) = (1 - e^{-jT})(J'' - jJ')$$

or, as $J'(T) = H(T) = qg(S(T))e^{-jT}$ and $J'' = -qjg(S(T))e^{-jT} + qg'(S(T))[dS(T)/dT]e^{-jT}$:

$$N = (1 - e^{-jT})qe^{-jT}g'(S(T))[dS(T)/dT].$$

The sign of N is the same as the sign of $dS(T)/dT$ because the remaining RHS terms are positive. As $dS(T)/dT$ is negative due to the fact that, in the final phase, $dS/dT = -f(K, Q, q, t)$, it turns out that N is negative, then the second derivative of $M(T)$ is negative and, finally, the extremum of $M(T)$ 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:

$$\mathbf{E} \left(\int_0^T [-u(t) - \alpha e^{-\lambda S_t}] e^{-jt} dt \right).$$

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

$$J(t, x) = \max_{u, u(s) \in [0, m]} \left\{ \mathbf{E}_{t,x} \left(\int_t^T [-u(s) - \alpha e^{-\lambda S_s}] e^{-j(s-t)} ds \right) \right\},$$

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

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

now with $u \in \mathbf{R}$, $x \in \mathbf{R}$, $t \leq T$ 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 Δx is given, the time-step Δt 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 $J_x(k(q,t)u(t) - f(q,t))$: given $A = (-k(q,t)u + f(q,t)) / \Delta x$, we effectively approximate $J_x(t, x)$ by $[J(t, x + \Delta x) - J(t, x)] / \Delta x$ if $A \geq 0$ and by $[J(t, x) - J(t, x - \Delta x)] / \Delta x$ if $A < 0$, switching regimes in accordance with the sign of A . For $J_{xx}(t, x)$, we use $[J(t, x + \Delta x) - 2J(t, x) + J(t, x - \Delta x)] / \Delta x^2$, 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 Δx and Δt approach 0.

11. Appendix 3. Adding speed to maintenance cost models

Table 13. Phase B Maintenance cost u with max. speed by train type v_c (580 obs., 2007)

A. (common weighted speed)		1		2		3.A		4.A	
Elasticity $\eta(u), \lambda, (t\text{-stat.}=0)^*, [t\text{-stat.}=1]$		$\eta(u)$	λ	$\eta(u)$	λ	$\eta(u)$	λ	$\eta(u)$	λ
u	Total maintenance cost per km (dependent variable) $(t=0)$ $[t=1]$	n.a. (11.53)	0	n.a. (9.53)	0,25 (19.34) [-59.48]	n.a. (8.66)	0,24 (18.04) [-56.82]	n.a. (10.02)	0,22 (16.82) [-59.14]
β_0	Intercept $(t=0)$	n.a. (11.53)		n.a. (9.53)		n.a. (8.66)		n.a. (10.02)	
S	Target Δ service : 2006-2005 [E(S _{ct-1}) - E(S _{ct-2})] $(t=0)$					0.08** (4.42)	1	0.10** (5.06)	1
	Trajectory correction: (obs.-target) ₂₀₀₅ [S _{t-2} - E(S _{ct-2})] $(t=0)$					-0.002 (-3.11)		-0.002 (-3.47)	
e₀	Segment length $(t=0)$	-0.125 (-3.00)	1	-0.096 (-2.31)	0.52 (8.29) [-7.63]	-0.082 (-2.07)	0.58 (8.64) [-6.28]	-0.039 (-0.99)	0.49 (8.23) [-8.46]
	Track length $(t=0)$	0.116 (1.05)		0.179 (2.32)		0.145 (1.87)		0.176 (2.30)	
	Number of switches $(t=0)$	0.190 (9.95)		0.415 (19.96)		0.394 (19.34)		0.439 (19.75)	
(W+w)	Cumulative+ current total tons (W+w) $(t=0)$ $[t=1]$	0.280 (8.22)	0	0.254 (9.33)	0.64 (3.01) [-1.67]	0.277 (10.37)	0.40 (2.70) (-4.03]	0.245 (8.82)	0.42 (2.48) [-3.53]
V	Speed weighted by (W _c +w _c) shares $(t=0)$ $[t=1]$	0.032 (0.54)		0.067 (1.88)	2.65 (1.19) [0.74]	0.017 (1.04)	3.62 (0.43) [0.31]	0.017 (0.84)	3.52 (0.41) [0.30]
t	Time since last regeneration (agerail) $(t=0)$	-0.086 (-0.79)		0.021 (0.25)	0	0.051 (0.60)	0	0.059 (0.75)	0
R₁	Same region: ρ_1 $(t=0)$							0.589 (6.89)	
	Log likelihood	-6977.45		-6866.26		-6856.44		-6825.26	
	Number of β_k estimated	7		7		9		9	
	Number of λ_k estimated	0		0		3		3	
	Difference in degrees of freedom	0		4		6		6	
	Variant run number	191		192		193		194	
B.(speed disaggregated by train type)		3.B				4.B			
Elasticity $\eta(u), \lambda, (t\text{-stat.}=0)^*, [t\text{-stat.}=1]$		$\eta(u)$	λ	$\eta(u)$	λ	$\eta(u)$	λ	$\eta(u)$	λ
<i>Disaggregated variables replacing W+w and V in Columns 3.A and 4.A of Table 13.A are shown.</i>		Total tons (W _c +w _c)		Speed (v _c)		Total tons (W _c +w _c)		Speed (v _c)	
(W_c+w_c) & (v_c)	TGV: high speed trains $(t=0)$	0.097 (2.66)	0.67 (3.94) [-1.91]	0.096 (1.69)	2.18 (1.52) [0.82]	0.070 (1.92)	0.71 (3.19) [-1.33]	0.137 (2.20)	1.63 (1.59) [0.62]
	GL: classic intercity trains (Corail) $(t=0)$	0.004 (0.11)		-0.540 (-2.48)		0.010 (0.24)		-0.436 (-2.60)	
	TER: regional trains $(t=0)$	0.052 (1.69)		0.588 (1.89)		0.049 (1.77)		0.358 (1.48)	
	IdF: Île-de-France trains $(t=0)$	0.004 (0.11)		-0.028 (-1.48)		-0.002 (-0.05)		-0.017 (-0.66)	
	FRT: freight trains $(t=0)$	0.089 (3.30)		-0.018 (-0.09)		0.070 (2.47)		-0.005 (-0.03)	
	HLP: locomotives $(t=0)$	0.051 (1.77)		0.188 (0.95)		0.045 (1.55)		0.135 (0.85)	
t	Time since last regeneration (agerail) $(t=0)$	0.017 (0.19)	0	0	0.037 (0.41)	0	0	0	
R₁	Same region: ρ_1 $(t=0)$							0.587 (6.51)	
	Log likelihood	-6850.68				-6820.58			
	Number of β_k estimated	19				19			
	Number of λ_k estimated	3				3			
	Difference in degrees of freedom	0				1			
	Variant run number	92				96			